Driver's Head Detection Model in Color Image for Driver's Status Monitoring

Woong Jae Won, Man-Ho Kim, and Joon-Woo Son

Abstract—**In this paper, we proposed a simple driver's head detection model for driver's status monitoring, in order to implement interactive safety driving agent system. For intensifying the driver's head region, we consider normalized color feature combination method with skin color filter. And, the center-surround difference and normalization algorithms are applied for making saliency map of driver's head in order to nt only reduce the background noise, but also more reinforce the intensity of driver's head regions than non-head area. The driver's head regions are simply selected through searching local maximum energy with histogram projection method. And, we adopt the 2D+3D active appearance model(AAM) for tracking driver's head region and estimating driver's head pose. Experimental results show that proposed model can successfully detect driver's head region and estimate driver's head pose.**

I. Introduction

Recently the National Highway Traffic Safety Administration(NHTSA) reported that the 25~35% of the traffic accidents or 1.2 million vehicle crashes per year in the United State are resulted from distraction and inattention[1, 2]. And, the Treat et. al. reported that the human error mainly cause the vehicle crash[3]. Furthermore, drivers are faced to an increasing number of on-board functions and in a near future of the massive introduction of in-vehicle systems such as vehicle information, entertainment, and assistance systems(audio, navigation, mobile phone, lane departure warning system, and etc). On the other hand, the driver is not always in good situation to receive and understand the messages that are given to him by the vehicle/system. Major are his current status like tired, distraction, absent minded, and etc[4]. Therefore, it's known that more increasing number of these in-vehicle systems dramatically increase car accidents as resulting by decreasing driver's inattention.

Thus, the adaptive human machine interface(AHMI) technologies such as interactive workload manager and advanced assistance system have been received more attention from many intelligent vehicle research community to support and disburden driver to significantly increase driving safety and comfort[4, 5, 6]. In order to implement these systems driver's status monitoring technology has been important issues for research in computer vision research community [4 - 7].

Therefore, the driver's status monitoring research bas been tremendously conducted as much as an amount of its importance[4 - 9]. Even through those model have good performance, because those model still have troubled with influence of illumination change, image distortions, and affine transform problem in real situation, many researchers agree that the developed algorithms are still exists additional effort to develop a new model with better performance than currently developed one[7, 8, 9]. Thus, some researchers recently have proposed saliency map models based on biologically motivated selective attention, which are imitating human-like early visual processing, to overcome those problems[10, 11, 12]. However, because these model need much computational time, it's also hard to apply for the real-time system[10, 11, 12].

In this paper, we focus on simple and robust driver's head detection model to implement the driver's status monitoring system. In order to detect driver's head regions, we consider normalized color feature combination with skin filter to intensify driver's head regions and separate between head regions and background. In addition, we adopt the center surround difference and normalization (CSD&N) algorithms with Gaussian pyramid processing to reduce noise influence and illumination change, and more reinforce the intensity of driver's head region than non-head regions[10, 12, 13]. And, the driver's head regions are simply selected by local maximum searching with histogram projection method[12, 14]. Then, we apply 2D+3D active appearance model(AAM) for tracking driver's head region and estimating driver's head pose[15, 16].

This paper is organized as follows; Section 2 describes the proposed new color based driver's head detection model. The experimental results will be followed in Section 3. Section 4 presents our conclusions and discussion.

II. Driver's Head Detection Model

Fig. 1 shows the proposed simple driver's head detection model. After extracting the red(R), green(G), and blue(B) color feature from the input color image, the normalized red(r), green(g), blue(b), and yellow(y) color features are extracted form the R, G, and B plan. Then, the driver's face

This work was supported by the Daegu Gyeongbuk Institute of Science and Technology(DGIST) Basic Research Program of the MOST.

