2008 IEEE Intelligent Vehicles Symposium
Eindhoven University of Technology
Eindhoven, The Netherlands, June 4-6, 2008

Real-time Vision-based Infotainment User Determination for Driver Assistance

Shinko Y. Cheng and Mohan M. Trivedi

Abstract— **Knowledge of driver body pose can be used in many applications. In this paper, we develop and evaluate a novel real-time computer vision algorithm to robustly discriminate which of the front-row seat occupants is accessing the infotainment controls. The knowledge of user type can alleviate driver distraction and maximize passenger infotainment experience. The system consists of a visible and near-infrared imaging device observing the front-row seat area in the vehicle. Using histogram-of-oriented-gradients to describe the image area over the controls, a support-vector-machine was shown to be able to provide 96.8% average correct classification rate. This approach represents an alternative of detecting and tracking the hand movements and then classifying the hands into the respective classes.**

I. Introduction

A broad new array of information devices are finding its place in today's vehicles. The infotainment device has graduated from a term referring to the radio to a collective word describing the navigational, vehicle status, climate control, personal cell-phone control, MP3 player control, web-browsing, and even television functionalities of the front console area of the vehicle [1], [2]. With all of these opportunities for the drivers to be distracted, the solution has been to limit the functionality or the information content from these infotainment systems, and make them less distracting. Often, the information provided by these devices become oversimplified and less useful for passengers. It is far more desirable to be able to both alleviate driver distraction and provide the passenger access to better information.

We propose a novel real-time Vision-based User Determination (VUD) system to determine which front-row seat occupant is accessing the infotainment controls. The controls of the infotainment system consist of buttons and a knob with rotational and directional degrees-of-freedom. They are centrally located in the aisle area between the driver and the passenger. The VUD is intended to improve vehicle safety by providing information about the class of the user (driver or passenger) so that the infotainment system can in turn provide an adequate level of information. Driver distraction can be mitigated and the passenger can be granted full-access to the information device.

The challenges of developing such a system center on developing a robust vision-based classification algorithm that is capable of maintaining high performance in all operating conditions of the vehicle. The performance should be maintained through changes in the appearance of people, changes in lighting from different times of the day, and changes in the camera position. Because vehicles are likely to receive maintenance only between several months of operation, if at all, much of the functionality must also require little or no maintenance. These attributes were achieved with appropriate choices of the pattern classifier and system components.

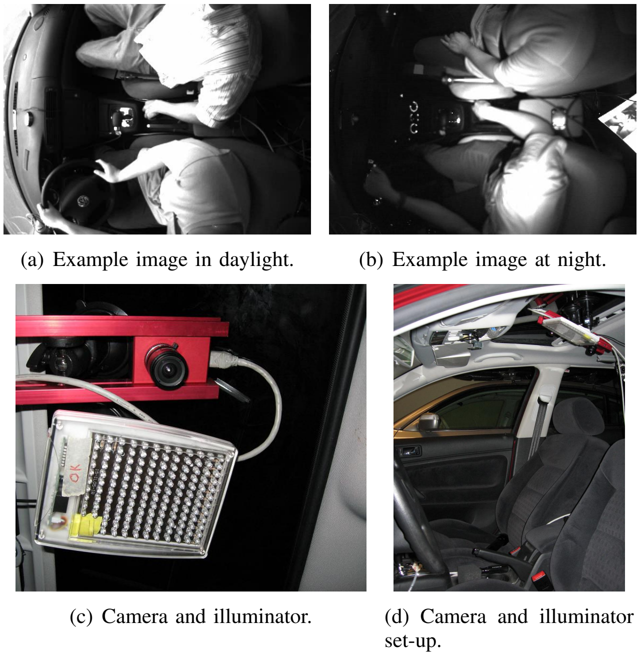

(a) Example image in daylight. (b) Example image at night.
(c) Camera and illuminator. (d) Camera and illuminator set-up.

Fig. 1. Example images captured during the day and night, and the positions of the camera and illuminator in the LISA-P test-bed for the VUD system.

