A Computer Vision System for Automatic Analysis of the Cephalo-Ocular Behaviour of Drivers

Frédéric Ntawiniga[1], **Denis Laurendeau**[1], **Normand Teasdale**[2], **Martin Simoneau**[2]

[1]*Dept. of Electrical and Computer Engineering* [2]*Social and Preventive Medicine – Kinesiology*
Laval University, Quebec City, QC, Canada G1K 7P4
[ntawini,denis.laurendeau]@gel.ulaval.ca [normand.teasdale,martin.simoneau]@kin.ulaval.ca

Abstract: A computer vision system for automatic analysis of the cephalo-ocular behaviour of drivers is presented. Driving is an important activity for a large percentage of the population. In Canada, one out of five drivers is over 65 years old. In the US, there are indications that 50% of older drivers suffering from mild dementia still continue to drive for several years. In this context, our group is currently developing an intelligent system composed of a driving simulator offering a rich collection of scenarios for assessing driving abilities combined with a computer vision system for extracting information on cephalo-ocular behaviour and visual search patterns of drivers. This paper describes the head motion tracking algorithm in 3D space, which is an important component of the computer vision system that is responsible for estimating the driver's head pose as a function of time.

Keywords: 3D head tracking, detection, tracking, stereo, matching, 3D reconstruction, monochrome, Kalman filter, neural network.

1. INTRODUCTION

Various techniques for tracking head pose in 3D space have been proposed. This demonstrates that head tracking is an active topic in computer vision research. The systems based on these techniques have a great range of applications such as virtual reality, entertainment, physiological sciences studies, transportation and computer-human interaction.

The majority of head tracking systems can be classified in two categories: systems that rely mostly on hardware such as wearable devices installed on the subject's head [8-11] and systems with a different approach of using mostly software combined with one or more video cameras [1-7]. Although the systems in the first category may be very accurate, they are often cumbersome for the user and/or very costly.

To extract information on cephalo-ocular behaviour of car drivers, the pose of the head in 3D space is needed. In computer vision, the use of two cameras and stereo vision allows the pose of the head to be recovered from synchronized video images. Not all of the systems based on computer vision use stereo vision. For instance, some track the head in 2D space [5-7]. Other systems use 3D models for tracking [3-4]. However, stereo is still needed for achieving better precision since the use of static head models is limited because head size varies with every subject. In [2] stereo is used with a colour camera. The system presented in this paper uses monochrome cameras because infrared illumination is also used in the environment as explained below. Gorodnichy[1] uses stereo combined with Epipolar Geometry for increasing tracking robustness but the automatic detection of facial features is not considered.

The system proposed in this paper offers the advantages enumerated above and, in addition, can be used in real time. It combines the following processing stages: image acquisition, head detection, facial feature extraction using Epipolar geometry, facial feature detection using Neural networks and Kalman filtering, and, finally, 3D reconstruction of the facial features that are being tracked as shown in Fig. 1. The validation of the results is achieved by running our algorithms on pre-recorded video sequences in order to obtain 3D pose estimates to be compared with the 3D measurements provided by a motion tracking device installed on the driver's head.

This paper is organized as follows: section 2 describes the Image acquisition system, section 3 describes the Head detection system, section 4 presents the facial features extraction and detection stages. Section 5 describes the 3D reconstruction of the head pose while section 6 presents the execution strategy of the different processing stages. Section 7 presents experimental results. The conclusion and future work are discussed in section 8.

2. IMAGE ACQUISITION

For image processing to provide good performance, the quality of input images at the image acquisition stage is very important [12]; every piece of

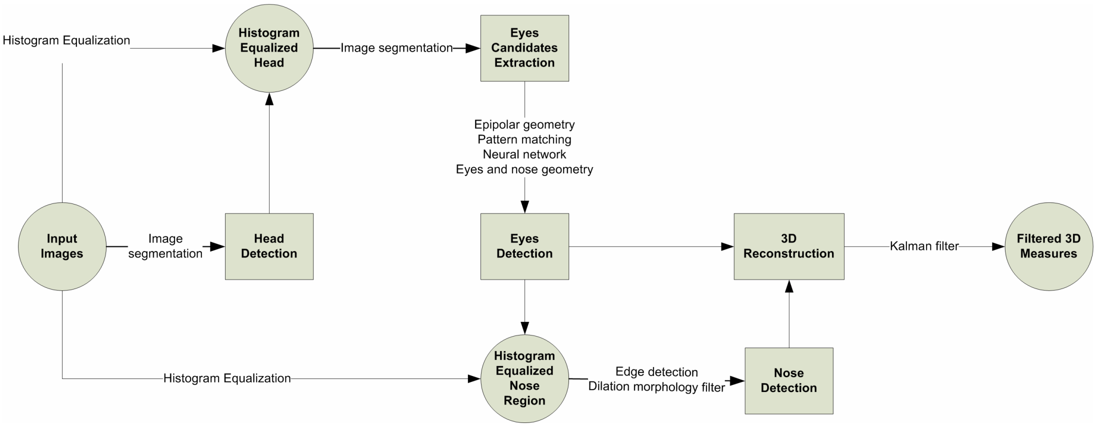

Fig. 1. System Block Diagram

information that is lost at this stage is difficult to compensate for at subsequent stages of the processing chain. To achieve this goal many concepts such as controlled lighting, light spectrum, colour space selection, cameras position, acquisition rate (FPS) and camera synchronization must be taken into consideration during system design.

