Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

Grayscale Correlation based 3D Model Fitting for Occupant Head Detection and Tracking

Zhencheng Hu, *Member, IEEE*, Tetsuya Kawamura and Keiichi Uchimura

***Abstract*— Occupants inside the vehicle can be deadly injured by the deployment of airbag at the time of crash. New collision safety technology requires classifying the occupant and tracking their position in real-time in order to adaptively deploy the air bag. This paper presents a fast 3D model fitting algorithm based on grayscale correlation of stereo disparity data, to detect and track occupant head position. The proposed system uses stereo vision with IR illumination for depth data acquisition. By detecting body center line and extra-near disparity calculation, this method is proven to be robust and accurate in variant lighting condition and occupant movement. Evaluation of the method shows over 98% correct head detection and near 100% correctness with head tracking.**

I. INTRODUCTION

WITH the development of collision safety technology in recent years, delicate control of air bag deployment which adaptively deploys the airbag depending on occupants' body shape, weight and position, has being intensively studied during past few years. The main purpose of the smart air bag system is to deal with the threat that occupants may be seriously injured by the deployment of an air bag at the time of crash if the occupant is too near to the airbag.

The National Highway Traffic Safety Administration (NHTSA)[1] specifies different classes for the occupancy including infants in rear facing infant seats, children and small adults, and out-of-position zones for the human occupants, on which the air bag deployment has to be controlled.

Research on detecting the type and position of occupant can be divided into 3 main categories based on different sensing technologies: I) Weight sensors on the seat measure the pressure distribution and classify the occupant into different types[2][3]; II) Electric-magnetic or ultrasound sensors that detect the change in the electric-magnetic field to confirm occupant type and position[4]; III) Computer vision sensors that directly detect occupant head and body position with 2D or 3D information, and classify the occupants[5]-[9]. Category I and II are the most popular sensors in the market in current stage of air bag control, which requires a reliable classification of adults, children and rear-faced child seats. However they are not adaptable for precisely detection of occupant position and posture, which is vital to the delicate control of air bag deployment.

Vision sensor provides the richest information of occupant position and posture. Depending on the number of cameras used, these studies can be further divided into two categories: monocular camera based methods and stereo vision based methods. Monocular camera always employs edge, contour and other image features to detect ellipse-liked shapes for head detection. By combining with the infrared detector, single camera solution can also obtain satisfied result in some well-controlled environment. However, it suffers from strong shadows, hot weather and insufficient 3D information which is necessary for functions such as the *out-of-position detection*. Stereo vision based methods use two co-planar cameras to calculate the disparity data and detect occupant head position and posture. Many algorithms employ the general 3D model fitting method to detect the ellipsoid-like 3D shape from a range image obtained from the stereo rig. M. Trivedi [5]-[7] uses shape and size constraints to eliminate search regions for less computation purpose, which may have serious side-effects that the head region can be also eliminated when it appears relative smaller than other ellipsoid-like shapes such as waving arms and shoulders. B.Alefs [9] uses depth data to recovery the occupant body surface and edge data to generate head candidate. Head recognition was carried out with a large trained dataset.

To achieve real-time performance while keeping high accuracy of occupant head detection, this paper presents a fast 3D parametric model fitting algorithm based on grayscale correlation of range data. Comparing with the traditional 3D parametric model fitting algorithms, this method simplifies the problem of searching 3D model from depth image into 2D grayscale correlation problem, which simultaneously determine all parameters with the best fitting model. By applying the proposed algorithm into occupant head detection application, this paper also proposes a body centerline segmentation method as well as a multi-resolution disparity

Manuscript received December 30, 2006.

Z. Hu is with the CSEE Dept., Graduate School of Science and Technology, Kumamoto University, 2-39-1, Kurokami, Kumamoto, Japan (Tel: 81-96-3423894; e-mail: hu@cs.kumamoto-u.ac.jp).

T.Kawamura and K.Uchimura are with the Graduate School of Science and Technology, Kumamoto University (e-mail: tetsuya@navi.cs. kumamoto-u. ac.jp, Uchimura@cs.kumamoto-u.ac.jp).

generation algorithm in order to deal with body occlusion and extra-near disparity calculation problems.