The proposed system takes as input visible and near-infrared images of the front-seat and center-console area illuminated with a bank of near-infrared LEDs. The system then uses these images to determine which front-row occupant is accessing the device, if anyone at all. The histogram of oriented gradients (HOG) image descriptor was chosen to create the feature vectors [3]. The system then utilizes the kernel support-vector-machine (SVM) to classify the observed image features into the three classes: driver, passenger or no-one. The processing takes on average 10.7 milliseconds per frame on an Intel Pentium D 3.2GHz PC.

The evaluation of this approach uses 2 metrics: the correct classification rates of each class forming the confusion matrix, and the average correct detection rate of the three classes. In the training process, care was taken to ensure that a representative data-set was used, and the usual cross-

S. Y. Cheng and M. M. Trivedi is with the University of California, San Diego, La Jolla CA, 92037 USA e-mail: {sycheng,mtrivedi}@ucsd.edu www:cvrr.ucsd.edu/lisa. S. Y. Cheng is presently at HRL Laboratories, LLC

978-1-4244-2569-3/08/$20.00 ©2008 IEEE.

validation techniques were employed to gauge the generalizability of the pattern classifier [4], [5]. Data was collected at 4 different times of day with 8 different individuals for a total of 18 test-runs, over 1-hour of data at 30 frames-per-second. A large representative data-set allows for an understanding of the performance on a wider range of operating conditions. We also analyzed the system's invariance to translation in the x- and y-directions of the image patch, where the features are extracted. These qualities influences the flexibility in camera placement during the installation process. The resulting trained system can correctly recognize whether the driver, passenger, or no one has their hand over the infotainment controls with better than 95% average correct classification rate.

The results of this research show that it is possible to achieve extremely reliable pattern recognition using a visible-wavelength imaging sensor in the volatile environment of the vehicle where illumination is changing constantly and spread widely in intensity. The results demonstrates a way to extract discriminant information for the abstract idea determining when a system should provide less distracting or more informative driving assistance.

II. Related Research

The idea of tailoring vehicle information system functions, input and output devices, and user interfaces based upon whether the user is the driver or passenger is not a new one. Chou *et al.* [6] have proposed the use of weight sensors to determine the presence of a passenger before enabling full-functionality of an infotainment system. Harter *et al.* [7] too proposed to switch between "enhanced functionality" and "base functionality" of the information system by determining the presence of a passenger but use proximity sensors instead. Furthermore, Harter *et al.* take a step further to determine when to engage "base functionality" by determining whether the driver has gazed into the infotainment monitor more than 2 seconds (considered too long) using a vision-based eye gaze tracking system. We propose to use the same vision modality but analyze the hands of the occupants rather than the driver's head to determine when to switch between functionality modes. Our proposed solution can replace or complement these other systems by providing the following advantages:

1) The proposed system is arguably simpler to implement and maintain than an eye gaze tracking system. The proposed solution requires no camera nor person calibration.
2) The proposed system actively monitors the hands to detect the intent to access the information system as soon as the hand nears the controls.
3) The proposed system actively detects the case when no one is accessing the information system. This case can be used to automatically show and hide access controls in the display, and can be interpreted as having a more attentive driver, therefore require less driver assistance. Weight and proximity sensor-based systems can detect if no-one is present, but cannot detect if someone is there, but no-one is accessing the system.
4) The proposed system can detect difficult situations with occlusion, including the partial occlusion from the other occupant's hand.

Tab. I summarizes the related work in user determination systems.

The problem of hand image based user determination can be approached in two ways: 1) Active tracking of occupant hands as the hand passes into and out of the area over the infotainment controls (or region-of-interest) to detect intent to access. 2) Learn the appearance of the driver's hand, the passenger's hand or no one's hand over the the region-of-interest. A number of works have addressed the first approach.

The first type consists of a detector which locates the hand in the images, and then tracks the hand. Tracking is associating one hand detection with the next over time. The challenge of this approach is in obtaining a good description of the appearance of the hand in its various poses, and a way to efficiently check all areas of the image for the existence of a hand. The characteristics of a good descriptor is one that would correctly associate two hand detections of the same hand in different positions and poses.

One such hand detection algorithm devised by Kölsch *et al.* [8] employs a cascade of boosted classifiers using Haar-wavelet-like image features and their extensions to determine whether an image patch, among all possible patches in an image, consists of a hand or not. The rates reported for real-time operation were very good (92% detection rate with a false positive rate of 1e-8), but the approach detected hands in a standard canonical position: fingers up and thumb to the left and 7 other similar forms. Kölsch *et al.* proposed using a flock-of-features approach to track the positions of the hands after initial detection, thus addressing the problem of maintaining track of hands through its many poses.

