2004 IEEE Intelligent Transportation Systems Conference
Washington, D.C., USA, October 3-6, 2004

Face Detection and Head Tracking using Stereo and Thermal Infrared Cameras for "Smart" Airbags: A Comparative Analysis

Stephen J. Krotosky, Shinko Y. Cheng, *Student Member, IEEE* and Mohan M. Trivedi, *Senior Member, IEEE*

***Abstract*—Vision based systems for "smart" airbag systems aim to give precise information about occupant pose and location. This information can be used to make intelligent airbag deployment decisions. This paper reviews both a stereo-based and long wave infrared-based system for "smart" airbag deployment. The algorithms are systematically evaluated through extensive real-time occupant tests. Data for both algorithms is simultaneously collected and evaluated for viability in an intelligent airbag system. Results of these experimental trials show the feasibility of each video based occupant position analysis system. Advantages and drawbacks of each method are discussed and potential solutions are suggested.**

I. INTRODUCTION

AIRBAG deployments during automobile accidents can cause injury and even death if the occupant is of an improper size or in an improper position. Because of this, steps have been taken to make airbag deployment safer by adding adaptive deployment decision capabilities. One method of determining these deployment decisions is by using a vision system to detect and track occupant position, namely head position, in the automobile cabin. Depending on the occupant's position, the airbag can be fully deployed, partially deployed or not deployed at all.

It is the objective of this research to offer a comparative analysis of two vision methods that detect and track the occupant head. Specifically, a stereo-based and long wave infrared-based method will be tested through an extensive series of experiments to determine each method's viability for use in an intelligent airbag deployment system.

Manuscript received April 1, 2004. This work was supported in part by a University of California Discovery Grant with Volkswagen Research Laboratory under the Digital Media Innovation Program.

S. J. Krotosky, S.Y. Cheng and M. M. Trivedi are with the Computer Vision and Robotics Research Laboratory at the University of California, San Diego, La Jolla, CA 92093 USA (858-822-0002; fax: 858-822-5336; e-mail: krotosky@ucsd.edu, sycheng@ucsd.edu, trivedi@ucsd.edu).

II. RESEARCH OBJECTIVES AND APPROACH

The ultimate goal of this research is the development of a robust real-time vision system for sensing occupant body posture in vehicles and providing safe airbag deployment. The head location is the target of our initial focus, as it provides a good estimate of the occupant position and is the most critical part of the body in airbag deployment.

To determine whether a person is in the right position for airbag deployment, the area between the seat and dashboard can be divided into three sections: In-Position, Out-of-Position, and Critically-Out-of-Position. A diagram detailing these divisions is shown in Fig. 1.

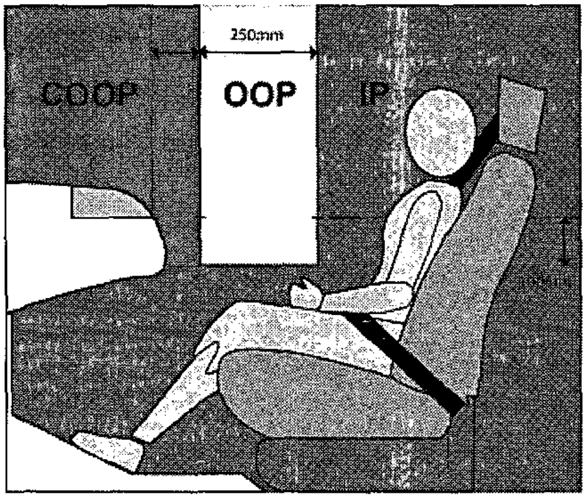

Fig. 1. Schematic showing the In-Position (IP), Out-of-Position (OOP), and Critically Out-of-Position (COOP) regions of the passenger seat.

These positions, and corresponding airbag deployment decisions, can be equated to occupant head location in many seating scenarios, including sitting normally, leaning forward, reaching down, sitting on the edge of the seat, and sitting reclined. In the cases where head location is not enough to make the correct decision, such as the occupant sitting with his feet on the dashboard, we believe that the head location still gives important information that can used

0-7803-8500-4/04/$20.00 ©2004 IEEE

to extend the body model in future research. Examples of occupant positions considered in development of these intelligent airbag systems are displayed in Fig. 2.

