Real-Time Stereo-Based Vehicle Occupant Posture Determination for Intelligent Airbag Deployment

Timothy D. Schoenmackers and Mohan M. Trivedi
Department of Electrical And Computer Engineering
Computer Vision and Robotics Research Laboratory
University of California at San Diego
La Jolla, CA, USA
tschoenm@ucsd.edu, trivedi@ece.ucsd.edu

Abstract:

From 1990 to April 2002, according to the National Highway Traffic Safety Administration (NHTSA) airbag deployment has been responsible for 208 deaths in low-severity crashes. Of these deaths, 107 were children between the ages of 1 and 11, 22 were infants, and 32 were women under the height of five feet two inches. In addition to deaths, airbags are responsible for many more incidental injuries and broken bones. These injuries could be avoided if the system had an accurate model of the actual pose of the occupant. This paper discusses a system we are developing to accurately determine the position of the head, torso, and arms of the front-seat occupant of a vehicle.

Introduction:

It has been shown that vision systems and principal component analysis can be used to determine the presence or absence of a human occupant, as well as determine if the occupant is an infant [1,2]. Other research has focused on determining the seat position first, then minimizing an energy function to determine whether the passenger occupied a zone too near the airbag [3]. Currently in the automotive industry, vehicles are equipped with varying degrees of intelligent airbag deployment schemes.

The simplest systems take into account whether the seatbelt has been engaged and use this information to determine the strength at which to deploy the airbag [4]. Some of the most complicated systems take into account the weight, seatbelt information, and data from bend-sensors in the seat in order to determine whether the seat occupant is an adult, a child, or a bag of groceries [5,6]. Our research focuses on using a real-time stereo camera to acquire a depth map of the interior of a vehicle.

Stereo algorithms are well suited to in-vehicle situations for numerous reasons. Stereo depth maps are less sensitive to illumination changes than intensity images [1]. Secondly, because stereo provides information about the depth of the associated scene, points that lie outside the vehicle can be filtered out easily. From the depth information the passenger head can then be localized, and tracked from frame to frame using the relatively simple and fast techniques described below. Because airbags deploy in approximately 30mS after the time of impact, an intelligent airbag system must be able to make a decision about the pose of the driver at these rates.

System Flow and Algorithmic Details:

In order to determine the posture of the driver in a car we have developed an algorithm to find and track the head of the driver. From the head, the approximate position of the torso can be determined by projecting a line down from the centroid of the head to the waist, which is assumed to be positioned at the intersection of the seat bottom and seat back. Currently this intersection is manually initialized, however we hope to automate this process in the future. Figure 3 shows the overall flow of the algorithmic modules.

Figure 1: Subject in test vehicle and corresponding depth map

Figure 2: Subject sitting in Infinity and corresponding depth image.