W. J. Won with Division of Advanced Industrial Science and Technology, Daegu Gyeongbuk Institute Science and Technology, 75, Gondanbu2gil, Dalseo-Gu, Taegu, 704-230, Korea (wwj@dgist.ac.kr)

M. H. Kim with Division of Advanced Industrial Science and Technology, Daegu Gyeongbuk Institute Science and Technology, 75, Gondanbu2gil, Dalseo-Gu, Taegu, 704-230, Korea (mhkim@dgist.ac.kr)

J. W. Son with Division of Advanced Industrial Science and Technology, Daegu Gyeongbuk Institute Science and Technology, 75, Gondanbu2gil, Dalseo-Gu, Taegu, 704-230, Korea (json@dgist.ac.kr)

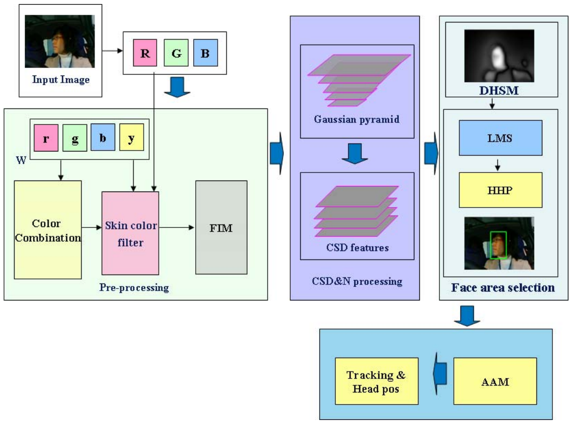

Fig. 1. The proposed driver's head detection model. R: red, G: green, B: blue, r, g, b, and y: normalized red, green, blue, and yellow color feature, W: weight, CSD&N: center-surround difference and normalized algorithms, DHSM: driver's head saliency map, LMS: local maximum searching, HHP: hybrid histogram projection function, AAM: active appearance model

intensified map(FIM) is extracted from combination of normalized color feature with skin color filter. The driver's head saliency map(DHSM) is constructed through CSD&N algorithms with Gaussian pyramid images. After constructing the DHSM, the candidate driver's head regions are simply selected by local search for maximum energy. Then, the driver's head areas are getting from hybrid histogram projection and variance histogram projection method. And, the driver's head regions are decided by constrain of the size and local energy threshold. Finally, the 2D+3D AAM is used for tracking driver's head region and estimating driver's head pose.

A. Extraction of Driver's Head Saliency Map

In order to make the DHSM, we consider the normalized color features which is known to effect on reducing influence of luminance like human visual system do[10, 12, 17]. For extracting the normalized r, g, b, and y, the R, G, and B color features are extracted from the input color image. Then, the r, g, b, and y color features are extracted by Eq. (1).

$$r=\begin{cases} R-\dfrac{G+B}{2}, & r>0 \\ 0 & ,\ r\le 0 \end{cases},\quad g=\begin{cases} G-\dfrac{R+B}{2}, & g>0 \\ 0 & ,\ g\le 0 \end{cases}$$
$$b=\begin{cases} R-\dfrac{G+B}{2}, & r>0 \\ 0 & ,\ r\le 0 \end{cases} \qquad (1)$$
$$y=\begin{cases} \dfrac{R+G}{2}-B-\dfrac{|R-G|}{2}, & y>0 \\ 0 & ,\ y\le 0 \end{cases},$$

Even through the skin color intensity is finer in r and y feature channel then g and b features, the skin color intensity is weaken in each normalized color features then R, G, and B color features. In order to intensify the skin color regions, we consider weighted color combination method as reflecting these characteristic. Therefore, the weighted normalized color combination map(WNCCM) is constructed by Eq. (2).

$$WNCCM=\begin{cases} (r+2\times y)-3\times(g+b), & 0<WNCCM<255 \\ 0 & ,WNCCM<0 \\ 255 & ,WNCCM>255 \end{cases} \qquad (2)$$