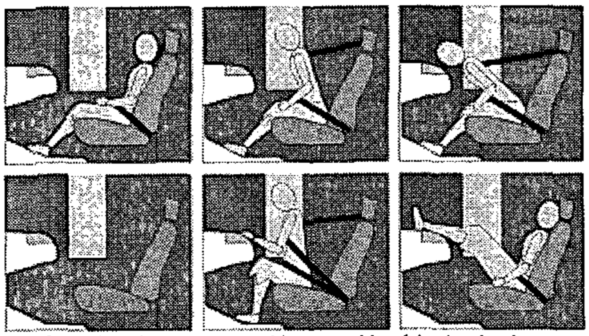

Fig. 2. Selected occupant positions considered in the development of "Smart Airbag" Systems. Head location alone can make the correct decision in the first five examples. In the last example, the occupant's legs present a potential danger despite the head being located in a safe position. In this case, further body modeling is necessary.

III. Related Studies

In previous experiments [1], active illumination, such as near-infrared LEDs, is used to capture the scene features. While active illumination can reduce camera illumination sensitivities, we show that these extra lighting schemes are superfluous for estimation of occupant posture.

Stereo-based methods have been previously investigated in [2], [3], [4], [5] and [6]. However, these methods suffer from either having limited occupancy classification [3] or requiring an extensive training set of data for proper operation [4], [5]. Our method, however, requires no such training and can provide more detailed occupant information than previous methods. And unlike [6], which relies on the occupant's surface and dense stereo reconstruction, our effort relies on the data fitting a body model that we can track frame to frame. Similar to [2], our method relies on more than just stereo disparities. By using the raw reflectance and stereo disparities, our method can elicit reliable occupant posture and pose estimation. This research also extends the methods discussed in [7] by providing more accurate stereo disparities, more varied head detection templates and decreased illumination sensitivity.

Thermal infrared-based methods have been investigated in [8]. This paper however, extends the investigation in [8] by moving the experiments outside of a controlled indoor laboratory environment and into the uncontrolled climate of a moving vehicle. The experiments presented in this paper are also more extensive in terms of the volume of frames analyzed as well as the complexity of the subject's movement throughout the experiment.

This research is also novel in that it offers a direct comparison of two competing face detection and head tracking algorithms. The extensive test bed developed in [12] allows for a frame-by-frame analysis of each algorithm, as data is captured for the stereo and thermal IR-based methods simultaneously. This unique setup will also allow for similar comparable testing and analysis of newer capture modalities and algorithms as they are developed.

IV. Face Detection and Head Tracking Algorithms

The Stereo and LWIR based face detection algorithms are both derived from Eleftheriadis, et al [9]. An edge-based face detection algorithm is described that provides head pose and size estimates that can be implemented in real time. The basic edge-based algorithm was first detailed in [12] and its specific stereoand LWIR-based instantiations are summarized below.

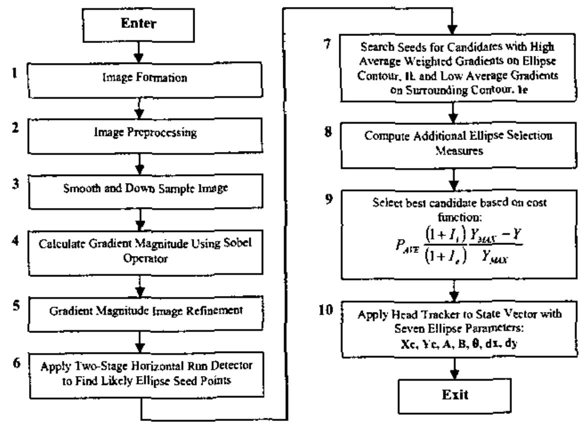

Fig. 3. Flowchart for the basic edge-based face detection algorithm. Details of steps 1, 2, 5, and 8 are specific to the capture modality.