In the remainder of this paper we will present a brief overview of traditional 3D parametric model fitting and our new approach based on grayscale correlation of range image (Section 2), a detail implementation of our approach (Section 3), and experimental results in the purview of an occupant head detection system (Section 4).

II. 3D PARAMETRIC MODEL FITTING ALGORITHM

A. Problem Description

Given an image frame (e.g. range image or edge image), the 3D parametric model fitting problem is to find the 3D parameters (e.g. 3D position and orientation, scale factor, intrinsic parameters, etc.) of the model. Figure 1 shows an example of finding an ellipsoid in a range image. The total number of ellipsoid 3D parameters is 9 including 3 rotation and 3 translation parameters, as well as 3 scaling factors along X, Y and Z-axis.

Research on 3D model fitting leveraged earlier work done in (Lowe[10], 1991) for generic 3D parametric model fitting. Image formation is modeled as a mapping of a 3D model into the image. Although the inverse mapping is non-linear due to the trigonometric functions of perspective projection, the resulting image changes smoothly as the parameters are changed. Therefore, local linearity can be assumed and several iterative methods can be employed for solving non-linear equations (e.g. Newton's method). Upon finding the solution for one frame, the parameters are used as the initial values for the next frame and the fitting procedure is repeated. The traditional approach can be very computational and time consuming and is not adaptive to the real-time required occupant head detection application.

B. Our Algorithm

With the assumption of local linearity, we can prepare a lookup table of possible combination of all parameters except 3D position (X, Y, Z), which will be determined by the later process of grayscale correlation. Rotation and scale parameters are used to generate the LUT in the case of ellipsoid detection. To simplify the process, only certain combination of rotation and scale parameters are adopted by the constraint of occupant physical position and posture. Here, 3 rotation angles $\{0, +45^\circ, -45^\circ\}$ along X and Z-axis are combined with 3 different ellipsoid shapes. Scale factors are defined by the possible movement range of the head.

Equation of 3D ellipsoid is shown as follows:

$$Z = Z_0 + c_i\sqrt{1-\frac{(X-X_0)^2}{a_i^{\,2}}-\frac{(Y-Y_0)^2}{b_i^{\,2}}} \qquad (1)$$

where a_i, b_i, c_i are scale parameters and X_0, Y_0, Z_0 are the 3D world coordinates of ellipsoid center.

Perspective projection equation (2) is adapted to project 3D ellipsoid surface points to the 2D image coordinates.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} & 0 \\ R_{21} & R_{22} & R_{23} & 0 \\ R_{31} & R_{32} & R_{33} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \qquad (2)$$

where camera's intrinsic parameters like lens focal (f_x, f_y) and optical center coordinates (u_0, v_0) are obtained through some preprocessing steps like camera calibration. Rotation matrix elements $R_{11} \sim R_{33}$ are retrieved from the parameter LUT.

To match with the disparity image, we use the following normalization equation to convert range data into intensity value.

$$I(x,y) = \frac{Z_{\min}}{Z} \times (2^N - 1) \qquad (3)$$

where $Z_{\min}$ is the minimum distance from ellipsoid surface points to camera. N is the bit-value of intensity image.

Figure 2 shows some examples of models generated from the parameter LUT.

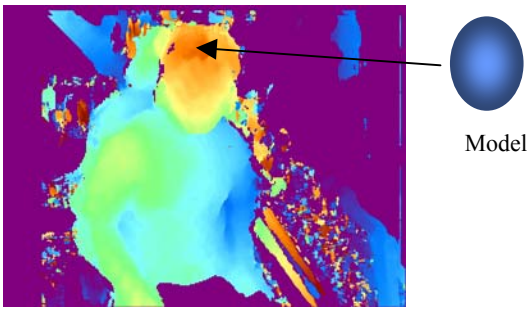


Figure 1. Searching a 3D model in a range image

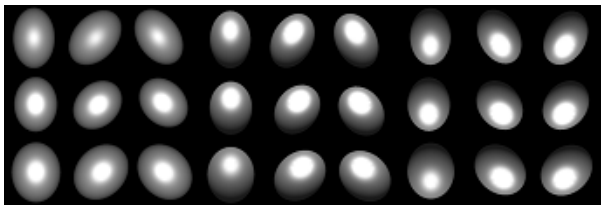


Figure 2. Parametric models of range data

C. 3D Model Fitting

Once the 3D parametric models are generated, we can simply adapt the traditional grayscale correlation algorithm to find a match between models and target range image.