Then, WNCCM is filtered for constructing face color intensified map(FIM) following the processing of skin color filter using Eq. (3).

$$\begin{aligned} & if\,((r_{mean}-\alpha\bullet\sigma_r^2\le r<r_{mean}+\alpha\bullet\sigma_r^2)\,\&\& \\ & \quad (g_{mean}-\beta\bullet\sigma_g^2\le g<g_{mean}-\beta\bullet\sigma_g^2)\,\&\& \\ & \quad (y_{mean}-\gamma\bullet\sigma_y^2\le b<y_{mean}-\gamma\bullet\sigma_y^2)\,\&\& \\ & \quad (b<b_{\max})) \\ & \ then \quad FIM=WNCCM \\ & else \quad FIM=0 \end{aligned} \qquad (3)$$

where r_{mean}, g_{mean}, and y_{mean} are the mean value of r, g, and y which are trained from the sample images. And, the b_{max} is the maximum intensity of b among the simple color range data. σ_r^2, σ_g^2, and σ_y^2 are the standard deviation of r, g, y.

In order to reduce influence of noise and illumination change in various scenes, and more reinforce the head region than non-head regions, we adopt the on-center and off-surround operation by the Gaussian pyramid images which are consisted of different scales for 0 to n^{th} level where by each level is made by the sub-sampling of 2^n [12, 13]. Then, the center-surround features are constructed by the difference operation between the fine and coarse scales in the Gaussian pyramid images. Consequently, the FIM center-surround features can be obtained by Eq. (4) [10, 12].

$$FIM(c,s)=|FIM(c)-FIM(s)| \qquad (4)$$

where the "c" is the finer scale pyramid image, the "s" is the coarse pyramid image, and "-" represents the interpolation to finer scale and point-by-point substraction.

For obtaining the DHSM, we extract the 7 different scales of Gaussian pyramid features from 2^{nd} level to 6^{th} level for the CSD&N algorithms. The selected Gaussian pyramid features are combined into DHSM using Eq. (5) [10, 12].

$$DHSM=\bigoplus_{c=2}^{3}\bigoplus_{s=c+2}^{3}N(FIM(c,s)) \qquad (5)$$

Fig. 2 shows the procedure of extracting the DHSM. As shown the Fig. 2, the face area intensity is more intensified through color combination. And, FIM is constructed form the

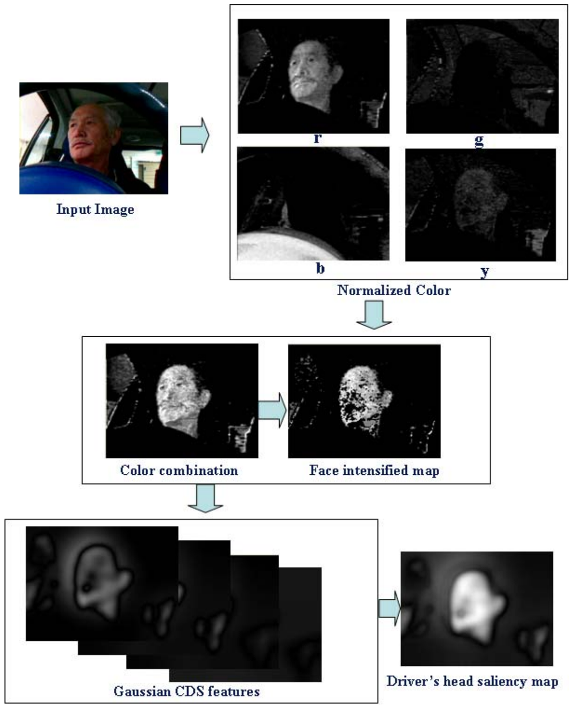

Fig. 2. The procedure of extracting driver's head saliency map.

color combination map with skin color filter. The 4 layers of CSD features for FIM are constructed from selected 5th layer of Gaussian pyramid features. Finally, the DHSM is constructed through the sum and normalized of CSD features.