We employed a similar detection technique with long-wavelength thermal infrared images of hands [9]. The thermal infrared modality is especially appropriate in the vehicular domain because image appearance is not at all affected by changing visible illumination conditions and pixel intensity of skin stays relatively constant. Because of the special quality of hands in thermal images, this detector was also effective in detecting hands in various poses. The detector is applied on each incoming frame and the multiple hands are tracked using the Kalman filter and Probabilistic Data Association Filter (PDAF) to disambiguate one track from another. This approach however suffers from the use of a thermal camera, which are still expensive as compared to the visible-wavelength camera.

We address the user determination problem using the second approach, which is the direct classification of images of the infotainment controls region. No tracking is required, although some rudimentary filtering of the classification responses will increase the correct classification rate; we describe later on. This approach takes advantage of the fact that the region-of-interest has a stable background, which is

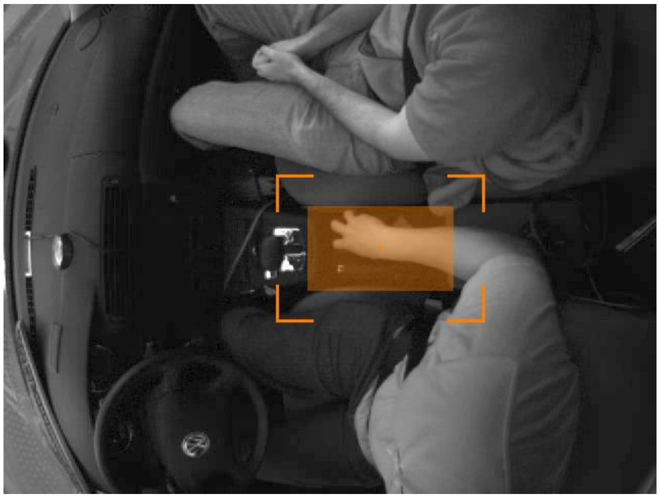


Fig. 2. Image region-of-interest used to determine the user in the VUD system. Patch as depicted is used for the final VUD system.

the vehicle interior. To the best of our knowledge, no other work addresses the user determination problem in a similar way.

III. Vision-based User Determination System

The Vision-based User Determination (VUD) system determines the user of the individual whose hand is accessing the infotainment device by classifying patches of captured images of the front-row seat area in a passenger vehicle. *User* is defined as one of three categories: 1) driver, 2) passenger, and 3) no-one. The infotainment controls are assumed to be positioned just aft of the gear-shift, forward of the hand-rest, and besides the hand-brake.

We adopt the visible and near-infrared spectrum imaging modality to provide the observations for determining whose hand is on the infotainment controls. The primary reasons are the passive nature of the camera; at night, the front-row seat area can still be captured by illuminating the area with near-infrared illuminators without distracting the occupants. Example images are shown in Fig. 1(a) and 1(b). The VidereDesign STH-MDCS2-VAR camera capturing 640x480 px gray-scale images at 30Hz, and the SUPERCIRCUITS IR14 infrared illuminator were used.

The overall system has three stages: data capture, feature extraction, and classification. This is the basic procedure for all pattern classification systems.

The system starts with the capture of monochrome images. A rectangular image patch that spans between the edges of the driver's and passenger's seat and the length between the gear-shifter and the hand rest is extracted. An example image captured from the front seat area and the image patch is shown in Fig. 2. The histogram of orientation gradients (HOG) description of the image patch is calculated, and then presented to the multi-class kernel support-vector-machine (SVM) classifier to determine which of 3 events occurred: the driver's hand, passenger's hand, or no one's hand accessed the infotainment controls.