A. Stereo-based Face Detection and Head Tracking

The stereo-based face detection and head tracking algorithm [11] is a modification of the edge-based face detection algorithm displayed in Fig. 3. A background model is first obtained using N frames of disparity data of an empty cabin. An example of the reflectance image and its corresponding disparity image is shown in Fig. 4. After the background image is computed, the current foreground data is generated. A threshold is applied to eliminate noisy pixel stemming from invalid stereo regions. Depths that fall outside the car are also removed. Simple background subtraction is applied and a median filter and morphological opening are performed to eliminate noise. Connected component analysis is also performed to remove areas smaller than a pre-defined minimum head size. This image is the current foreground disparity image.

Given the current foreground image and the raw reflectance image, the best head location can now be located. First edges are found in the reflectance image using Sobel operators. Edges that are contained in the current foreground disparity image are retained. Given the pre-computed set of ellipse templates of various sizes, angles and eccentricities the best fit ellipse/head position is found by maximizing the cost function in step 9 of Fig. 3. P_{AVE} is the average depth in the bounding box enclosing the ellipse candidate. Ii is the value on the interior contour. Ie is the value along the exterior contour and Y is the vertical position in the image. A best ellipse candidate will have a high value Ii, a low value Ie, a high value P_{AVE} and a low value of Y.

Fig. 4. Example of reflectance and corresponding disparity images for a subject in the LISA-P test bed

B. *Thermal IR-based Face Detection and Head Tracking*

This algorithm is based on the edge-based face detection algorithm displayed in Fig. 3. The input image is first remapped from LWIR to a probability of human skin temperature based on [10] using a simple Gaussian PDF, with the mean and variance manually, empirically set. An example of this remapping is shown in Fig. 5. An iteration of grayscale erosion is then performed on the image to eliminate noise.

Using this image, the best-fit ellipse/head location is determined similarly to the stereo-based method. Edges are found in the image using Sobel operators. Given the pre-computed set of ellipse templates of various sizes, angles and eccentricities the best fit ellipse/head position is found by maximizing the cost function in step 9 of Fig. 3. P_{AVE} is the average probability of skin temperature in the bounding box enclosing the ellipse candidate. Ii is the value on the interior contour. Ie is the value along the exterior contour and Y is the vertical position in the image. A best ellipse candidate will have a high value Ii, a low value Ie, a high value P_{AVE} and a low value of Y.

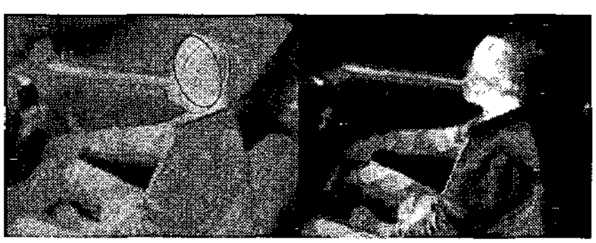

Fig. 5. Example of LWIR image and corresponding remapped to probability of skin temperature image.

V. Stereo vs. LWIR Head Detection and Tracking Comparison

A. *Experimental Setup*

Experiments were performed using the LISA-P test bed described in [12]. The LISA-P is a Volkswagen Passat, shown in Fig. 6, equipped with various cameras that can be synchronously captured and saved for data analysis. A series of experiments were conducted where data was captured for both the stereo and IR methods simultaneously. The desire is to have a direct comparison of the two head tracking methods on a frame-by-frame basis.

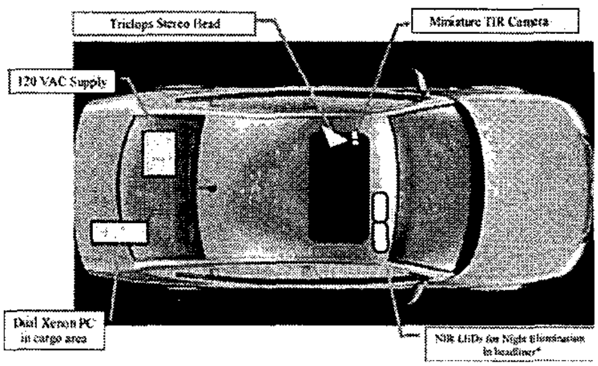

Fig. 6. Multimodal video cameras, synchronized capture, computing, storage, and power modules in the LISA-P instrumented vehicle testbed. Items with asterisks are proposed but not yet installed.