B. Detection of Driver's Head Regions

Since the DHSM intensify the driver's head regions and diminish complex background, we can simply select a candidate driver's head regions(DHA) through searching the local maximum energy with fixed window by shifting the pixels using Eq. (6) in the DHSM.

$$DHA_k(x,y) = \arg\max_k \left(\sum_{i=-win/2}^{i=win/2} \sum_{j=-win/2}^{j=win/2} DHSM(x+i, y+j) \right) \tag{6}$$

where k is the number of candidate face region, and *win* is the size of the search window.

For obtaining the proper scale of driver's head regions, the function of vertical integral, variance, and hybrid histogram projection function (*VIHP'(x), VVHP(x), VHHP(x)*) are calculated for the DHSM using Eq. (7) - (9) at first.

$$VIHP'(x) = \frac{1}{d_y} \sum_{y=0}^{y=d_y} [DHSM(x, y_i)] \tag{7}$$

$$VVHP(x) = \frac{1}{d_y} \sum_{y=0}^{y=d_y} [DHSM(x, y_i) - VIHP'(x)] \tag{8}$$

$$VHHP(x) = \alpha \bullet VVHP + (1-\alpha) VIHP' \tag{9}$$

where d is image size and α present the weight factor for *VHHP*

Then, the proper horizontal scale of driver's head region is selected through searching left and right local minimum point with threshold of magnitude of *VHHP* from the center point of candidate driver's head area(DHA_k)

After selecting the regions of horizontal driver's head, the function of horizontal integral, variance, and hybrid histogram function (*HIHP'(x), HVHP(x), HHHP(x)*) are calculated from Eq. (10) - (12) within the selected the horizontal driver's head region.

$$HIHP'(y) = \frac{1}{s_x} \sum_{x=left}^{x=right} [DSHM(x_i, y)] \tag{10}$$

$$HVHP(y) = \frac{1}{s_x} \sum_{x=left}^{x=right} [DHSM(x_i, y) - IHHP'(x)] \tag{11}$$

$$HHHP = \beta \bullet HVHP + (1-\beta) IHHP' \tag{12}$$

Where s_x present the horizontal size of selected driver's head, and β is the weight factor for *HHHP*.

And, the proper vertical scale of driver's head region is selected through searching local minimum point with threshold of magnitude of *HHHP* from DHA_k

As shown the Fig. 3, the proposed model can simply select the driver's head regions through window based local energy maximum searching. And, the horizontal proper scale of driver's head area is obtained from hybrid vertical projection function. Then, the vertical proper scale of driver's head area is getting from hybrid horizontal projection function.

C. Tracking Driver's Head and Estimating Driver's Head Pose

For the tracking driver's head and estimating driver's head pose, we consider 2D+3D active appearance model(AAM)

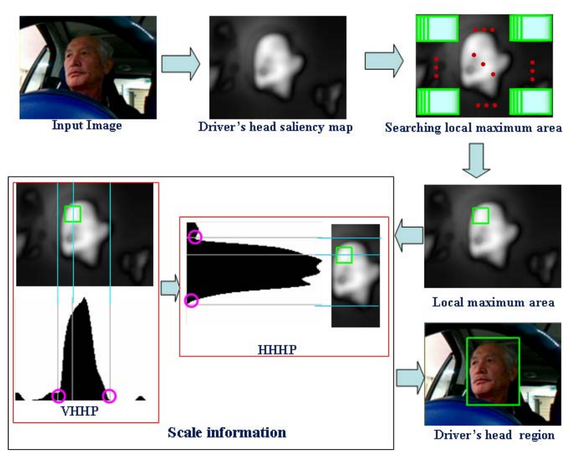

Fig. 3. The procedure in experimental result for detection of driver's head region.

[15, 16].

