2004 IEEE Intelligent Vehicles Symposium
University of Parma
Parma, Italy • June 14-17, 2004

Occupant Posture Analysis using Reflectance and Stereo Images for "Smart" Airbag Deployment

Stephen J. Krotosky

Computer Vision and Robotics Research Laboratory
University of California, San Diego
La Jolla, CA 92093
krotosky@ucsd.edu

Mohan M. Trivedi

Computer Vision and Robotics Research Laboratory
University of California, San Diego
La Jolla, CA 92093
mtrivedi@ucsd.edu

Abstract

Robust detection of vehicle occupant posture is necessary for intelligent airbag deployment. This paper presents a vision-based method of estimating the size, posture and pose of the occupant. Utilizing raw reflectance and stereo disparity images, this algorithm presents a mixed-mode approach to finding occupant features. Extensive experiments show the viability of this mixed-mode method in identifying the occupant's head location and suggest the feasibility of extending the analysis system to include detection of features such as occupant arms and foreign objects.

1. Introduction

In recent years, occupant safety in automobiles has greatly improved with the advent of airbag systems, crumple zones, and other safety enhancements. Unfortunately, these enhancements, specifically airbags, have the added cost of potential injury to persons of improper size or in improper positions.

It is the primary objective of this research to develop the framework and algorithms for a vision-based system that can accurately and robustly describe occupant size, pose and posture to aid in the control of a safer intelligent airbag deployment system.

1.1 Previous Work

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 a low level of detail [3] or requiring an extensive training set of data for proper operation [4], [5]. Our method, however, requires no 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.

2. Head Detection and Tracking

First, a background model is generated based on N frames of disparity data captured in an empty cabin. Using the disparity image in the background model allows for a larger degree of lighting changes than using the reflectance images. Once the background model is obtained, the current foreground data is computed. The SVS API can give high-valued noisy response in regions where the stereo is incomputable, so disparity values that are too high are removed from the current image. Disparity values that fall outside the car, specifically those in the car's window region are also removed from the current disparity image. This refinement helps to remove extraneous and invalid foreground data.

After this threshold is applied, a foreground map is generated through background subtraction. To eliminate more of the extraneous stereo data, a median filter operation is performed, followed by a morphological opening. Then, connected component analysis removes all areas smaller than the minimum head size. This final binary foreground map is combined by a logical AND with the current disparity image. The result is the current foreground disparity image.

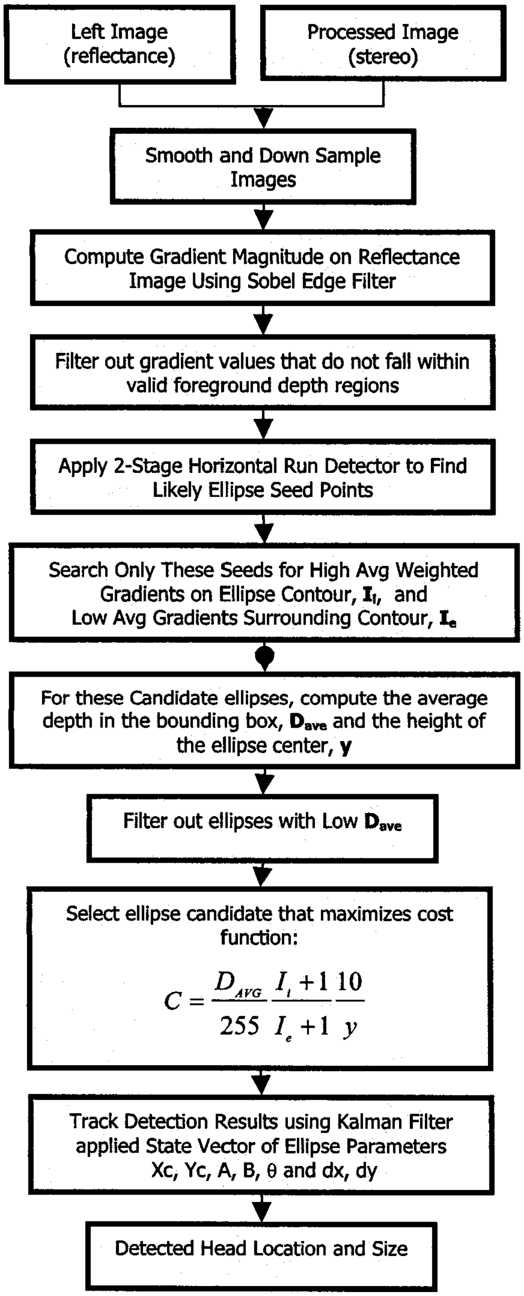

Figure 1 – Head Detection Algorithm Flowchart

With the foreground disparities computed, attention can now be given to detecting the head position. The flowchart of the following algorithm is displayed in Figure 1. Based on the method by [11], edge data from the reflectance image and depth information are used to yield a robust head detection algorithm for stereo imagery. Edges are found in the left raw image using Sobel operators. Only edges that fall within the current disparity-based foreground map are kept. Given a pre-computed set of ellipse templates of various sizes, angles, and eccentricities that represent appropriate head sizes and positions, the best-fit ellipse/head position is found by maximizing the following cost function:

