# Head Detection and Tracking for the Car Occupant's Pose Recognition

Jeong-Eom Lee[1], Yong-Guk Kim[3], Sang-Jun Kim[1,2], Min-Soo Jang[1], Seok-Joo Lee[4], Min Chul Park[5], and Gwi-Tae Park[1,*]

[1] Dept. of Electrical Engineering,
[2] Interdisciplinary Programs of Mechatronics,
Korea University, Seoul, Korea
*gtpark@korea.ac.krT
[3] School of Computer Engineering, Sejong University, Seoul, Korea
ykim@sejong.ac.krT
[4] Hyundai Autonet Co. Ltd., Korea
[5] Systems Technology Division, KIST, Korea

**Abstract.** This paper describes a Vision-based Occupant Pose Recognition (VOPR) system, which can ensure a safe airbag deployment. Head detection and its tracking are necessary for occupant's pose recognition in the car, since the position of occupant's head provides valuable information, such as his pose, size, position, and so on. We use the stereo cameras to extract a disparity map. Against variable lighting conditions including the night drive, we adopt infrared illumination as well as normal one. Results suggest that VOPR system is reliable and performs reasonably well.

## 1 Introduction

Airbags have played an important role for automotive safety. Although airbags have saved many lives, wrong deployment of them can lead to a fatal injury. Recently, National Highway Traffic Safety Administration (NHTSA) has amended Federal Motor Vehicle Safety Standard (FMVSS) No. 208, asking to install some smart airbags in the cars. This law aims to improve protection of the occupants for different height and weight, regardless of whether they use their seat belts, while minimizing the risk to infants, children, and other occupants of deaths and injuries caused by airbags [1]. In fact, risk to out-of-position occupants has been reduced, since various occupant sensing technologies are adopted into vehicle occupant protection systems by major automakers and suppliers [10]. But to improve safety inside the car, it is necessary to develop more reliable methods. In this study, we present a vision-based technology that enables to capture diverse information about occupant such as class, pose, distance from occupant to the dash board.

Several researchers have studied vision-based technology and proposed algorithms that can classify the occupant and control intensity of airbag deployment according to class, posture of occupant [6, 7, 8, 9]. In this paper, we particularly focus on head

M. Ali and R. Dapoigny (Eds.): IEA/AIE 2006, LNAI 4031, pp. 540 – 547, 2006.

detection and tracking method for the occupant's pose recognition to ensure safe airbag deployments. Detecting the head is relatively easier than other human body and its position provides valuable information about the occupant, such as pose, size, position, and so on. So, measuring the accurate position of occupant's head is critical for the smart airbag system. Our system for pose recognition consists of three parts: head detection, head tracking and head localization. The head detection is based on motion information and uses contour models and support vector machines (SVM) classifier [2]. The head tracking is based on edge information and use template matching. As the position of the detected or tracked head provides only 2-D information, we use 3-D disparity information extracted from the stereo images for head localization.

Section 2 of this paper provides more details of head detection and tracking method. And section 2 also describes how the position of the head on the image coordinates (2-D coordinate) is transformed into a camera coordinate (3-D coordinate). Result of experiments is described in section 3. Finally, we summarize our results and discuss the performance of the whole system in section 4.

## 2 Vision-Based Occupant Pose Recognition (VOPR) System

Our VOPR system consists of three parts: head detection, tracking and localization part. For the head detection and tracking in a car environment, several methods based on motion and color information are proposed [6, 7, 8, 9]. Considering diverse illumination condition within the car, it is difficult to take advantage of the color information. Moreover, since the infrared illumination is utilized to capture the occupant in the night, our method is based upon using only the grey image.