The tests were conducted on three non-consecutive mornings in sunny weather from 8-11 am using the LISA-P test bed. A Point Grey Digiclops stereo camera with a 10cm baseline [13] was placed on the driver's side roof rack, looking down on the passenger's seating area. This position was chosen to allow the passenger positions to lie within disparity regions that are greater than the minimum calculable depth (~3 feet) given the stereo baseline. The mounting of the camera was such that driver occupancy does not introduce occlusion into the image frames. An overhead camera mount is chosen as to minimize the potential occlusions of the head both by the driver, the passenger, and any objects the passenger may carry into the car.

Stereo pairs were captured at 320x240 resolution at 15fps. Stereo data was computed using SRI's Small Vision System API [14]. For these tests, the IR camera by Raytheon [15] was placed next to the stereo camera in the same orientation, so that approximate depth from the dashboard could be inferred. The long-wave infrared camera has a spectral response of 7-14 μm. Infrared images were captured simultaneous to the stereo images at a resolution of 640x480.

TABLE I

STEREO AND THERMAL IR FACE DETECTION AND HEAD TRACKING COMPARISON

Occupant Task	Male 1, 5'8"		Female 1, 5'8"		Female 2, 5'11"		All Occupants	
	Stereo	IR	Stereo	IR	Stereo	IR	Stereo	IR
Sit Normal	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Lean Halfway	100.0%	73.0%	100.0%	92.9%	X	X	100.0%	82.8%
Lean Forward	76.4%	0.9%	X	X	X	X	76.4%	0.9%
Return to Normal 1	100.0%	95.9%	98.0%	98.0%	100.0%	100.0%	99.6%	97.4%
Lean Back	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Return to Normal 2	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Lean Right	100.0%	52.1%	100.0%	100.0%	97.8%	96.7%	99.1%	92.1%
Lean Left	100.0%	98.9%	X	X	97.7%	100.0%	98.4%	99.7%
Return to Normal 3	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Position Test Totals	97.3%	80.3%	99.8%	98.7%	98.7%	99.1%	98.4%	91.7%
(Number of Frames)	(940)	(776)	(537)	(531)	(676)	(679)	(2153)	(1986)
Move Hands about cabin	78.1%	100.0%	100.0%	97.4%	97.8%	99.1%	91.6%	99.2%
Open the glove box	100.0%	100.0%	100.0%	95.5%	74.3%	97.6%	91.2%	97.8%
Put hands on face & stretch	81.7%	100.0%	100.0%	85.2%	87.8%	89.4%	90.0%	91.3%
Adjust car radio	100.0%	100.0%	100.0%	100.0%	99.4%	100.0%	99.8%	100.0%
Place hat in lap	100.0%	100.0%	100.0%	97.5%	100.0%	97.7%	100.0%	97.9%
Put hat on head	90.0%	84.3%	90.5%	35.7%	100.0%	93.3%	95.2%	85.2%
Move with hat	98.8%	87.9%	100.0%	68.3%	92.6%	62.8%	96.5%	71.0%
Remove Hat	100.0%	100.0%	100.0%	62.1%	100.0%	100.0%	100.0%	94.9%
Feet on Dashboard	100.0%	94.5%	100.0%	76.4%	93.9%	100.0%	98.3%	87.3%
Hand Motion & Object Test Totals	92.6%	97.4%	99.8%	85.7%	92.0%	90.5%	94.8%	90.9%
(Number of Frames)	(1399)	(1471)	(1939)	(1665)	(2258)	(2221)	(5596)	(5357)
Free Motion Test	100.0%	87.4%	99.8%	95.5%	95.8%	86.1%	97.9%	88.9%
(Number of Frames)	(493)	(431)	(470)	(450)	(942)	(846)	(1905)	(1727)
All Test Totals	95.4%	90.2%	99.8%	89.6%	94.0%	90.9%	96.2%	90.3%
(Number of Frames)	(2832)	(2678)	(2946)	(2646)	(3876)	(3746)	(9654)	(9070)

X denotes that the subject moved out of the camera frame for this test, and the results were invalid.