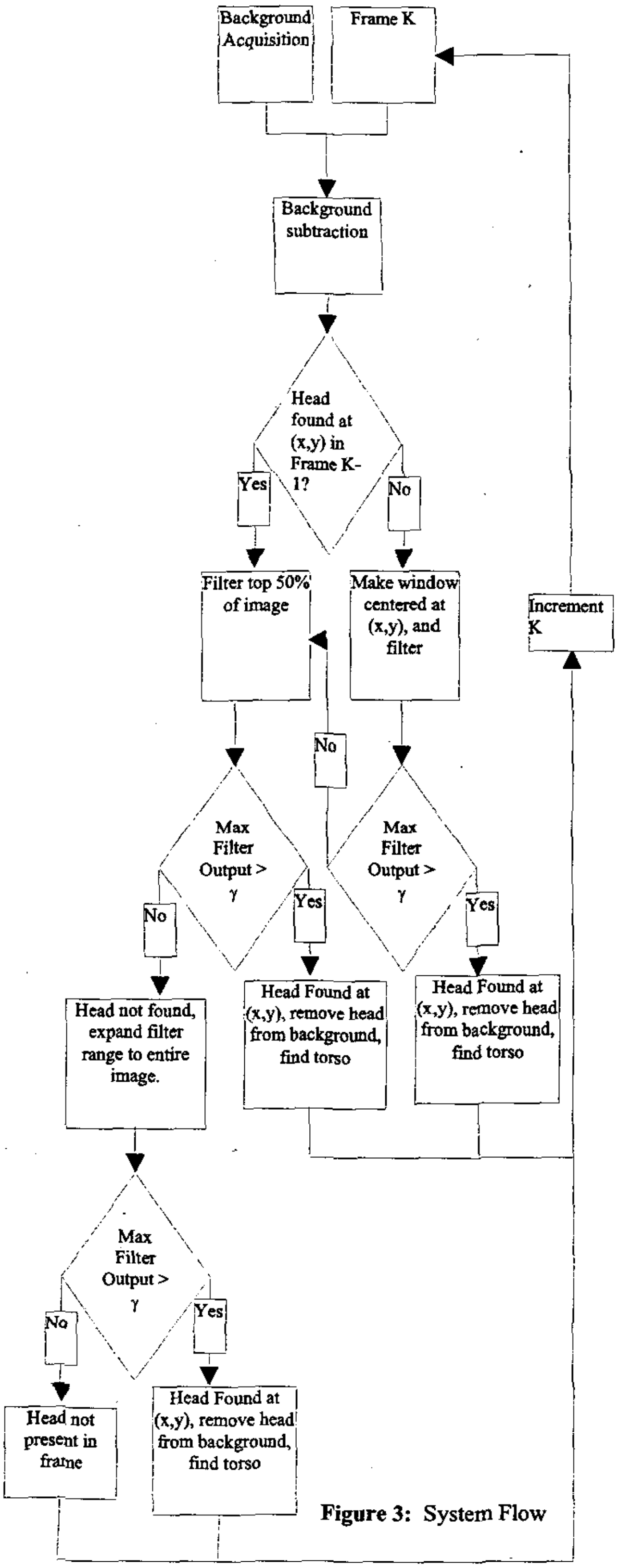


Figure 3: System Flow

Once the stereo camera grabs an image, the corresponding disparity image is computed. Using the disparity map, the system then filters out any points that fall outside the car. Once the in-car points have been determined, a background model of the car is determined by averaging the first 10 frames of the disparity image. This assumes that the occupant has not yet been seated in the vehicle for the first ten frames. Once the background model is calculated, it can then be used to subtract from new images in order to segment an occupant of the seat. After the first 10 frames, the background is adaptively updated every frame according to the formula:

$$B_{K+1} = \alpha B_K + (1-\alpha)F$$

where B_{K+1} is the new background, B_K is the old background, F is the current frame, and α is a scalar greater than 0 and less than 1 that controls the rate at which new frames are incorporated into the background. After the initial frames, as a new frame is acquired the background is subtracted, and the resulting image is converted into a binary image. To ensure that a still passenger is not incorporated into the background, once the body of the passenger is found, that portion of the image is subtracted back out of the background. We then apply the simple zero-mean oval filter shown in Figure 4.

(a) (b) (c)

Figure 4: (a) Filter for head finding. (b and c) Three-dimensional reconstruction of driver

The filter is 20pixels wide, and 30 pixels long. The gray portion of the filter represents negative values, while the white portion of the filter represents positive values. After filtering the image, the algorithm looks for the maximal value of the filter outputs. If the maximal filter output exceeds a threshold, we determine that a head is present, and call the x and y coordinates of that point the centroid of the head. From the x and y, we begin calculating a running average of the torso length by calculating the distance from the head down to the hips. Once the length of the torso is known, it can be used to constrain where the head may be located in future frames, and thereby increase the accuracy of the system. The current system is able to report the average length of the torso, and a future implementation will use this information to constrain the head location.

Once the head is found, tracking is accomplished by using a rectangular search window limit the possible locations for the head in the successive frame. Using the window allows the algorithm to perform the head filter operation on a much smaller portion of the image and thereby increases the speed of the algorithm. If the head is not found in the smaller window, then the algorithm attempts to locate the head by filtering the entire image, and searching for the maximal filter outputs in the larger area.

Once the head and torso have been found, these portions of the image are removed from the background to ensure that they do not get incorporated into the background model. If they were not removed, a person sitting relatively still would become incorporated into the background, and the pose determination would fail.