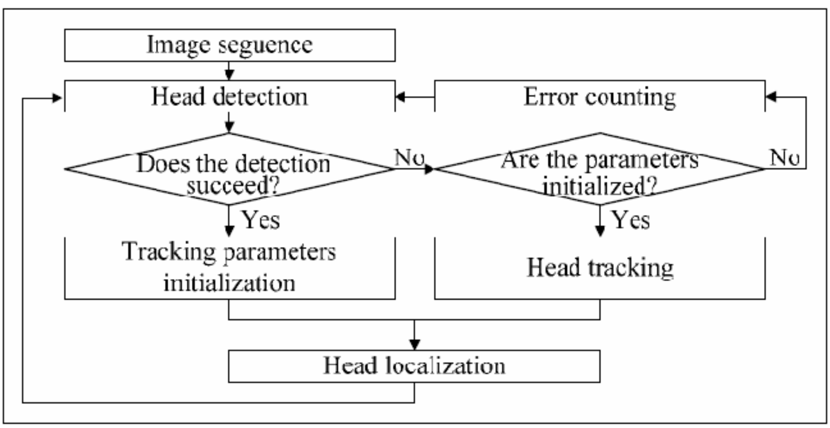


**Fig. 1.** A block diagram of the pose recognition system architecture

Difference images from the motion information are used for the head detection task and then, accurate head position is obtained by applying contour models and consequently an SVM classifier. However, when the occupant's motion is not significant, it is difficult to detect the head. On the other hand, head tracking works pretty well in such

case if parameters for head tracking have been initialized. Therefore, head detection and tracking should be used in complementary ways. **Fig. 1** shows a schematic diagram of our VOPR system architecture. In this figure, an error means that VOPR system is not initialized or it doesn't work. It is just prepared against the extremeness cases.

And the position of the detected or tracked head doesn't present a real position in the car since the detection and tracking are carried out only on 2-D image of a left charge coupled device (CCD) camera. By combining them with stereo information, we can obtain 3-D information of the occupant's head and then can control triggering and intensity of the air bag deployments.

The following sections describe the details of the head detection, head tracking and head localization.

## 2.1 Head Detection

This paper augments our previous work [15] which used head-shoulder contour model and support vector machines (SVM) classifier for head detection. In this previous paper, to get motion information, we first make an accumulative difference image of sequential images from the single camera. And a silhouette image is obtained after applying binary morphological operations. And then, feature points to consist of a contour are extracted from silhouette. These points are used as input for the SVM classifier. The public domain implementation of SVM, called LibSVM, was used for this study [3]. The head-shoulder contour model can be derived from the feature points and describe the shape of occupant's head, regardless of the orientation of the occupant's face. A procedure of head detection is depicted in **Fig. 2**.

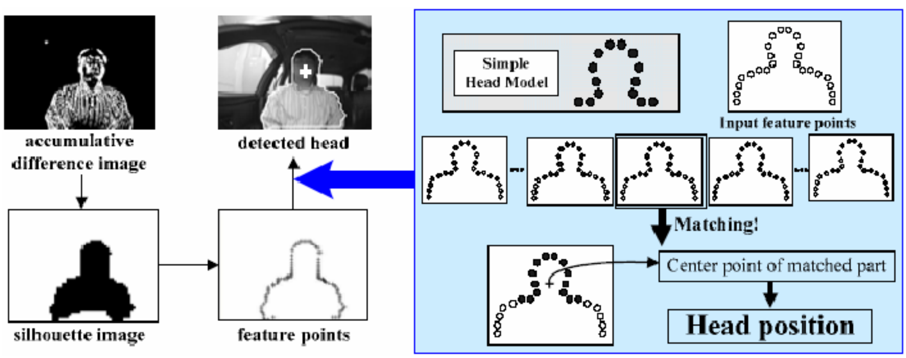


**Fig. 2.** A procedure of Head detection

## 2.2 Head Tracking

Our head tracking algorithm consists of two parts: initialization part and tracking part as shown in **Fig. 3**. In the first part, initial searching area for the template model is set up by using the position of the detected head. In the second part, a Ω shaped template is fitted to the head area and then updates the searching area.

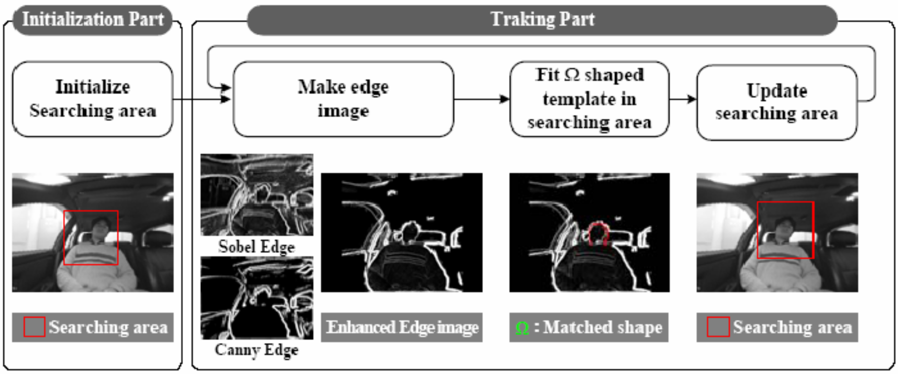


**Fig. 3.** A procedure of Head tracking

Here, the head tracking algorithm is based on edge method to improve computational efficiency. As shown **Fig. 3**, the edge detector is a combination of the *Sobel* [5] and the *Canny* operation [14]. In such way, we can obtain an edge image which contains the fine details around occupant's head and yet removes unwanted edge fragments. As a result, the enhanced edge image has only gray level which is similar to skin color.