The driving simulator environment shown in Fig. 2 is illuminated with infrared lights. Two synchronized monochrome video cameras operated at 30 FPS are used to allow stereoscopic capture. Infrared spots are used to illuminate the simulator environment because visible light would lower the quality of immersion by “bleaching” the scene that is being projected onto the screen.

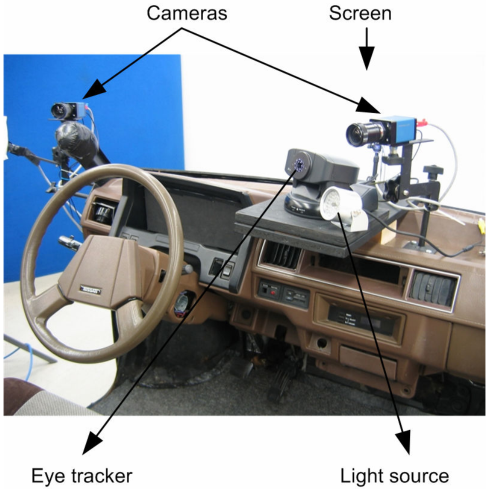

Fig. 2. Simulator Cockpit

3. HEAD DETECTION

In our system, head detection amounts to foreground detection. As the system is in a controlled lighting environment with a dark background, head detection is implemented so as to reduce the area of the image where facial features used for tracking are searched for. Head detection is a three-step process:

. Image thresholding and binarization (for both the left and right image).
. Detection of the largest blob in the resulting binary image.
. Assignment of a rectangle (defined as the “head bounding box”) to the top part of the largest blob.

The head detection step is followed by the detection and extraction of facial features. The two eyes and the tip of the nose were found mostly invariant to the head movement, rotation and facial expression and were thus selected as reliable facial features used for tracking [1], [16].

4. FACIAL FEATURES EXTRACTION AND DETECTION

When two stereo images of a driver are observed, only one eye can be easily detected by applying image processing techniques. Most of the time, the other eye is in contact with the background which makes segmentation difficult. To overcome this, the epipolar geometry of the stereo pair is exploited. Two images of the same scene are related by the epipolar geometry: an object in the first image has a match in the second image on the line called the epipolar line. Since the eyes that are easy to segment are always located on the opposite side in the stereo images (Fig. 3), they are detected using image processing algorithms described below and the epipolar geometry is used to locate them in the other image (close to the background).

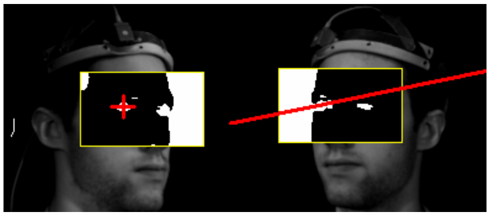

Fig. 3. Simulator Cockpit

The potential “eye” candidates are extracted as follows:

. the histogram of the image is equalized.
. the resulting image is thresholded and binarized.

- the binarized images are segmented into connected blobs.
- blobs with an area larger than a given value are detected and have to match an “average eye pattern” to be considered as “eye candidates”.

As a result two sets of eye candidates are obtained for every image.

The eye detection stage follows in 5 steps:

- A corresponding epipolar line for every eye candidate in both images is obtained.
- A candidate that matches a candidate in another image on the corresponding epipolar line is detected by using pattern matching, the pattern being a small window extracted from the input images.
- A test is performed to check if the match found is in the candidates set of the second image.
- A test is performed to make sure that the candidate is not outside of the head bounding boxes in both images.
- A validation is done on the results with a neural network using the distance information between the two eyes in every image.

The neural network validates whether the distance between the two eyes is realistic or not based on prior training by a set of sequences. After the detection of the two eyes, the brightest pixel at the tip of the nose is selected as a third facial feature [16].
The next step consists in forming a triangle (in 3D) composed of the position of the two eyes and the tip of the nose.

5. 3D RECONSTRUCTION

The 3D reconstruction is achieved by estimating the pose of the triangle formed by the two eyes and the tip of the nose in 3D space. The coordinates of the computed 3D points are filtered with a Kalman filter to improve smoothness of pose with respect to time. A 3D head model in OPENGL is superimposed to the 3D triangle for visualization as shown in Fig. 4.

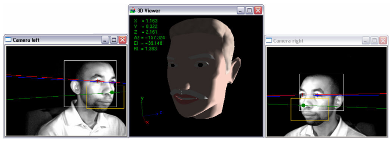

Fig. 4. 3D Model

6. RUNNING MODES

The system is optimized in order to avoid the search for facial features in each frame. In the detection mode all of the above stages above are implemented. Then, once the first pose estimates become available, the processing of the whole blob corresponding to the head is avoided by using previous 2D location of the eyes in the stereo images as seed regions for detecting the eyes in the next frames. This strategy is shown in Fig. 5.