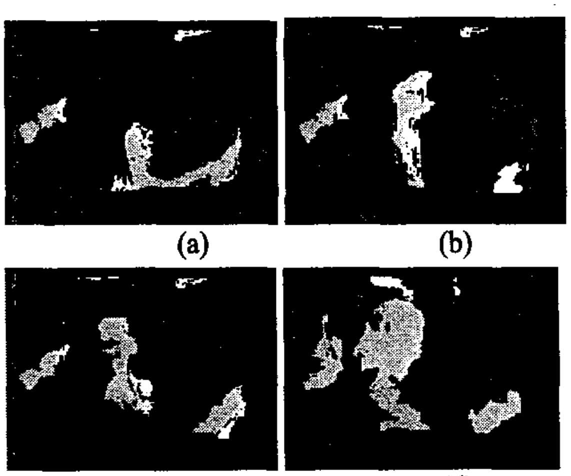
If an adult head is not found in the entire image, the size of the head filter can be adaptively modified to search for a smaller head in the same image. Using the techniques described above, we can determine the torso length of the smaller sized passenger, and continue to track the body parts using the same algorithm. This feature will be implemented in a future iteration of the program.

Experimental Setup and Evaluation:

Our current experimental setup consists of a trinocular stereo camera capable of delivering 15fps stereo images to a computer, an experimental frame of a Mercedes S-Class vehicle, an infrared illuminator, and a 650Mhz computer. The trinocular stereo camera consists of 3 black and white calibrated cameras capable of producing images up to 640x480 pixels, and has integrated hardware to deliver the corresponding depth maps at framerate. Because these cameras are black and white, they are also sensitive to near-infrared light. Using infrared illumination, the system is able to perform both in daylight as well nighttime conditions with no modification. Our system is capable of acquiring video from the stereo camera at 15fps. Once the data is acquired, the head finding algorithm runs at an average of 30fps on it. The current implementation is also capable of determining the torso length, however, it does not yet use this information to constrain where the head can be located. We expect when this feature is implemented the results will improve.

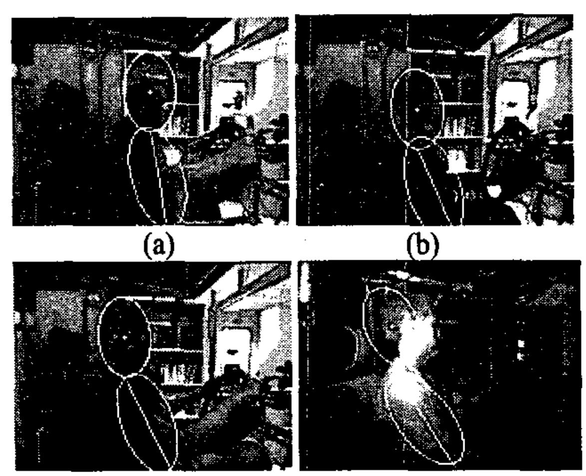
To run this experiment we acquired sequences of 500 frames at 15fps as a subject climbs into the vehicle, moves forward (toward the front of the vehicle) and backward twice. The subject is then told to move side to side twice (toward and away from the camera mounted on the B-pillar), and then rotates his or her head twice to determine if the system is sensitive to head rotation. Of these sequences one was run in complete darkness with only the infrared illuminator to light the scene. The tests were run on five different subjects, and the experimental results are shown below. The head and torso of the occupants are tracked correctly across multiple frames (Figures 5,6).

A second set of data was collected outdoors on a sunny afternoon between the hours of four and six o-clock pm with the vehicle's occupant sitting in a 2002 Infinity Q45 sedan. This series of data consists of 900 frames of data in which the subject enters the vehicle after the background has been acquired. The occupant is free to move about the cabin has they please, and the results are tabulated and shown below. In the second of the tests, the subject opened and closed the sunroof to test the system's performance to lighting changes. Between these tests, the angle between the subject and the camera was adjusted to determine the optimal camera positioning. The first two sequences were taken with the headrest outside the view of the camera, while the next four were taken with the headrest in the image (see Figure 7).