The next step is to fit a model to the enhanced edge image. The model adopted here is a Ω-shaped model proposed in [12]. In fact, the Ω-shaped model is very similar to the head-shoulder contour model. The goal of such fitting is to find the most similar part by modifying scale and translation parameters. Namely, a matched template is adjusted to maximize a value between the Ω-shaped model and the edge image within the searching area. Such value can be calculated by using an equation:

$$f(I) = \sum_{x,y \in \mathbf{T}_O} |I(x,y)| + \sum_{x,y \in \mathbf{T}_I} |255 - I(x,y)| \tag{1}$$

$\mathbf{T}_O$ *: a set of pixels on template,* $\mathbf{T}_I$ *: a set of pixels on interior of template*

Here, *f(I)* is increased when pixels on a boundary of template are close to white value and pixels within the template are close to black value. After the matched template is found, a center of the searching area is changed into a center of the matched template and its size is updated in proportion to the size of the matched template.

## 2.3 Head Localization

In this section, we describe how real position of the occupant's head can be obtained from the detected (or tracked) head. As stated in section 2.1 and 2.2, the position of the detected head is the central point of head-shoulder contour model or Ω-shaped model. Although the head detection and tracking method gives us useful information on the present position of the occupant's head, it doesn't provide 3-D pose information of the head, since the detection and tracking are carried out only on 2-D image. In the present study, we aim to combine the head detection with stereo information to acquire the pose information of the occupant within the car.

An object seen from a pair of stereo cameras leads to a visual difference, which depends on the baseline length between two cameras as well as 3-D shape of the object. Disparity of an arbitrary point is defined as follow:

$$d = u - u' \tag{2}$$

It can be calculated using a stereo matching algorithm [11]. Since disparity provides 3-D information of the object [6], we can recognize the real position of occupant's head in a car. To display the location that the head belongs to at a certain moment, a camera coordinate can be marked on a top-view picture as illustrate on the right of **Fig. 4**, in which a circle designates the head location.

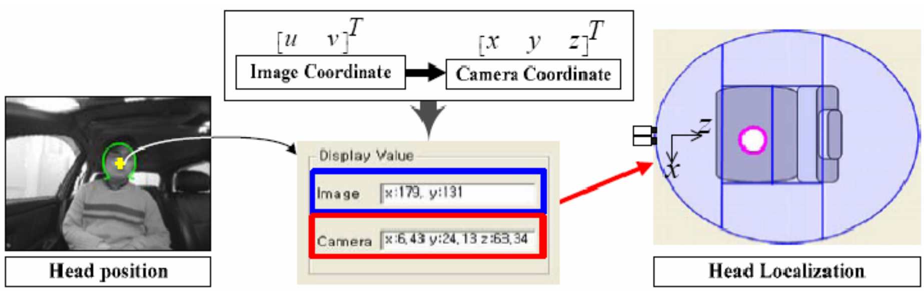


**Fig. 4.** Head localization

## 3 Results

### 3.1 Video Database for Test

The performance of the VOPR system is evaluated with our video database, which consists of sequential image sets captured within a car at 15fps (frame per second). **Table 1** shows the image database adopted in this study. In order to include the night drive case, we also took the infrared images for Image Sequence Ⅲ and Image Sequence Ⅳ.

**Table 1.** The video database (unit:frame)

| | Number of frame | Remark |
|---|---|---|
| Image Sequence Ⅰ | 235 | • |
| Image Sequence Ⅱ | 212 | • |
| Image Sequence Ⅲ | 122 | Infrared |
| Image Sequence Ⅳ | 196 | Infrared |
| Total | 765 | |

### 3.2 Performance of the Algorithm

In real-time demo, the head indicated as a circle drifts around different areas as the occupant move his head and body around the seat. However, the experiments were carried out with an off-line basis for four different video databases. We only counted it as a correct image frame when the system is able to assign the occupant's head position to the ground-truth among those 6 areas (see **Fig. 4**). Since the tolerance of the error was 7.5 cm that is radius of the typical human head, any image frames having an error larger than this tolerance are counted as the incorrect frames. **Table 2** summarizes the result of the analysis. **Fig. 5** shows graphs for the trial error, where the error is varied as the frame goes by time.