$$C = \frac{D_{AVG}}{255}\frac{I_i+1}{I_e+1}\frac{10}{y} \qquad (1)$$

D_{AVG} is the average disparity within the rectangular region enclosing the candidate ellipse, I_i is the value on the interior contour of the ellipse, I_e is the value on the exterior contour of the ellipse, and y is the vertical center position of the ellipse. This cost function gives preference to good fitting ellipses that have consistent depth values and are in the upper part of the image where the head is likely to be found.

To reduce the effects of measurement noise, seven ellipse parameters are tracked using a Kalman Filter. The tracked parameters are the ellipse center coordinate, ellipse axes size, the inclination angle, as well as the position change from the previous detected location.

3. Occupant Pose Extensions

The design and testing of the head detection and tracking algorithm leads one to ask how much more of the occupant's pose and position can be accurately captured and classified. To more thoroughly describe the occupant posture and events in the cabin, a sequence of algorithms is investigated. Using the head detection location as a starting point, the area between the occupant and the dash can be analyzed for foreign objects. In addition, the location of the top of the arm can be estimated and used to estimate the occupant's left arm pose.

3.1 Foreign Object Detector

In addition to locating the head, it is important to determine if there are potentially dangerous foreign objects located between the head position and the dashboard. Such objects could inflict injury on the passenger in the event of an accident. To determine if such a potential exists, a region of interest is defined from the left-most valid stereo region to a point that is an ellipse width away from the detected head center. If the number of active foreground pixels in this region exceeds a given threshold, then the region is filled with a potentially dangerous object. Such detectable objects include legs on the dashboard, hands and arms as well as any object the subject may be holding (e.g. laptop, coffee, etc).

3.2 Top of the Arm Detector

To further analyze occupant posture, it is necessary to detect more feature points. One point of interest is the top of the left arm. This point can be used in arm pose detection. A rectangular region is defined so that it is half the height of the detected ellipse and 1.5 times the width. The region is centered at a point one-ellipse height below the center of the detected ellipse. The maximum disparity value in this region is the point closest to the camera and is postulated to be location of the top of the arm.

3.3 Arm Pose Detector

Using the top of the arm as a seed point, the Hough transform is used to estimate the pose of the left arm up to the elbow. This is done by first performing Canny edge detection on the full-sized raw image. The Hough transform is then performed on the edge map and lines that exceed a manually selected threshold are kept. The line closest to the top of the arm point is kept as the first arm contour. The next closest line parallel to the first line is estimated as the second arm contour. Candidate line pairs can then be evaluated based on the disparity data. It is desirable that the area between the two lines has relatively constant depth values. This would reinforce the validity of the chosen lines. The development of this algorithm is still in its infancy, but shows great promise that it can robustly detect the arm.

4. Testbed and Evaluation

All experiments were performed using the LISA-P test bed described in [8]. The LISA-P is a Volkswagen Passat wagon equipped with multiple cameras and a computer to synchronously capture data from multiple modalities and execute real-time processing algorithms. A Digiclops Stereo Camera [9] was used to capture the stereo data (left & right only) and SRI's Small Vision System API [10] was used to compute the depth disparities from the captured stereo data. The Digiclops Camera is attached to the driver's side roof rack, positioned so that the camera is looking through the sunroof into the passenger side cabin. The stereo images are captured at 320x240 pixels at a frame rate of 15 fps. An example of the left reflectance image and its corresponding stereo disparity image can be seen in Figure 2.

Figure 2 – Reflectance and disparity images of a subject in the LISA-P testbed

Experiments were performed in the LISA-P on three non-consecutive mornings. The occupants were asked to enter the car and perform a series of movements and tasks while the car was driven at road speeds. The occupants' movements were captured and saved to disk. These tests were conducted with five subjects for a total of 21,379 frames of collected data.

The occupant script was divided into a Position Test, which tests the algorithm's ability to detect the head at various positions in the cabin and track the head's movement; a Hand Motion and Object Test, which was designed to evaluate the algorithm's robustness to competing objects and hand motion in the scene; and a Free Motion Test that was designed to catch other potential situations for detection error, as the subject was free to move as they wish during this test.

Test result summaries are listed in Table 1 and a specific breakdown of the three tests for a certain subject is outlined in Table 2.