AAM is a stochastic model which can be learned from large sample data. Therefore, we were manually marking the predefined 66 facial shape feature points and triangular mesh to make facial shape and texture database for 2D AAM on the sample images. Then, the marked shape points are represented as one dimensional vector using Eq. (13).

$$A=\begin{bmatrix} s_1 & s_2 & \cdots & s_N \end{bmatrix},\\ s_i=\begin{bmatrix} x_i^1 & y_i^1 & x_i^2 & y_i^2 & \ldots & x_i^{66} & y_i^{66} \end{bmatrix} \qquad (13)$$

where A is the shape data matrix, s_i is the i^{th} shape data vector, and N is the number of samples

And, the facial shape(s) and texture(t) are represented as the linear summation of mean and principal components using Eq. (14) - (15) for large sample data of facial feature points[15].

$$s=m_s+\sum_i \alpha_i \hat{\mathbf{s}}_i \qquad (14)$$

$$t=m_t+\sum_i \beta_i \hat{\mathbf{t}}_i \qquad (15)$$

where m_s and m_t are mean shape and mean texture vector which are calculated from Eq. (16) – (17). α_i and β_i is coefficient of shape ($\hat{\mathbf{s}}_i$) and texture ($\hat{\mathbf{t}}_i$) bases vector which are obtained from principle component analysis(PCA) [15].

$$m_s=\frac{1}{N}\sum_1^N s_i \qquad (16)$$

$$B=\begin{bmatrix} t_1 & t_2 & \ldots & t_N \end{bmatrix},\\ t_i=\begin{bmatrix} R_i^1 & G_i^1 & B_i^1 & R_i^2 & \ldots & G_i^n & B_i^n \end{bmatrix},\\ m_t=\frac{1}{N}\sum_1^N t_i \qquad (17)$$

where m_s mean shape vector, B is the texture data matrix, t_i is the i^{th} texture data vector, N is the number of pixels, and m_t is the mean texture vector.

As shown the Fig. 4, after making the facial shape feature points manually, the facial feature models are obtained from PCA with similarity transform. Then, the facial textures are warping to mean shape to make the facial texture models. Finally, the facial texture models are getting from PCA[15].

Because 2D AAM is hard to obtain the 3D head pose, we consider additional 3D shape model, which can be generated by the 3D reconstruction method[16]. By using this model, 3D shape data could be recovered from 2D shape data. Then, the 3D shape model which consists of 3D mean shape and principle component are constructed by the PCA[16]. As the

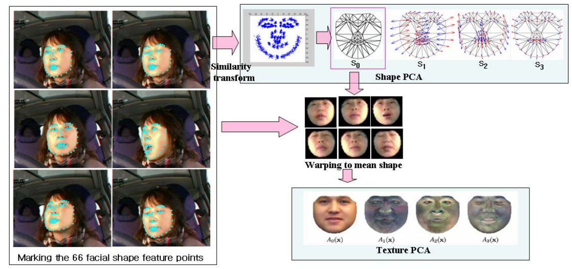

Fig. 4. The procedure of 2D AAM construction for shape and texture model (the shape PCA and texture PCA are obtained from Matthews and Backer[15]).

2D shape case, we assume that any 3D mean shape can be represented as a linear summation of 3D mean shape and bases like Eq. (18).

$$s_{3D}=m_{3D}+\sum \gamma_i \hat{\mathbf{s}}_i \qquad (18)$$

Finding the 3D head pose is what parameters are minimizing the squares error which is the summation of two squares errors, one is the error between the image produced by 2D texture model and 2D input image, the other is the error between 2D shape and the projected 2D shape of 3D shape [16].

Fig. 5 show the procedure of active appearance model for tracking and estimating driver's head pose. After selecting the driver's head area, the head feature points are initialized. Then the driver's head is tracked and driver's head pose is estimated from AAM processing.