A. HOG Feature Extraction

The HOG descriptor for an image patch is created by first taking the gradient of the patch. The resulting gradient image is then divided into smaller rectangular patches of pixels specified by the number of x-bins and y-bins in the x and y directions. Within each rectangle, a histogram is collected of the angles of the gradient vectors of each pixel. In generating this orientation histogram, the number of o-bins that span gradient orientations of 0 and 360 degrees describes the length of each histogram for each rectangle. All the orientation histograms are then vectorized and concatenated to form the feature vector $\mathbf{x}$. Altogether, 3 parameters determine the dimensions of the final feature vector: the number of divisions along the x, y, and the number of bins in the orientation histogram. In the VUD, a 2x2 grid of bins with 8 slices in the range of possible gradient orientations results in a 32-dimensional feature vector (2x2x8) for each image patch. The HOG descriptor is also referred to as the SIFT descriptor [10].

B. SVM Classifier

The objective of any classifier is to correctly assign the observed feature vector $\mathbf{x}$ to its corresponding label or class k of K classes. Mathematically, this refers to creating a set of discriminant functions $g_i(\mathbf{x})$ for $i \in 1, ..., K$ such that $g_k(\mathbf{x})$ produces the highest value when $\mathbf{x}$ corresponds to class k. The discriminant function is parameterized by a set of variables represented as θ. Training a classifier refers to optimizing the parameters θ such that classifier will correctly classify as many input features vectors $\mathbf{x}$ as possible in a given training data-set. The data-set contains example feature vectors and their associated class values or target values, which are manually assigned.

The support-vector-machine was chosen to be the underlying classifier for this application. The SVM classifier takes the form

$$g(\mathbf{x}) = \text{sign}\left(\sum_{n=1}^{N} a_n t_n k(\mathbf{x}, \mathbf{x}_n) + b\right) \qquad (1)$$

where $t_n \in \{-1, 1\}$ is the target value for feature $\mathbf{x}_n$, while a_n and b are the weights to be optimized. The key result of SVM is the sparse solutions for a_n, i.e. many terms are zero. The effect of this is efficiency in classification; only a very small subset of N examples actually need to be retained to calculate $g(\mathbf{x})$. The feature vectors $\mathbf{x}_n$ for which the corresponding a_n is non-zero are also called the "support vectors."

The vector of coefficients a_n can be seen as a hyper-plane (n-dimensional plane) separating the two sets of features (one corresponding to $t = -1$, and the other $t = +1$) in the feature space. The optimal condition is when the two sets of features are separated by this hyper-plane with the largest possible spacing, or margin, between the hyper-plane and the closest feature vectors in this feature space. For a more detailed treatment of the SVM, please refer to [5], [11], [12].

SVM by itself is a two-class classifier. Multi-class SVM is used to classify the feature vectors into the 3 classes for the

TABLE I

RELATED WORK ON USER DETERMINATION FOR INFORMATION SYSTEM MODE CONTROL.

	Objective	Method	Result	Cues used	Comment
Chou *et al.* ('99) [6]	User Discrimination Control of Vehicle Information Systems	Occupant Presence using weight sensor.	N/A	Weight sensor	-
Harter *et al.* ('02) [7]	User Discrimination Control of Vehicle Information System	Prolonged Driver Eye Gaze Detection Passenger Presence using multi-modal sensors	N/A	Weight sensor Driver eye-gaze sensor Proximity sensor seat-belt tension sensor	-
Approach presented in this paper	Determine whether Driver/ Passenger/ No-one's hand is present.	HOG feature, 3-class SVM classifier	96.8% Avg. CCR	Hand imaging sensor	Analyzed sample-by-sample correct detection rate.

VUD system. This extension is achieved by training 3 SVM classifiers with the one-versus-rest approach. Each two-class SVM classifier will be trained with the feature vectors annotated as one class with a target value $t = +1$, and the other feature vectors grouped together with target value $t = -1$. There are three two-class SVM classifiers in all. Each SVM classifier can be seen as one of 3 discriminant functions, and the final classification is done by determining which of the 3 functions yields the highest value, i.e. determine which of 3 regions in feature space does the feature vector lie in the deepest. Formally, the classified result C is given by

$$C = \arg\max_{C_k} g_{C_k}(\mathbf{x}) \tag{2}$$

Because the raw features do not neatly lie in their respective sides of this hyper-plane in the raw feature space, i.e. the raw features $\mathbf{x}$ are not "linearly separable," we embed these features onto a higher-dimensional kernel space $k(\mathbf{x}_i, \mathbf{x}_j)$. Depending on the chosen kernel function, different aspects of the spatial configuration of the raw features are emphasized by the kernel function. The radial-basis-function (RBF) for example is one type of kernel function where each kernel function depends on the distance (usually Euclidean), from a specified mean μ_i. The means are set to the feature points, thus producing as many kernel functions as there are example feature points. The use of the kernel function effectively projects the raw feature space from a d-dimensional space onto an N-dimensional space, where $d << N$. Arriving at a sparse solution where only a small subset of N is retained as part of the classifier is the problem that SVM solves.