4.1 Head Detection and Tracking Results

Head detection results were considered correct if the algorithm placed an ellipse center point somewhere on the subject's head. For a successful detection, the estimated ellipse size needed to be comparable to occupant's head size. Examples of successful detection are shown in Figure 3.

Similarly, head tracking results were considered correct if the algorithm placed the tracked ellipse center point somewhere on the subject's head.

Subject	Position Test	Hand Motion & Object Test	Free Motion Test	All Tests
Male 5' 8" 	*97.3%*	92.6%	100.0%	95.4%
Male 5' 9" 	99.7%	93.4%	84.5%	94.9%
Female 5' 0" 	99.3%	96.3%	99.9%	98.3%
Female 5' 8" 	98.7%	92.0%	95.8%	94.0%
Female 5' 11" 	99.8%	99.8%	99.8%	99.8%
All Subjects	99.1%	94.8%	96.9%	96.4%

Table 1 – Head Detection Results Summary

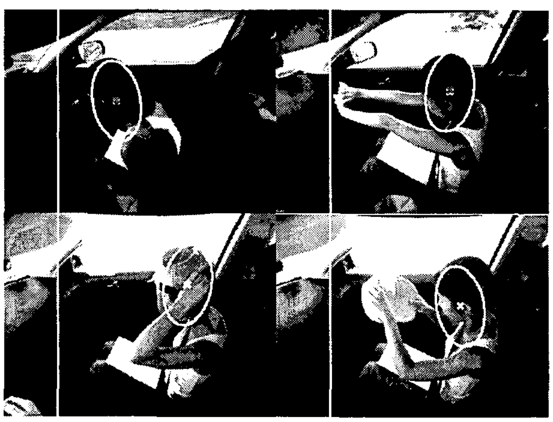

Figure 3 – Illustrations of successful detection and tracking of occupant in four different situations

Occupant Task	% Detected	% Tracked
Sit Normal	100.0%	100.0%
Lean Halfway Forward	100.0%	100.0%
Lean Completely Forward	93.7%	94.3%
Return to Normal 1	100.0%	100.0%
Lean Back	100.0%	100.0%
Return to Normal 2	100.0%	100.0%
Lean Right	99.6%	100.0%
Lean Left	95.4%	95.4%
Return To Normal 3	100.0%	100.0%
Position Test Totals	*99.2%*	*99.3%*
Move Hands About Cabin	98.5%	99.1%
Open Glove Box	100.0%	100.0%
Put hands on face & stretch	78.8%	80.0%
Adjust Radio	99.6%	100.0%
Place hat in lap	99.1%	100.0%
Put hat on head	92.9%	94.6%
Move With Hat	99.3%	100.0%
Remove Hat	83.6%	85.9%
Feet On Dashboard	96.2%	96.5%
Hand Motion & Object Totals	*95.6%*	*96.3%*
Free Motion Test	*99.5%*	*99.9%*
All Test Totals	*97.9%*	*98.3%*

Table 2 – Detailed Head Detection and Tracking Results for Female 5'0"

4.2 Head Detection Error Analysis

With an overall detection rate of 96.4%, this head detection algorithm is very successful. There are however, certain instances where the algorithm could be improved. Notice in Table 2 that the detection rates drop when the occupant leans forward or to her left. These drops can be explained by the nature of the stereo data and the camera setup.

When using the SVS API, the left most 64 columns are invalidated in the disparity image, as it is the value of the disparity search range used in this test. This corresponds to the area in the left image that is presumed to not overlap with the right image. If the head falls in or near this region, denoted by the white vertical line in Figures 3, 4 and 5, the invalid stereo data region can distort the head contour, causing detection problems.

Similarly, if the occupant leans too close to the camera, the head can fall out of the range of valid disparities. The

SVS software will return invalid data in that area where the head exceeds the minimum distance from the camera where stereo data is computable. Naturally, this invalid data can cause problems detecting the head. These out of position errors are indicated in Figure 4.

Figure 4 – Illustrations of unsuccessful head detection due to invalid disparity data

These errors are due entirely to camera selection and placement. This can easily be resolved by selecting a camera with a field of view and baseline that contains the entire range of potential occupant positions. Under the current setup, the Digiclops' enclosure size and baseline limits the camera positioning options and the position selected for these tests is the most optimal. These errors directly contribute to lower results for leaning forward and leaning left, as well as the low numbers associated with the Free Motion Test for Male 5'9" as the subject repeatedly moved out of frame.

The biggest falloffs in detection rates occur in the hand motion and object tests. Specifically, when the occupants put on or remove the hat, or put their hands on their face, the detection rates drop significantly. This occurs because other objects in the cabin, namely the hat and hands, look similar to the head in the disparity image and may maximize the detector cost function better than the head location for that frame. Examples of these errors are in Figure 5.