III. Experimental Results

For the experiment, we installed Logitech's Notebook-Pro Cam at Mercedes-Benz's Smart Car based simulator which was developed to analyze driving characteristic of older-drivers in our laboratory. And, we also installed lighting control system at the right and left area for the reflecting the influence of lighting change while driving as shown Fig. 6. The forty-eight participants are participated this experiment. And, we took a video for making data-base of driver's head in five lighting condition which are turned-on each left and right lighting controller, right and left lighting

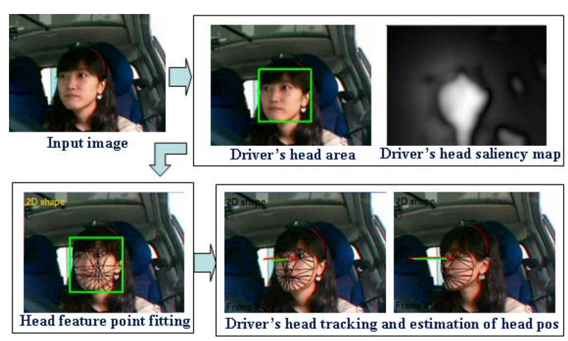

Fig. 5. The procedure of active appearance model for tracking driver's head and estimating driver's head pose.

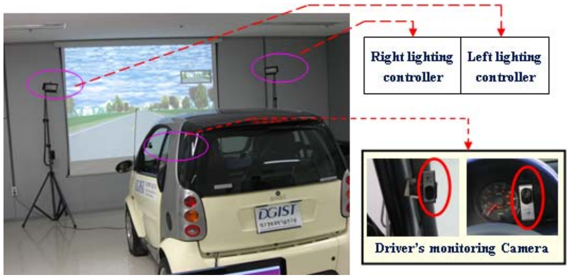

Fig. 6. The Mercedes-Benz's smart car based simulator system.

controller together, and turn-off the right and left lighting controller together, and only screen condition.

And, we captured 1,552 images from 48 video data-bases for the computer simulation experiment. Among 1,552 images, the 128 images are used as a training data for skin color filter, and the remained images are used as test data. Then, we select the 593 samples of driver skin regions, which are segmented by hands, and the each r, g, and y mean value of each 593 samples regions used for one-sample data. Therefore, we could obtain mean and standard deviation of each r, g, and y component which are 100.9 for r_{mean} and 36.6 for σ_r^2, 9.36 for g_{mean} and 8.27 for σ_g^2, 72.08 for y_{mean} and 26.62 for σ_y^2. Furthermore, we could get 3.06 for b_{max}. Then, we set α, β, and γ to 2 for skin color filter which is obtained from the distribution of each sample r, g, and b color intensity. Also, we set the fixed window size as 17x17, which is obtained from the experiment, for searching local maximum energy to select driver's head area.

In computational experiment, the driver's head detection accuracy rate of proposed driver's head model is 98.66% for 1, 242 test images. Moreover, the proposed face detection model can perform within 0.098 sec at Genuine Intel CPU T2400-1.83Ghz.

And, in order to making the training data of shape and texture for AAM, we also captured 3, 150 images, which include varying head poses and illumination conditions, for 21 video data-base. And, we marked 66 shape feature points on the images manually. Then, we construct mean shape, 8 shape bases, mean texture, and 10 texture bases.

Fig. 7 shows the experimental results of the proposed driver's head detection model. In here, normalized r, g, b, and y are considered as input features for color combination map. Then, the FIM is extracted by skin color filtering process for eliminating background. And, the DSHM is consisted through CSD&N. And, the DHSM is consisted through CSD&N. Then, the candidate region of driver's head region is selected through searching local maximum point with histogram projection method. The proposed model presents larger salient value in driver's head area than another area. Moreover, we can get a proper scale of driver's head regions in the DHSM. And, Fig. 8 show some computer simulation results of proposed driver's head detection model in our simulator study with various lighting condition.