The proposed system was prototyped with the SVM implementation in OpenCV Machine Learning Library [13].

IV. EXPERIMENTAL EVALUATION

The experimental evaluation consists of the definition of relevant performance metrics, validation of the performance evaluation as being representative of the true performance, validation of the optimality of the algorithm parameters, and finally discussion of the results pertaining to robustness of the system in conditions in which they may fail.

A. Performance Metric

The evaluation of the VUD system utilizes 2 metrics: 1) the confusion-matrix summarizing the classification rate of a feature vector of a given class as a given class, and 2) the average correct classification rate among the three classes. The confusion matrix consists of 3 rows and columns for each of the 3 classes. The row represents the actual class of novel examples (excluding examples used for the training of the classifier), and the columns represent the predicted category of those examples. Perfect recognition will yield a confusion matrix with zero values in the off-diagonal elements. A normalized confusion-matrix represents the percentage of, or the probability that a particular class will be predicted as one of the classes. The normalized confusion-matrix is calculated by dividing each element of the row by the sum of that row in the confusion matrix. Worst performance is considered the performance that can be achieved by random guessing, which generates a normalized confusion matrix with 33% classification rate in each matrix element.

B. Validation of Performance

During the training of any pattern classifier, there is a risk of over-fitting, where the classifier is trained to the point when the training set is classified perfectly, but when presented with novel examples the correct classification rate is worse than can otherwise be achieved. This is due to noise associated with overlapping class regions in feature-space. A classifier that has been over-fitted would create decision boundaries to classify these noisy samples correctly and thereby degrade classification performance for non-noisy samples.

To address this, 5-fold cross-validation is used to estimate the expected recognition rate, a better indicator to the generalizability of the pattern classifier to as-yet-unseen data. That is to say in order to generate a performance measure that represents more closely the true classification performance on as-yet-unseen data, we use cross-validation to estimate the correct classification rates. The data-set was divided into 5 sub-sets. A multi-class SVM classifier was trained on all but 1 of the 5 sub-sets of examples, and the classification rates are calculated from the remaining sub-set to produce 5 normalized confusion matrices. The average of all 5 recognition rates are found and reported. The standard deviation of each element in the confusion matrix is also found and were always less than $.5\%$ difference.

To ensure that the trained results would perform well in real situations, the data-set was collected at various times-

TABLE II

ATTRIBUTES SUMMARY OF THE 18 IMAGE SEQUENCES USED FOR TRAINING AND TESTING THE VUD CLASSIFIER. THERE ARE 8 UNIQUE SUBJECTS AND 1 HOUR OF 30FPS VIDEO.

	Sunny	Overcast	Night	Indoor
Veh. Stationary		11,036		10,814
Veh. Moving	46,209	11,740	35,087	
Subjects	5	1	4	1

TABLE III

SUMMARY OF USERS.

Class	Description	Examples
1	No One	68,467
2	Driver	20,179
3	Passenger	25,340

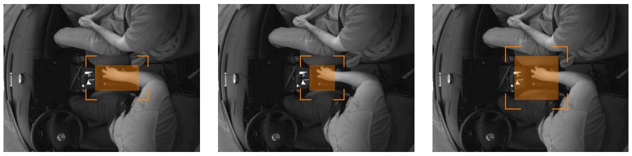
of-day (noon, afternoon, twilight, night) with various individuals (8 male individuals of average build, 5'0" to 6'0" in height) in both the driver and passenger position. One sequence was captured with a variety of clutter (flashlight, card-board, paper, mouse-pad, tools, cups) introduced into the region-of-interest to capture the statistics of feature vectors of those instances. A total of 18 video sequences containing a 114,886 examples and 63 minutes of video in various illumination conditions were used for training and testing. The conditions under which the data-set was collected is summarized in tab II and III. Individuals wore short and long-sleeve shirts. For most sequences, the data was captured while the vehicle was in motion, driven along a circular route in which the direction of sunlight could shine into the vehicle from every direction at least once. Different times of day yielded different angles of the sun and character of the sun.