2D grayscale correlation algorithms are well studied for decades and many acceleration techniques like multi-level and pyramid sub-sampling technologies have been proposed. To add tolerance to intensity change, we use the normalized

grayscale correlation (NGC) equation to find the best matching from multiple models.

$$r = \frac{S\sum I(x,y)d'(x,y) - (\sum I(x,y))\sum d'(x,y)}{\sqrt{\{S\sum I^2(x,y) - [\sum I(x,y)]^2\}\{S\sum d'^2(x,y) - [\sum d'(x,y)]^2\}}} \quad (4)$$

where $I(x,y)$ and $d'(x,y)$ are intensity values of model image and normalized target image respectively. S is the effective pixel number. Matching score $r = 1$ refers to the perfect match and $r = 0$ means not match at all.

The normalization process on the target image is a general histogram smoothing as described in equation (5).

$$d'(x,y) = (2^N - 1)\frac{d(x,y) - d_{\min}}{d_{\max} - d_{\min}} \quad (5)$$

where $d(x,y)$ is the original disparity value on pixel (x, y), $d_{\min}, d_{\max}$ are the minimum and the maximum disparity value in the region. N is the bit-value of disparity map.

Result of our grayscale correlation algorithm presents not only the position but also the best model, which indicates the rotation and scale parameters simultaneously.

III. SYSTEM IMPLEMENTATION DETAILS

The system is designed as a co-planar stereo camera with constructive infrared illumination light source. The stereo rig is mounted on the center roof console near the back mirror. Generally it should have few centimeters baseline and wide-angle lens that can overview the whole passenger's cabinet.

A. Constructive Illumination Lighting System

A fast stereo algorithm [11] is adapted to generate disparity map with two synchronized video source input at 30 frames per second. To overcome the uneven illumination and shadow problem for real outdoor environment, an infrared pulsed illumination lighting system is installed, combining with band-pass filtered lens to cutoff all un- necessary wavelength light.

A disadvantage of block matching based dense disparity algorithm is the aperture problem. The aperture problem arises as a consequence of the ambiguity of one-dimensional intensity on left and right image through out the horizontal Epipolar line. No disparity data can be derived for an even intensity region like dark or over lighted regions.

We tested different kinds of light patterns, and the cross pattern of light-dark-light with an angle of ±45 degree showed the best performance. Figure 3 shows an example of disparity map result without/with constructive light.

B. Background Subtraction

To eliminate passenger's seat, door and other interior regions from the range image, background subtraction is carried out for every new frame. Background range image were generated as an average of 30 frames' range image for the empty seat. Automatic background generation will be further implemented according to the sensors' output of seat lateral position and reclining angle.

Post-processing includes binarizing, morphological process, and blob analysis. The biggest blob that satisfies the position and area constraints will be extracted as occupant body's candidate region. Figure 4 shows the background subtraction results.