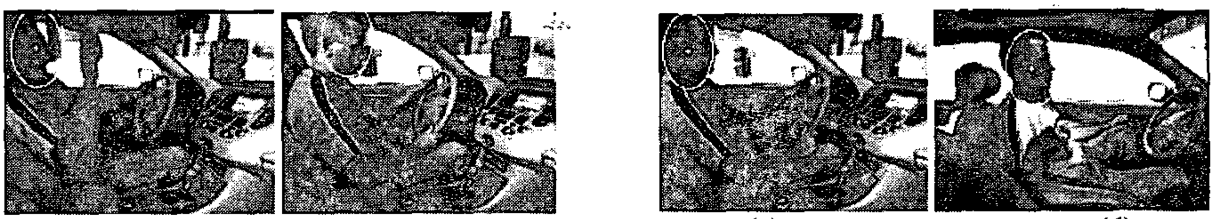
(a) (b) (c) (d)

Figure 5: Disparity maps for the frames shown in Figure 5 with points lying outside the test vehicle filtered out.



(a) (b) (c) (d)

Figure 6: Results of the Head tracking algorithm: (a) and (b) Subject leaning away from camera (c) Subject rotating head (d) Subject in darkened room, scene illuminated by infrared illuminators.



(a) (b) (c) (d)

Figure 7: Results of the Head tracking algorithm inside a vehicle: (a) Subject with sunroof closed, (b) Subject with sunroof open (c) Subject with sunroof closed again (d) Camera view with headrest in scene.

Using these sequences of videos, the following table was constructed to show the error rates. The head was considered found in the experiment if the center of the oval was located within a 10-pixel radius of the center of the occupant's ear. This condition is sufficient because once one point on the head is known, the position of the rest of the head can be determined based upon the background-subtracted depth image. The first row of the table shows the number of frames where the position of the head was incorrectly determined. The second row shows the total number of frames in which the head was present in the video, and the third row of the table is the calculated percentage of frames in which the head position was correctly determined.

Indoor Experiment	Test 1	Test 2	Test 3	Test 4	Test 5 (Complete darkness)
Frames Incorrect	3	10	30	15	106
Total Frames	230	226	190	169	341
Percent Correct	98.7	95.6	81.0	91.1	68.9

Table 1: Results of Indoor experiment

Outdoor Experiment	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Frames Incorrect	70	60	106	125	60	206
Total Frames	494	782	785	482	435	778
Percent Correct	85.8	92.3	86.5	74.0	86.2	73.5

Table 2: Results of Outdoor experiment

The indoor experimental results showed relatively good performance in well-lit conditions. Further investigation into the reasons why the head was not correctly identified in some frames revealed that the head of the vehicle occupant was not always resolved in the depth image (see example in Figure 8). This is partially due to the fact that the occupant occasionally leaned outside of the vehicle, causing the head to be filtered out of the image, and partially due to the fact that the edges and textures used by the stereo hardware could not find a proper correspondence in the other images due to poor lighting and other factors.

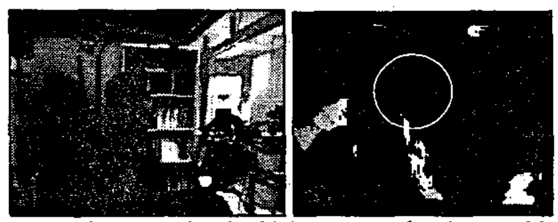


Figure 8: Example of vehicle occupant leaning outside of car. In this depth image, the head is at a depth outside of the vehicle, and has been filtered out (circle drawn in to emphasize where head should be located)

In the test sequence conducted using only the infrared illuminator, the results were less accurate than those conducted in full light. We attribute this to the fact that the illuminator had a spotlight effect, and once the subject left the point of bright illumination, the stereo camera was unable to solve the correspondence problem, and therefore the head did not appear in the dept image. This problem can be solved by either using a light diffuser to help spread the illumination more evenly across the entire scene, or by using multiple illuminators placed around the cabin of the vehicle.