Each frame of the video is manually annotated with the category to which it belongs. Namely, each frame may show either no one, the driver, or the passenger is placing their hand over the infotainment controls area. There are a total of 68,467, 20,179 and 25,340 unique frames collected for each of the three classes, respectively.

C. System Parameters Optimization

Three feature types were analyzed to validate our choice of image patch dimensions. Intuitively, the forearm is a good indicator of whose hand is accessing the infotainment controls. A rectangular image ROI of size 140x80 as depicted in Fig. 3(a) appears to capture both the hand and the forearm compactly. The other two image patches consisted of a square image patch of sizes 80x80 and 140x140, centered around the hand as shown in Fig. 3(b), and 3(c).

The multi-class SVM classifier with the RBF kernel has 2 parameters that require tuning: the slack parameter C, and the RBF kernel width γ. This is done by searching a grid of values for the optimal tuple (C, γ) that yields the highest average correct detection rate (mean of the diagonal elements of the normalized confusion matrix). A subset of the complete data-set was used to efficiently generate the values in the grid-search: 5000 examples of each class were randomly selected for a total of 15,000 examples in the new training set. The results of the grid-search for all three feature-types indicate that the optimal values are $C = 25$ and $\gamma = 100$.



(a) Exp 1 (b) Exp 2 (c) Exp 3

Fig. 3. Various image patch sizes were used in evaluating the VUD system.

The results of the parameter optimized kernel SVM classifiers are summarized in Tab. IV. The differences in percentage points were subtle, but the rectangular image patch produced the best results of the three with 96.79%. Each percentage in the matrix represents the proportion of each case predicted as any of the 3 cases. To give a better sense of the performance as a function of time, the duration of time in error was calculated. The amount of time when the system was in error in one minute is calculated by average percentage of time in error multiplied by 60 seconds. The average number of seconds in error in one minute for the three types of features were 1.926, 2.544, 2.526 seconds respectively.

V. CONCLUDING REMARKS

We presented a real-time vision-based user determination system. The system consisted of a visible and near-infrared imaging device observing the front-row seat area in the vehicle. Using histogram of oriented gradients image descriptor to describe the area over the controls, a support-vector-machine was shown to be able to provide 96.8% average correct classification rate over the three classes: driver, passenger or no-one is accessing the infotainment controls.

The system is intended to improve the safety and comfort of the vehicle by enabling the vehicle to determine which occupant is accessing the vehicle's infotainment controls, often characterized as one of the more distracting elements in a vehicle. It is a safety device in the sense that the vehicle would know whether there is a potential of the driver to be distracted in a critical situation, taking one more step towards a fully automatic driver's situational awareness estimation system [14], [15]. It is a comfort device in the sense the passenger can still be allowed to access the infotainment controls to aide the driver in navigational and convenience needs.

A. Future Work

In consideration of future work, upgrading the algorithm to affine-invariance can be investigated further. To increase the affine-invariance of the image descriptors, the image ROI can be repositioned (re-calibrated) when the controls are visible

TABLE IV

SUMMARY OF VUD PERFORMANCE. THE PERFORMANCE IS DESCRIBED USING THE CONFUSION MATRIX AND AVERAGE PROPORTION OF 1 MINUTE IN ERROR. THE CLASSIFIERS ARE MULTI-CLASS KERNEL SVMS (C=25, RBF, GAMMA=100) AND THE CONFUSION MATRICES ARE AVERAGES OF 5 TRAINING RESULTS VIA 5-FOLD CROSS VALIDATION. 5000 RANDOMLY SELECTED EXAMPLES PER CLASS WAS SELECTED FOR EACH TRAINING.