**Table 2.** Experimental Result (unit: frame)

| | Video frames | Correct frames | Incorrect frames | Success rate |
|---|---|---|---|---|
| Image Sequence Ⅰ | 235 | 218 | 17 | 92.8% |
| Image Sequence Ⅱ | 212 | 180 | 32 | 84.9% |
| Image Sequence Ⅲ | 122 | 109 | 13 | 89.3% |
| Image Sequence Ⅳ | 196 | 163 | 33 | 83.2% |
| Total | 765 | 670 | 95 | 87.6% |

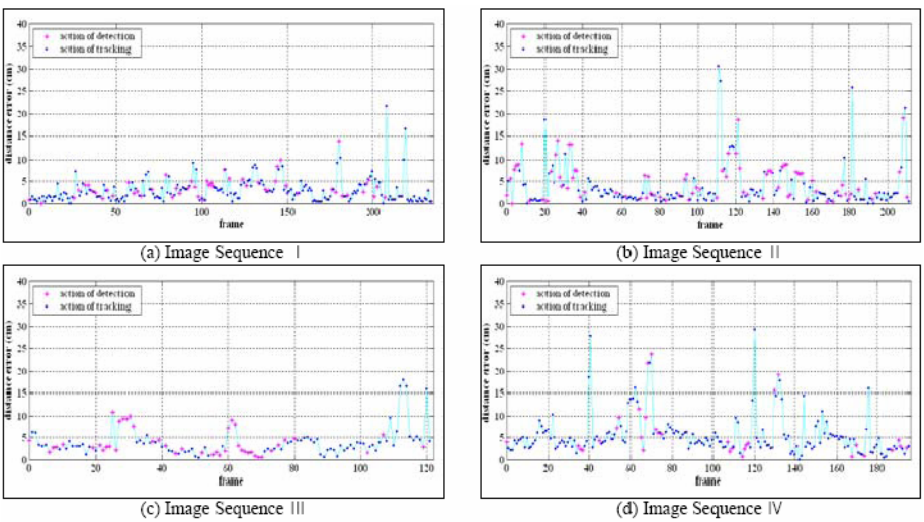


**Fig. 5.** Graphs for the trial errors

As showed in **Table 2**, the total success rate of our system was 87.6%. Spikes in **Fig. 5** indicate false positives that are happened by wrong detection. However, you can observe that such spikes are corrected within a few frames. Besides, occlusion is the cause of errors. Although detection or tracking is accurate in 2-D images, occlusion caused errors when the image coordinates (2-D coordinate) is transformed into the camera coordinate (3-D coordinate). When the occupant's head is located closely to camera, a few pixel differences in 2-D images bring about wide differences in camera coordinate. This is the reason why errors occurred from 20$^{th}$ to 40$^{th}$ frames for Image Sequence Ⅱ. Demonstration of the VOPR system is shown in **Fig. 6**. **Fig. 6**(a) shows some errors caused by occlusion. **Fig. 6**(b) shows ability to re-acquire the subject by head detection when the occupant returned to the camera's field of view. **Fig. 6**(c) is some examples of edge images obtained from input images.