Fig. 9 shows the experimental results of considered 2D+3

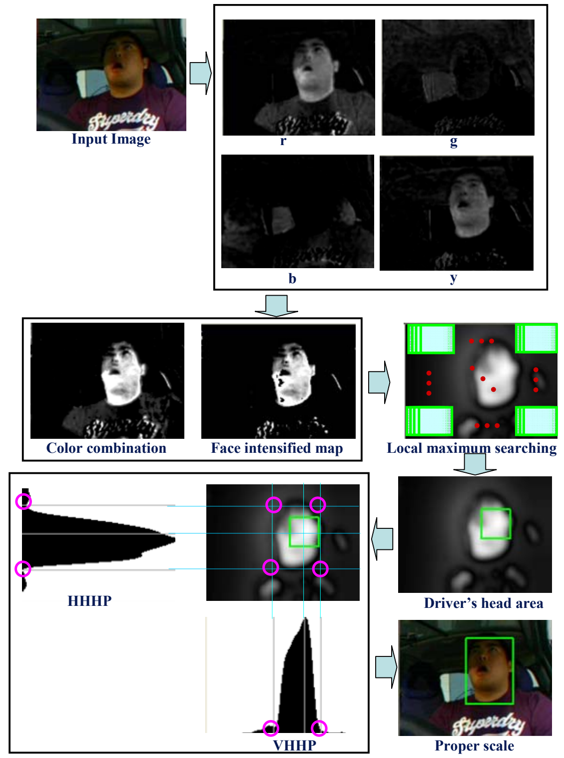

Fig. 7. Overall procedure in experimental results for the proposed driver's head detection model.

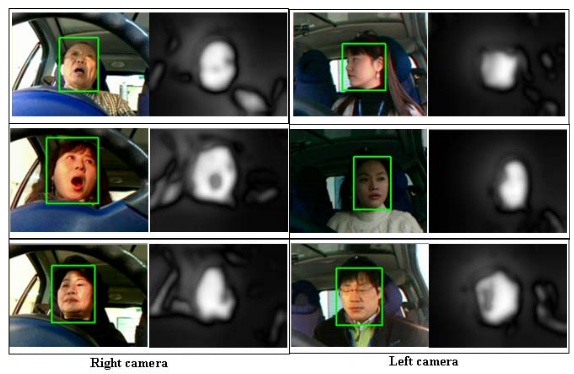

Fig. 8. The results of proposed driver's head detection model.

AAM. After obtained driver's head region, the driver's head feature points are initialized for tracking and estimation of driver head pose. Then, the driver's head and driver's head pose are tracked and estimated by 2D+3D AAM.

IV. Conclusion

We proposed a new driver's head detection model for monitoring driver's status. In order to more intensify driver's head regions than background and reduce background noise, we consider CDS&N algorithms with skin color filter. And, we can simply select the candidate region of driver's head region through searching the local maximum energy with vertical and horizontal histogram projection method. Moreover, we adopt 2D+3D AAM for tracking of driver's head and estimating driver's head pose. In computational and

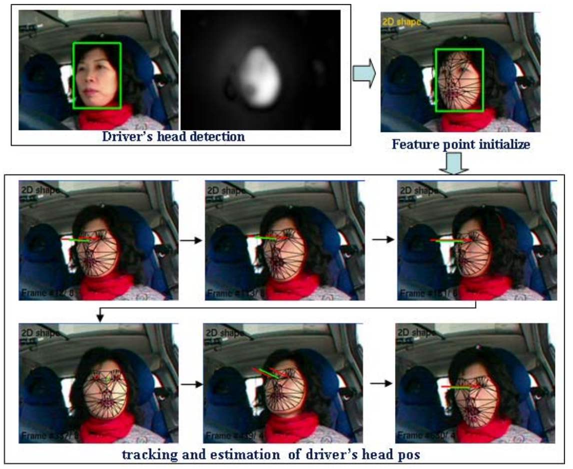

Fig. 9. The experimental results of tracking and estimating driver's head pose.

real experiment, the proposed model can successfully not only detect and track driver's head regions, but also estimate driver's head pose.