(a) Exp 1

	P(predicted|actual)			
	NoOne	Drver	Psngr	
NoOne	95.42	2.01	2.57	
Drver	1.82	97.63	0.54	
Psngr	2.12	0.57	97.31	
Average Correct Classification Rate				96.79%
Average Seconds per Minute in Error				1.926 sec

(b) Exp 2

	P(predicted|actual)			
	NoOne	Drver	Psngr	
NoOne	94.98	2.44	2.58	
Drver	2.66	96.41	0.93	
Psngr	3.02	1.08	95.90	
Average Correct Classification Rate				95.76%
Average Seconds per Minute in Error				2.544 sec

(c) Exp 3

	P(predicted|actual)			
	NoOne	Drver	Psngr	
NoOne	93.73	2.77	3.49	
Drver	2.40	97.18	0.42	
Psngr	2.94	0.59	96.47	
Average Correct Classification Rate				95.79%
Average Seconds per Minute in Error				2.526 sec

on a regular basis to correct for any vibrations of the camera over time. A scheme as simple as template matching of the gradient images with the stored image may be used to align the ROI to the location that produces the best classification rates.

Although data collection was carefully performed to ensure obtaining a representative training sample set, there are additional variations to be considered. Namely, all the test subjects were adults and no children nor 5-th or 95-th percentile occupants were asked to be subjects. Although one subject had very long sleeves, covering half of his hand, gloves or jewelry were not used during the data capture. The performance is not expected to decrease by much with these variations, but having a representative data-set is critical to ensure that the measured performance is the performance of a deployed system. We reserve this task for future work as well.

Finally, one aspect of the VUD that was not studied in detail is the ability of the system to correctly determine the user when both the passenger's and driver's hands are in the ROI. Anecdotally, the system was observed to be able to determine the correct user even after the other occupant's hand occluded nearly 50% of the ROI, either reaching for the dome light or radio controls. This indicates that the feature descriptor is probably rich enough to classify between the three cases in these situations, and the (future) task is to collect and train on sufficient instances where occlusion occurs.

ACKNOWLEDGMENT

The authors would like acknowledge the sponsorship of this project by Volkswagen of America, Electronics Research Lab, Palo Alto CA and by UC Discovery, Digital Media Innovations program. We would also like to thank the kind assistance from members of the Computer Vision and Robotics Research Laboratory for their hand in data collection.

REFERENCES

[1] J. D. Lee, "Technology and teen drivers," *Journal of Safety Research*, vol. 38, pp. 203–213, 2007.

[2] M. Jerbi, S.-M. Senouci, R. Meraihi, and Y. Ghamri-Doudane, "An improved vehicular ad hoc routing protocol for city environments," in *IEEE Int'l Conf. on Communications*, Jun. 2007, pp. 3972–3979.

[3] E. Murphy-Chutorian, A. Doshi, and M. M. Trivedi, "Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation," in *IEEE International Conference on Intelligent Transportation Systems 2007*, September 2007.

[4] R. Duda, P. Hart, and E. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[5] C. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. Springer Sciences and Business Media, LLC, 2006.

[6] P. B.-L. Chou, J. Lai, A. Levas, and P. A. Moskowitz, "System for controlling vehicle information user interfaces," *United States Patent*, 2001, 6,181,996.

[7] J. E. Harter, Jr., G. K. Scharenbroch, W. W. Fultz, D. P. Griffin, and G. J. Witt, "User discrimination control of vehicle infotainment system," *United States Patent*, 2003, 6,668,221.

[8] M. Kölsch and M. Turk, "Analysis of rotational robustness of hand detection with viola & jones' method," in *IAPR International Conference on Pattern Recognition*, 2004.

[9] S. Y. Cheng, S. Park, and M. M. Trivedi, "Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis," *Computer Vision and Image Understanding*, vol. 106, no. 2–3, pp. 245–257, May–Jun. 2007, doi: 10.1016/j.cviu.2006.08.010.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[11] C. Burges, "A tutorial on support vector machines for pattern recognition," *Knowledge Discovery and Data Mining*, vol. 2, no. 2, 1998.

[12] C.-C. Chang and C.-J. Lin, "A library for support vector machines," website, 2007.

[13] IntelCorporation, "Open Computer Vision Library," http://sourceforge.net/projects/opencvlibrary, 2007.

[14] M. M. Trivedi and S. Y. Cheng, "Holistic sensing and active displays for intelligent driver support systems," *IEEE Computer Magazine*, pp. 60–68, May 2007.

[15] S. Y. Cheng and M. M. Trivedi, "Turn-intent analysis using body pose for intelligent driver assistance," *IEEE Pervasive Computing Magazine*, vol. 5, no. 4, pp. 28–37, 2006.