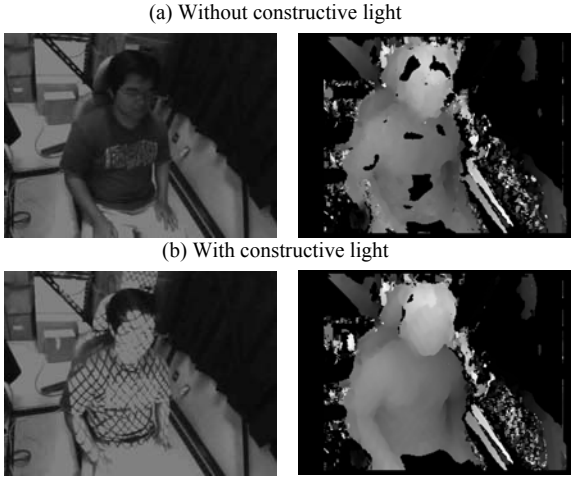


Figure 3. Disparity map result without/with constructive light

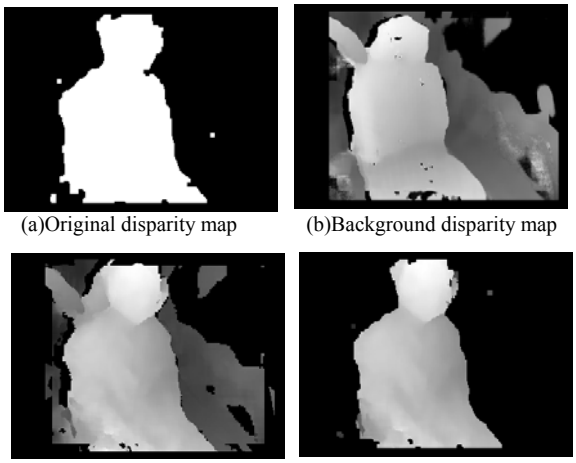


Figure 4. Background subtraction result

C. Composition of Multi-resolution Disparity Maps for Near Distance Disparity

Fast stereo processing algorithms [11] always use a fixed maximum disparity value to accelerate the matching process. For example, a maximum disparity of 32 pixels leads to the

maximum searching distance of 32 pixels. Disparities over the maximum disparity will be omitted.

According to the basic equation of stereo disparity shown in (6), the maximum distance leads to the minimum detection distance, as the baseline b and lens focal f is unchanged.

$$d = x_l - x_r = \frac{fb}{z} \quad (6)$$

Figure 5 shows an example of extra-near distance target that cannot obtain disparity data.

To enlarge the disparity range for extra-near target detection, we propose a composition algorithm of multiple resolution disparity maps. A lower resolution stereo image pair will generate a wider detection range disparity map since its pixel size is bigger than the general resolution image pair. Figure 6 shows the composition result of disparity maps generated from 160x120 and 320x240 stereo images.

D. Foreground Segmentation with Body Center Line

Extracted occupant body ROI may include multiple ellipsoid-liked regions which has similar size with the head, such as shoulders, waving arms, and other objects. Examples are shown in Figure 7. In this paper, we extended Russakoff's concept [12] of body center line to 3D region segmentation to eliminate the ambiguities.

Assuming passenger is always sitting on the seat, so that the lower part of body's ROI is relatively stable and can be used as the reference part to segment the ROI. Detail steps are shown as follows:

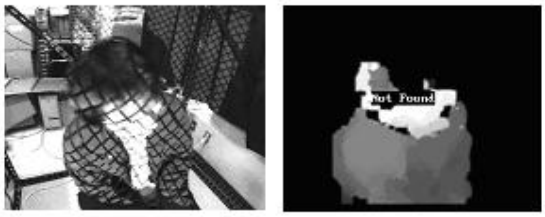
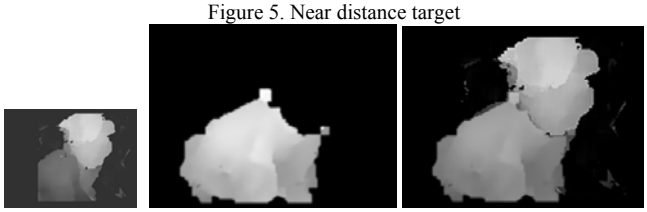


Figure 5. Near distance target



(a) (b) (c)

(a)disparity map generated from 160x120 stereo images, (b) from 320x240 stereo images, (c)Composition of above disparity maps

Figure 6. Composition of multiple resolution disparity maps

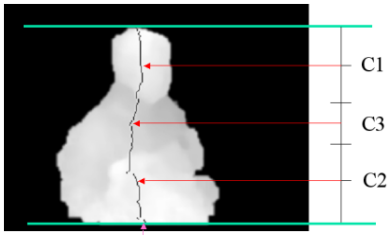
Figure 7. Examples of ellipsoid-liked objects extracted from body ROI

Step 1. After the preprocessing steps described in the above sections, calculate the so-called horizontal median points on each row of the binary ROI image based on Russakoff's algorithm.