As further works, we are considering simple and enhanced method for getting proper scale of driver's regions. Also, we'll consider estimating and analyzing method of driver's gaze and facial movement for monitoring of driver's status.

ACKNOWLEDGMENT

This work was supported by the Deagu Gyeongbuk Institute of Science and Technology(DGIST) Basic Research Program of the MOST.

REFERENCES

[1] L. R. Shelton, "Statement before the subcommittee on highways and transit, committee on transportation and infrastructure", U. S. house of representatives, May 9. 2001. Available: http://www.nhtsa.dot.gov/nhtsa/announce/.

[2] J. C. Stutts, D. W. Reinfurt, L. Staplin, and E. A. Rodgman, "The role driver distraction in traffic crashes," Report Prepared for AAA Foundation for Traffic Safety, June 10, 2003.

[3] J. R. Treat, N. S. Tumbas, S. T. McDonald, and etc., "Tri-Level study of the Causes of Traffic Accidents: Final Report-Executive Summary," Department of Transportation, National Highway Traffic Safety Administration, Washington D. C, Tech. Rep. DOT HS805 099, 1979.

[4] T. V. Helen, B. Thirry, B. Serge, and etc., "Development of a Driver Situation Assessment Module in the AIDE Project," *in Proc. IFAC 16th World Congress*, Czech Republic, 2005.

[5] P. Green, "Driver Distraction, Telematics Design, and Workload Managers Safety Issues and Solution," i*n Proc. The 2004 International Congress on Transportation Electronics*, Warrendale, 2004, pp. 165-180.

[6] SAVE_IT, Safety Vehicle(s) using Adaptive Interface Technology (SAVE_IT) Program, Volpe National Transportation Systems Center. Available: http://www.volpe.dot.gov/hf/roadway/saveit/.

[7] P. Smith, M. Shah, and N. V. Lobo, "Determining Driver Visual Attention With One Camera," *IEEE Trans. Intel. Transportation Systems*, vol.4, no.4, pp.205-218, 2003.

[8] R. L. Hsu, M. A. Mottableb, and A. K. Jain, "Face detection in Color Images," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, No. 5, pp.696-706, 2002.

[9] P. Viola, and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol.57, pp.137-154, 2004

[10] L. Itti, C. Kock, "Computational Modeling of Visual Attention," *Nature Reviews Neuroscience*, vol.2, no.3, pp.194-203, 2002.

[11] C. Siagian and L. Itti, "Biologically-Inspired Face Detection: Non-Brute-Force-Search Approach," *in Proc. CVPRW'04*, Washington, DC USA, 2004, pp.62-69.

[12] W. J. Won and J. Y. Yeo, S. W. Ban, and M. Lee, "Biologically Motivated Incremental Object Perception Based on Selective Attention," *International Journal Pattern Recognition and Artificial Intelligence*, vol.21, no.8, pp.1293-1305, 2007.

[13] J. M. Ogden, E. H. Adelson, J. R. Bergen, and P. J. Burt, "Pyramid Methods in Image Processing," *RAC Engineer*, vol. 30, no.5, pp.33-41, 1981.

[14] Z. H. Zhou, and X. Geng, "Projection Functions for Eye detection," *Pattern Recognition*, vol.37, pp.1049-1056, 2004.

[15] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol.60, no.2, pp.135-164, 2004.

[16] I. Matthews, J Xiao, and S. Baker, "On the Dimensionality of Deformable Face Models," Tech. Report CMI-RI-TR-06-12, Robotics Institute, Carnegie Mellon University, 2006.

[17] E. B. Goldstein, Sensation and Perception 4th ed. An international Thomson publishing company, USA, 1996.