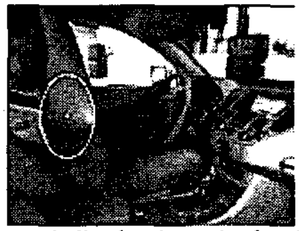


Figure 9: Arm occluding head, causing head to become less round than corner of arm.

Other errors occurred in the head detection algorithm due to the arms of the occupant either covering the head, or by forming a round shape as shown in Figure 9. When the arms occlude portions of the head, they cause the head to become less oval-shaped, and as a result, the filter output is not maximal at the center of the head. This results in an incorrect detection of the head's x and y coordinates.

The results for the outdoor experiment show that the highest percentage of correct detection occurred when the headrest was not in the scene. Additionally, the lighting change due to the sunroof being closed and opened had little effect on the algorithm.

These results can be improved with the use of more sophisticated tracking techniques such as Kalman filtering, which plan on investigating in the near future.

Conclusion and Future Work:

This paper discusses one method to locate and track the head and torso of a vehicle's occupant. Some benefits of the technique employed are that the method is computationally simple, and relatively fast. Analysis of program flow shows that the bottleneck of the system is in solving the stereo correspondence problem. This portion of the program is currently only able to capture depth images at 15 frames per second. In the upcoming months we will investigate techniques to improve the speed of both the stereo acquisition system, as well as the computation of the head location.

The system correctly detected and tracked the head on an average of 83.4% of the frames in the outdoor scene, and 92.1% of the time in the well-lit indoor scene. The darkened room performed worse than expected, and we are investigating methods to improve this such as using multiple sources of infrared light, as well as using a light diffuser to more evenly illuminate the scene. The majority of the incorrect head positioning resulted from the stereo camera not being able to solve the correspondence problem on the head. This resulted in the head being filtered out of the depth map, thereby causing the head detection to fail. We will examine possible solutions to this problem in order to increase the performance. We are also investigating methods of detecting the arms of the passenger in order to determine their positional relation to the airbag. Once this is known, a classification system can be implemented to determine the force at which the airbag should be deployed.

One potential problem of the system is that it will fail to correctly track the torso of the occupant if they are seated such that their bottom is not against the bottom back corner of the seat (e.g. when the occupant's feet are on the dashboard of the vehicle). This is a problem we hope to be able to confront in the future. In this case, it would be imprudent to fire the airbag at all, as doing so would cause the legs of the occupant to bend potentially breaking the kneecaps, leg bones, or causing the torso to impact the legs causing worse damage. As such, a simple solution would be to search the background subtracted image for objects on top of the airbag console, and if detected simply not fire the airbag regardless of the position of the occupant's torso.

In conclusion, a robust and practical "smart" airbag system may very well use other sensory signals in addition to the stereo described in this paper. Future studies will evaluate the relative merits of a multisensory system as well as focus on the development of a sensor-fusion and classification modules. In addition to the stereo methods presented in this paper, we are exploring thermal imaging and three-dimensional voxelization based techniques to determine the pose of the occupant.

References:

[1] J. Krumm, G. Kirk, "Video Occupant Detection for Airbag Deployment", IEEE Workshop on Applications of Computer Vision, October 1998

[2] I. Mikic, M. Trivedi, "Vehicle Occupant Posture Analysis Using Voxel Data," Submitted to the 9th World Congress on Intelligent Transport Systems, Chicago, Illinois, October, 2002.

[3] A. Marín-Hernández, M. Devy, "Application of a Stereovision Sensor for the Occupant Detection and Classification in a Car Cockpit", Proceedings *of "International Symposium on Robotics and Automation"*; Monterrey, Mexico, November 10-12, 2000

[4] http://www.edmunds.com/ownership/safety/articles/45863/article.html (12/11/2002)

[5]http://www.boschusa.com/AutoOrigEquip/Sensors/OccupantDetection/ (3/31/2003)

[6]http://www.delphi.com/news/solutions/monthly/ms11788-05012002 (3/31/2003)

[i] *This research is sponsored by a couple of UC Discovery Grants and by the Volkswagen-AudiCorporation. The authors also want to thank their colleagues in the CVRR Laboratory for their valuable assistance and cooperation.*