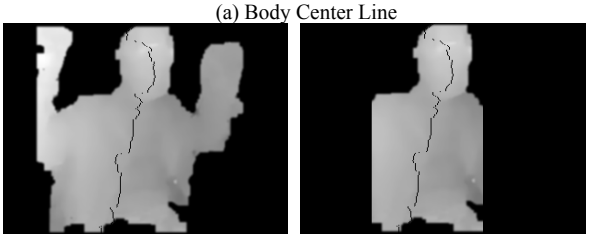
Step 2. Detect the upper center position C1 and lower center position C2 along the body center line, where C1 and C2 are on the rows of 1/5 and 4/5 of the ROI height respectively as shown in Figure 8(a).

Step 3. If the slope angle of line C1C2 is less than threshold k (the occupant is in the normal seating position), then we can simply vertically cut off the regions that are further than a predefined distance to C1C2's middle point C3. An example is shown in Figure 8(b).

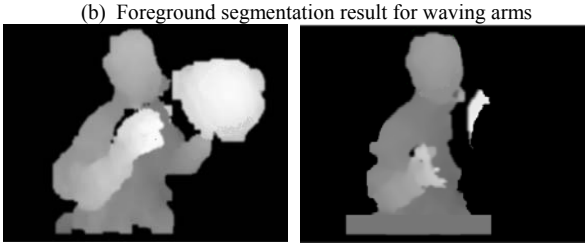
Step 4. If the slope angle is larger than threshold k (the occupant is in the leaning position), the cut-off lines will be parallel to line C1C2, while keeping the predefined distances.



(a) Body Center Line



(b) Foreground segmentation result for waving arms



(c) Filtering result by disparity constraint

Figure 8. Foreground segmentation with body center line

Step 5. Segmented foreground region will be further filtered by the constraint of disparity. The ideal disparity data on each row *i* can be calculated through the following linear interpolation equation.

$$d_i = d_1 - \frac{y_1 - i}{y_2 - y_1}(d_2 - d_1) \quad (7)$$

This constraint will eliminate most of the outliers and other objects in front of the body ROI. An example is shown in Figure 8(c).

Step 6. The result image will be further normalized for the 3D model fitting process described in Section II.

IV. EXPERIMENTAL RESULTS

The proposed algorithm was tested under various sizes of passengers and different postures that occupants may behave during the normal driving situations. The stereo vision system was equipped with two gen-locked CCD cameras. Stereo images were captured by a Matrox Meteor2/MC frame grabber board and all processing was done by a Pentium IV 2.66GHz PC. The stereo baseline is 64 mm, and the lens focal is 2.8 mm. 320x240 disparity maps were generated at the speed about 25 ms/frame with the maximum disparity of 32 pixels.