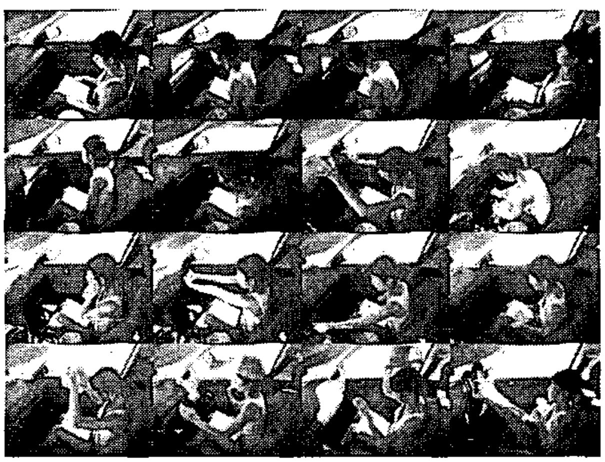

Fig. 7. Example images of occupant script poses. From top left: sitting normally, leaning halfway, leaning completely forward, leaning back, leaning right, leaning left, moving hands about cabin, opening glove box, hands on face, stretching, adjusting radio, hat in lap, putting on hat, moving while wearing hat, removing hat, feet on dashboard.

Occupants were asked to perform a scripted series of movements and tasks after entering the car. The occupant script was divided into three tests. First, a Position Test was conducted to tests each algorithm's ability to detect the face at various positions in the cabin and track the head's movement. Next, a Hand Motion and Object Test was designed to evaluate each algorithm's robustness to competing objects and hand motion in the scene. Finally, the subject was asked to move as they wish for a Free Motion Test, designed to catch other potential situations for detection error. Throughout the duration of the tests, the car is driven at road speeds. Example images of each of the occupant script poses are shown in Fig. 7.

Stereo and LWIR data was collected with three subjects for a total of 10,045 frames. Head Detection was considered successful when the ellipse center was placed somewhere on the occupant's head and the selected ellipse size was comparable to the occupant's face size. 80 ellipse templates were used ranging in size from 24-36 pixels, in eccentricity from 1.1-1.6 and in angle from 60-120 degrees. The ellipse templates operated on downsampled 60x80 pixel images. Table I lists the detection rates for the Stereo and LWIR head detection and tracking algorithms.

B. Results and Discussion

This test set up is particularly unique in that it allows for a direct comparison of the stereo and LWIR head detection methods on a frame-by-frame basis. Both algorithms achieve a high average accuracy in detecting and tracking the head. At success rates of 96.4% and 90.1% respectively for various occupant types, it can be concluded that both the stereo and LWIR systems can be used to robustly detect the location of the head with high reliability. The resulting head location can give the necessary information to decide the manner in which an airbag should be deployed. Examples of successful head detection in both algorithms are shown in Fig. 8. Although the algorithms achieve a high success rate, both suffer from certain drawbacks described below that should be resolved through further research.