Figure 5 – Illustrations of unsuccessful head detection due to competing objects

These competing object errors are critical since the head is not only being detected incorrectly, but there also may be potentially dangerous foreign objects in the scene that could cause further harm in the event of an improper airbag deployment.

To remedy these types of errors, further processing is necessary. It is not enough to only search for the head in the current proposed manner. To further validate the choice of head position, sub-facial cues such as eyes and noses could be searched for inside of the ellipse area. This would help to eliminate areas where the stereo data presents a valid option for a head location, since the reflectance data would invalidate it.

It may also prove useful to classify all the stereo disparity data as in a safe or unsafe position relative to the real world automobile cockpit. This would allow for validation of the occupant's safety by a method other than the head location and would help make correct decisions in situations where foreign objects are in the scene. This is critical because when foreign objects are in the scene, decision errors could occur both when the head location is detected correctly (by using the correct head location to decide it is safe when the foreign object could cause injury) and incorrectly (by detecting the head location in the incorrect position thereby giving an inaccurate decision).

Despite the potential for these detection errors, the overall error rate is still very low. The errors also seem to occur in short isolated bursts. Most of the time, errors occur in a single frame or two and are corrected in the next frame. The time for each error is a fraction of a second. Considering that the tracked head location is often correct when these errors occur, the time when both the detector and tracker are completely wrong is small.

4.3 Occupant Pose Extension Results

Figure 6 shows two example results of the three occupant pose extension algorithms. The raw image is on the left, and its corresponding foreground disparity image is on the right. The two vertical lines represent the region searched for potentially dangerous objects.

Note in the top image, the hands and arms are categorized as potentially dangerous, denoted by the diagonal line connecting the two vertical lines, while in the bottom image, there is no potentially dangerous object and therefore no diagonal demarcation.

The light gray cross on the arm denotes the detected location of the top of the arm and the light gray rectangle in the disparity image is the top of the arm search area. While not exactly placed, the point is sufficient enough to guide the arm pose detector to the correct set of lines in the Hough transform that define the arm pose.

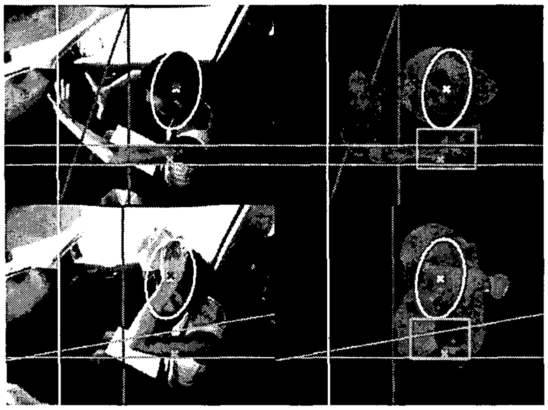

Figure 6 – Occupant Pose Extension Results for two different poses

Table 3 shows the detection rates for the head detector, potential danger detector and the top of the arm detector for the Female 5' 0" subject. The false positive rate for the foreign object detector was 3.8% and the miss rate was 15.6%.

Occupant Test	Head Detection	Foreign Objects	Top of Arm
All Tests	*98.6%*	*93.9%*	*82.2%*
Position Test	99.2%	98.0%	84.4%
Hand Motion & Object Test	96.8%	87.6%	75.9%
Free Motion Test	100.0%	95.3%	86.7%

Table 3 – Head Detection, Potential Danger Detection and Top of the Arm Detection Results for Female 5'0"

5. Summary and Conclusions

Occupant posture and pose information can be used to greatly enhance safety decisions in airbag deployment. The mixed-mode vision approach in this paper provides a good framework for robustly detecting and tracking the occupant head. Extensive experimental tests have proven the success and reliability of this method. The groundwork has also been laid to extend occupant feature detection to arms and foreign objects.

Acknowledgments

This research is sponsored by a UC Discovery Grant in partnership with Volkswagen Research Laboratory. The authors gratefully acknowledge the valuable assistance of Dr. Klaus Schaff, Dr. Arne Stoschek, and Mr. Sascha Sertel of VW Research Laboratory. The authors also offer their gratitude to the anonymous reviewers whose suggestions helped to refine and improve this paper. We also wish to thank Jean Trinh, Tasha Vanesian, Shinko Cheng and Leslie Wood for their 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] M. M. Trivedi, et.al., "Occupant Posture Analysis with Stereo and Thermal Infrared Video: Algorithms and Experimental Evaluation," *IEEE Transactions on Vehicular Technology, Special Issue on In-Vehicle Vision Systems,* 2003 (under review).

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

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

[11] A. Eleftheriadis, A. Jacquin, "Face Location Detection For Model-Assisted Rate Control In H.261-Compatible Coding Of Video", *Signal Processing: Image Communication*, Vol. 7, pp.435-55, 1995.