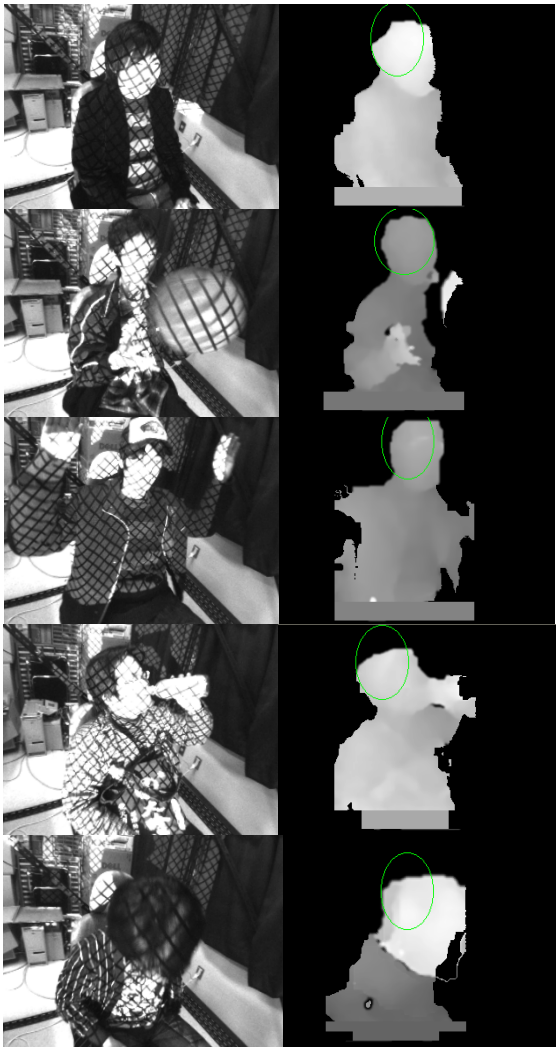
Totally 16 adults testers including 12 males and 4 females were chosen for the test. With their height distributed from 153cm to 183cm and weight distributed from 50kg to 80kg, the testers were supposed to cover the main range of adult passenger sizes. They were asked to perform all kinds of postures that could be happened during the real driving situations, like readings, waving arms, drinking, etc. Each test was continuously captured for 1500 frames. Table 1 shows the test results of different situations.

TABLE I
TEST RESULTS OF DIFFERENT SITUATIONS

Tester (Posture)	CORRECT DETECTED	False Detected	Not Detected	Rate of Correct(%)
1(N)	1500	0	0	100
2(N)	1500	0	0	100
3(N)	1500	0	0	100
4(N)	1500	0	0	100
5(N)	1500	0	0	100
6(N)	1500	0	0	100
7(N)	1500	0	0	100
8(N)	1478	9	13	98.5
9(A)	1500	0	0	100
10(B)	1477	17	6	98.5
11(C)	1416	7	77	94.4
12(D)	1369	2	129	91.3
13(E)	1500	0	0	100
14(F)	1497	0	3	99.8
15(G)	1460	0	40	97.3
16(H)	1484	8	8	98.9
TOTAL	23681	43	276	98.7

Postures: N=Normal siting position, A=Reading book,
B=Playing basketball, C=Moving body in different direction,
D=Waving arms around head, E=Reading newspaper,
F=Talking with a mobile phone, G=Drinking water, H=Wearing a cap.

Tester 1 to 8, who were sitting straightly in the normal position, showed the best performance near 100% correct detection rate. Tester 9 to 16, who were asked to perform different kinds of movement and postures, still showed a very high detection rate about 97.5%. The overall correct detection rate is 98.7%. Some very difficult situations like partially occluded target, extra-near target and multiple ambiguities were also correctly detected. Figure 9 shows some examples.



(a)original image (b)detection result
Figure 9. Occupant Head Detection Results

False detection (<0.2%) were happened under the situations of occlusion and head was not detected (<1.2%) mostly due to the situation that occupant was out of position.

Figure 10 shows some false examples.

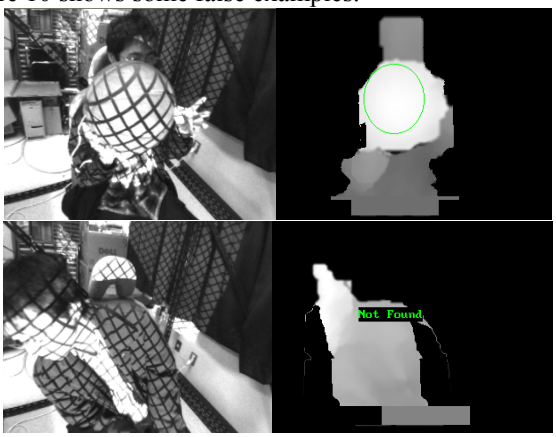


Figure 10. Falsely Detected Examples

False detection and miss detection generally happen within very short period of time. Tracking of head position in both intensity image and disparity map will largely help to locate the head position even for fully occlusion case. Tracking can also reduce searching area by predicating head position. Some preliminary tests were carried out and showed very satisfied results.