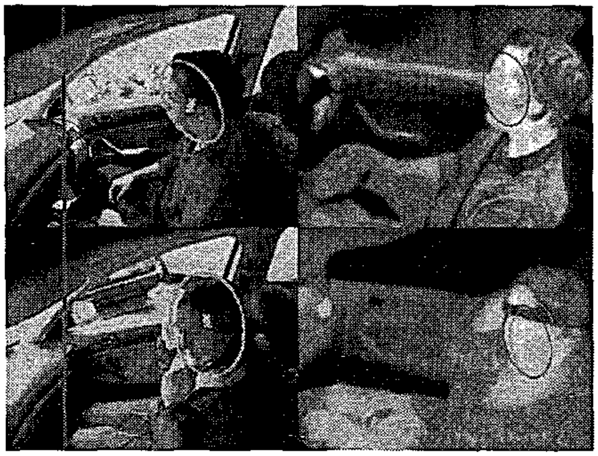

Fig. 8. Example images of successful head detection for both the stereo and LWIR based algorithms. The top pair shows successful detection for sitting normally, and the bottom pair shows successful detection when putting on a hat.

Stereo methods outperformed IR methods when the subject was wearing a hat. Because the head detection algorithm is searching for elliptical objects, the detection rate decreases in the IR case because the hat changes the emissive properties of the subject's head, making it appear less elliptical. Conversely, in the stereo case, we compute edges based on the reflectance data, where the occupant head appears similarly elliptical with or without a hat. This was especially true with the Female 1 occupant. An example of this error is shown in Fig. 9.

Fig. 9. The left image shows successful detection using the stereo-based head detection algorithm. The right image shows the IR-based algorithm's inability to detect a head location because too much of the face is obscured from the camera because the occupant is wearing a hat. This is a drawback of using skin temperature as a cue for detecting the head. If there is not enough skin in view, the head location will not be found.

Stereo methods also outperformed IR methods when the subject leans completely forward. This, however, is not a function of the occupant's position, but rather due to the subject's head being turned from the camera so that only the back of his head was visible. Naturally, the occupant's hair does not have the same temperature as the face, and resulted in the low detection rates. It is impossible to cue on skin temperature when the occupant is looking away from the camera. This was also the case for the low results for the lean right test for Male 1. This indicates that modeling only skin temperature may not be enough, and other occupant properties, such as hair and clothing should be taken into account. An example of this type of error is shown in Fig. 10.

Stereo methods rely on background modeling. Currently simple background subtraction is used and does not accommodate dynamic background movements, such as seat adjustments. More sophisticated methods of background modeling need to be implemented to account for such movements in the background. It remains untested, however, how much effect a seat movement would have on the head detection results using the current algorithm. One method of background modeling would be to have a database of stored background disparities for each seat position. Sensors that give the current seat orientation can be used to select the appropriate reference background disparity image for the head detection algorithm.

Fig. 10. The left image shows successful detection using the stereo-based head detection algorithm. The right image shows the IR-based algorithm's inability to detect a head location because too much of the face is obscured from the camera because the occupant is turned away from the camera. This is a drawback of using skin temperature as a cue for detecting the head. If there is not enough skin in view, the head location will not be found.

IR methods outperform stereo methods when dealing with competing elliptical objects in the scene, especially hands. This is because the skin temperature at the hands is usually different enough from the face as not to confuse the IR detector. However, if the hands are of a similar size and depth as the head, the stereo detector can give erroneous results. Potential solutions include using sub-facial cues to verify ellipse candidates, as well as introducing more sophisticated body modeling to account for hands and arms in the scene. An example of this error is shown in Fig. 11.

In general, IR would be more robust to changes in lighting conditions than the stereo-based method. The LWIR spectrum is not especially sensitive to visible light and could easily operate in day or night conditions. A similar robustness could be obtained in the stereo-based method, but would require near-IR illuminators and filters for the camera lenses.

Despite the success of these initial tests, further testing is imperative. This test of the stereo and LWIR systems included only three subjects at a particular time of day in fair weather. Clearly, many different subjects need to be tested on the system. It is still untested how well the algorithms will perform when subject's have features such as facial hair, large hats, are eating or drinking, are very large or very small, or are sitting in unconventional positions. The algorithms are also untested in driving conditions other than a sunny day. Although an exhaustive test of the permutations of subject type and driving condition is impractical, a much larger and extensive test of these variations is necessary to deem the algorithms reliable enough for commercial use.