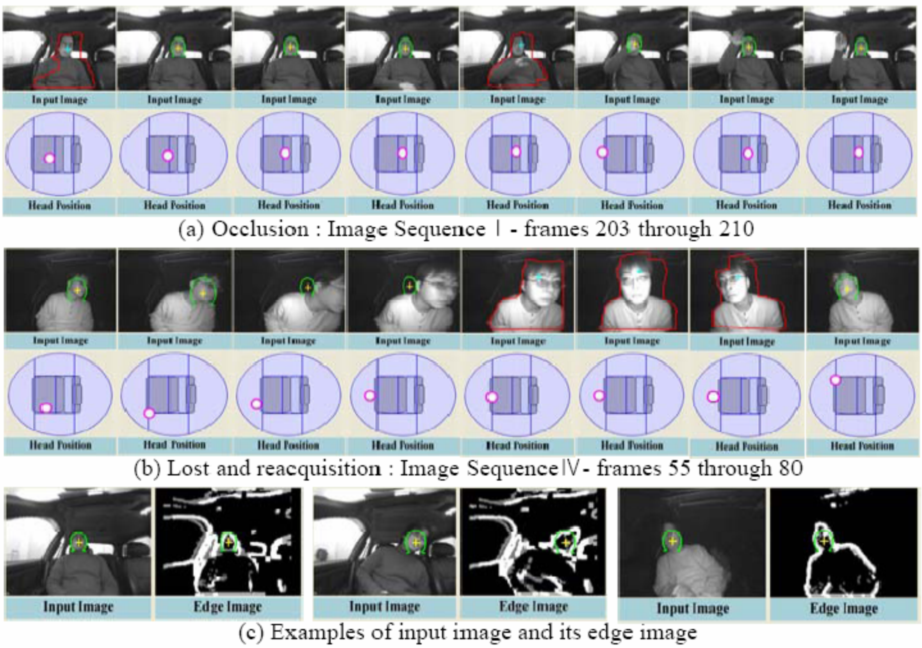


**Fig. 6.** Demonstration of the VOPR system's performance

## 4 Conclusions and Discussion

In this paper, we described a VOPR system that can be recognized the pose of the occupant. We use the head detection and tracking algorithms in complementary fashion to highlight strength and to cover shortcoming one another, and combine them with stereo information to obtain a camera coordinate (3-D coordinate) of the occupant's head position. The disparity of the occupant's head position is calculated from the stereo image captured using a stereo camera. Because of computation time, the stereo matching algorithm is performed in the determined region. We plan to carry out such intensive computation using a specialized hardware such as FPGA. Our VOPR system just consists of two CCD cameras and infrared illumination. In video databases for test, the VOPR system successfully found the occupant's head and maintained appropriate performance. Experimental result confirms that our system is feasible and the performance is satisfactory. In the future, we will work on adding prediction scheme to solve problems caused by occlusion and false positives.

## References


1. http://www.nhtsa.dot.gov
2. V. Vapnik, "The Nature of Statistical Learning Theory," Springer-Verlag, NY, USA, 1995
3. http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html
4. A. Blake and M. Isard, "Active Contours," Springer-Verlag, London, 1998
5. G. Baxes, "Digital Image Processing: principles and applications," John Wiley & Sons, New York, USA, 1994
6. Marcus Klomark, "Occupant Detection using computer vision," master's thesis, 2000
7. Y. Owechkp, N. Srinivasa, S. Medasani, and R. Boscolo, "Vision-Based Fusion System for Smart Airbag Applications," IEEE, Intelligent Vehicle Symposium, vol. 1, pp. 245-250, 2002
8. B. Alefs, M. Clabian, H. Bischof, W. Kropatsh, and F. Khairallah, "Robust Occupancy Detection from Stereo Images," IEEE Intelligent Transportation Systems Conference, 2004
9. M. M. Trivedi, S. Cheng, E. Childers, S. Krotosky, "Occupant Posture Analysis with Stereo and Thermal Infrared Video: Algorithms and Experimental Evaluation," IEEE Transactions on Vehicular Tech. Special Issue on In-Vehicle Vision Systems, Volume: 53, Issue: 6, 2004
10. L. Zhang, L. Chen, A. Vertiz, and R. Balci, "Survey of Front Passenger Posture Usage in Passenger Vehicles," 2004 SAE World Congress, 2004
11. C. Sun, "A Fast Stereo Matching Method," Digital Image Computing: Techniques and Applications, pp.95-100, 1997
12. R. Patil, P. E. Rybski, T. Kanade, and M. M. Veloso, "People Detection and Tracking in High Resolution Panoramic Video Mosaic," IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS 2004), vol. 1, pp. 1323-1328, 2004
13. S. Birchfield, "Elliptical Head Tracking Using Intensity Gradients and Color Histograms," IEEE Conference on Computer Vision and Pattern Recognition, pp. 232-237, June 1998
14. J. Canny, "A Computational Approach to Edge Detection," IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 8, pp. 679-698, 1986
15. YG Kim, JE Lee, SJ Kim, SM Choi, and GT Park, "Head Detection of the Car Occupant based on Contour Models and Support Vector Machines," LNAI, Vol. 3533, pp. 59-61, 2005