V. CONCLUSION

Occupant head detection is sensitive to the variation of illumination, occupant posture and body size. To achieve real-time performance while keeping a high accuracy of occupant head detection, this paper presents a fast 3D parametric model fitting algorithm base on grayscale correlation of range data. Evaluation of the method shows over 98% correct head detection. Combining with head tracking algorithm on intensity image and disparity map, the proposed algorithm will perform near 100% correct detection.

REFERENCES

[1] Federal Motor Vehicle Safety Standards; Occupant crash protection; final rule; Department of Transportation; Federal Register; vol. 65; no. 93; pp. 30680-30770; 2000.

[2] K.R. Kennedy, J.F.Nathan, M.Shridhar, "An LVQ-based Automotive Occupant Classification System", in Proc. 18th Intl. Conf. on Pattern Recognition, Vol.2, pages.662-665, 2006

[3] K.Lasten, *et al.*, "iBolt Technology – A Weight Sensing System for Advanced Passenger Safety", *Advanced Micosystems for Automotive Application 2006- Part 2,* VDI-Buch, Jurgen Valldorf and Wolfgang Gessner, pages 171-186, 2006

[4] R.Seip, B.Adamczyk, D.Rundell, "Use of Ultrasound in Automotive Interior Occupancy Sensing: Optimum frequency, beam width, and SNR from Empirical Data", in Proc. IEEE Ultrasonics Symposium, Vol.1, Pages. 749-752, 1999.

[5] Ivana Mikic, Mohan Trivedi, "Vehicle Occupant Posture Analysis Using Voxel Data", in *Proc. Ninth World Congress on Intelligent Transport Systems*, October 2002.

[6] S.Krotosky, S.Cheng, and M. Trivedi, "Real-Time Stereo-Based Head Detection using Size, Shape and Disparity Constraints", in *Proc. IEEE Symposium of Intelligent Vehicle 2005*, 2005

[7] M. Trivedi, S. Y. Cheng, E. Childers, and S. Krotosky, "Occupant posture analysis with stereo and thermal infrared video: Algorithms and experimental evaluation," *IEEE Trans. Veh. Technol.*, vol. 53, no. 6, pp.1968–1712, Nov. 2004.

[8] John Krumm, Greg Kirk, "Video Occupant Detection for Airbag Deployment", in *Proc. Fourth IEEE Workshop on Applications of Computer Vision*, October 1998

[9] B. Alefs *et al.*, "Robust occupancy detection from stereo images," in *Proc. IEEE Intelligent Transportation Systems Conference*, 2004.

[10] D.G. Lowe, "*Fitting parameterized three-dimensional models to images*," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 5, pp. 441-450, May 1991.

[11] K. Konolige, "Small vision systems: hardware and implementation," in *Eighth International Symposium on Robotics Research*, 1997.

[12] D.B.Russakoff, M.Herman, "Head tracking using stereo", *Machine Vision and Applications*, Vol. 2002, No.13, pp.164-173, 2002

[13] Stan Birchfield, "An Elliptical Head Tracker", in *Proc. 31st Asilomar Conference on Signals, System, and Computers*, November, 1997

[14] Ruigang Yang, "Model-based Head Pose Tracking With Stereo Vision", in *Proc. 5th IEEE Intl. Conf. On Automatic Face and Gesture Recognition*, pages 255-260, May 2002

[15] A.M. Hernandez, M. Devy, "Application of a Stereovision Sensor for the Occupant Detection and Classification in a Car Cockpit", in *Proc. 2nd International Symposium on Robotics and Automation*, LAAS No.00444, pp.491-496, November 2000

[16] Changming Sun, "Fast Stereo Matching Using Rectangular Subregioning and 3D Maximum-Surface Techniques", *International Journal of Computer Vision*, vol.47, No.1/2/3, pp.99-117, May 2002