Fig. 11. The left image shows unsuccessful detection using the stereo-based head detection algorithm because the occupant arm configuration confused the ellipse-based head detector. The right image shows the IR-based algorithm's ability to detect a head location because the skin temperature segmentation accentuates the face location more than the arm location.

VI. Conclusion

As intelligent airbag systems continue to develop, it is important to offer a method of comparison between algorithms in order to determine viability under various conditions. This paper presented the systematic evaluation of a stereo-based and LWIR-based face detection and head tracking algorithm that can be implemented in a "smart" airbag system. The work has shown that both methods can detect occupant head location with a high reliability. Advantages and drawbacks of each method were discussed and potential solutions to each algorithm's flaws were presented.

Acknowledgment

The authors gratefully acknowledge valuable inputs and assistance by Dr. Klaus Schaff and Dr. Arne Stoschek of VW Research Laboratory. We extend our thanks to Mr. Edwin Childers for his assistance in the development of our LISA-P test bed and initial efforts with the thermal IR cameras. We are also thankful for the assistance and support of our colleagues from the UCSD Computer Vision and Robotics Research Laboratory, especially Tasha Vanesian and Junwen Wu. We would also like to thank Leslie Wood for her valuable assistance in our experiments.

References

[1] C. Koch, T.J. Ellis and A. Georgiadis, "Real-time Occupant classification in High Dynamic Range Environments", *IEEE Intelligent Vehicle Symposium*, Vol. 2, pp. 284 -291, 2002.

[2] Y. Owechko, et. al., "Vision-Based Fusion System for Smart Airbag Applications", *IEEE Intelligent Vehicle Symposium*, Vol. 1, pp. 245 -250, 2002.

[3] P. Faber, "Seat Occupation Detection Inside Vehicles. Image Analysis and Interpretation", *Proc. 4th IEEE Southwest Symposium*, 2000.

[4] J. Krumm, G. Kirk, "Video Occupant Detection for Airbag Deployment. Applications of Computer Vision", *WACV '98*, 1998.

[5] R. Reyna, A. Giralt, D. Esteve, "Head Detection Inside Vehicles with a Modified SVM for Safer Airbags", *Proc. IEEE Intelligent Transportation Systems Conference*, 2001.

[6] M. Devy, A. Giralt, A. Marin-Hernandez, "Detection and Classification of Passenger Seat Occupancy Using Stereovision", *IEEE Intelligent Vehicles Symposium*, 2000.

[7] T. Schoenmackers, M.M. Trivedi. "Real-Time Stereo-Based Vehicle Occupant Posture Determination for Intelligent Airbag Deployment." *IEEE Intelligent Vehicle Symposoium*, 2003

[8] C. Eveland, D. Socolinsky, and L. Wolff, "Tracking human faces in infrared," CVPR Workshop on Computer Vision beyond the Visible Spectrum, Kauai, Dec. 2001.

[9] A. Eleftheriadis and A. Jacquin, "Face location detection for model-assisted rate control in H.261-compatible coding of video," *Signal Processing: Image Communication*, vol.7, no.4-6, pp.435-55, Nov. 1995.

[10] D. A. Socolinsky, L. B. Wolff, J. D. Neuheisel, and C. K. Eveland, "Illumination invariant face recognition using thermal infrared imagery" in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2001.

[11] S. J. Krotosky and M. M. Trivedi, "Occupant posture analysis using reflectance and stereo images for 'smart' airbag deployment", in *Proc. IEEE Intelligent Vehicles Symposium*, June 2004.

[12] M.M. Trivedi, S.Y. Cheng, E. M. C. Childers, and S. J. Krotosky, "Occupant Posture Analysis with Stereo and Thermal Infrared Video: Algorithms and Experimental Evaluation," *IEEE Trans. Veh. Technol. Special Issue on In-Vehicle Vision Systems*, To Appear 2004.

[13] Point Grey Research, "Digiclops Stereo Vision Camera", www.ptgrey.com.

[14] Videre Design, "SRI Small Vision Systems", www.videredesign.com.

[15] Raytheon, "Thermal-Eye 2000AS Camera", www.raytheoninfrared.com.