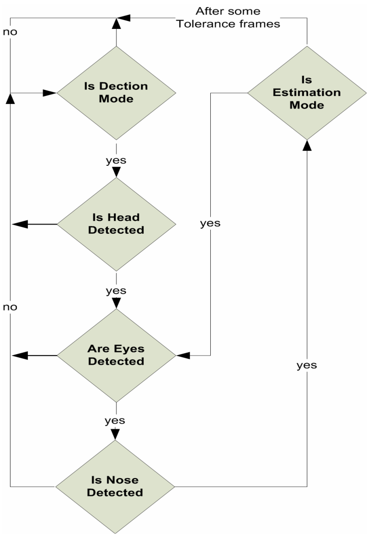

Fig. 5. Processing strategy

7. EXPERIMENTAL RESULTS

To validate the performance of the system, we calculate the average error and the maximum error between the estimated pose with the one obtained with a magnetic head tracker. The average error (1) is defined as an average distance between the estimated and measured pose for video frames for which stereo matching is achieved. The maximum error (2) is the error that represents the worst case pose in video sequences.

$$erraver = \frac{\sum_i |a_i - \tilde{a}_i|}{N} \qquad (1)$$

$$errmax = \max_i |a_i - \tilde{a}_i| \qquad (2)$$

Where a_i is the measured pose at time i and $\tilde{a}_i$ the estimated pose, N being the number of video frames. All results as shown in Fig. 6 are presented in Tab. 1, and are comparable to other video sequences.

Tab. 1. Translation and rotation errors

	Average Error	**Maximum Error**
X position	4.41mm	19.36mm
Y position	2.63mm	13.40mm
Z position	7.33mm	25.55mm
Roll	-0.98°	24.81 °
Elevation	2.59 °	12.32 °
Azimuth	0.23 °	10.13 °

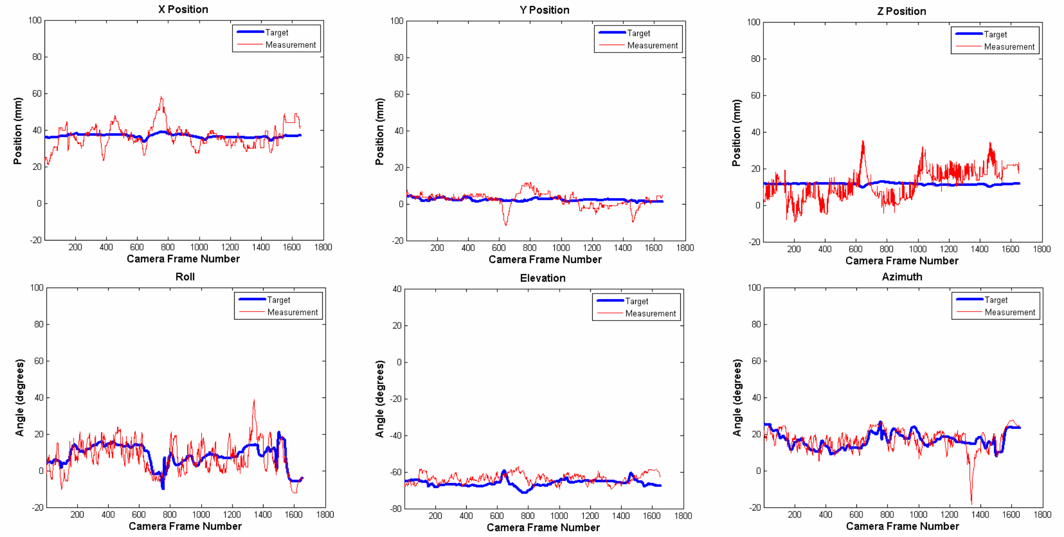

Fig. 6. Experimental Results

The orientation parameters are very close to the exact value while position parameters are satisfactory. The observed offset is due to the delay in the response of the Kalman Filter.

8. CONCLUSION AND FUTURE WORK

A computer vision based head motion tracker in 3D space was developed to help analyze the behaviour of drivers in a driving simulator.

The tracking system is made robust by using epipolar geometry that helps to cover a wide range of head rotation. The system also runs in real time at a frame rate of 13 FPS on an off-the-shelf 1.6 Ghz Centrino Duo processor.

Future work will consist in extending the algorithms to drivers wearing glasses. We also intend to install the tracking system on board a real car. In this case, the background subtraction algorithms will need to be revisited in order to be able to cope with varying background conditions, finally the neural network training will be done automatically to facilitate 2D tracking results whenever the cameras are moved.

REFERENCES

[1] D. Gorodnichy, S. Malik, and G. Roth. *Affordable 3D face tracking using projective vision*, In Proc. Intern. Conf. on Vision Interface (VI'2002), Calgary, May 2002.

[2] K. Terada,A. Oba, A. Ito, *3D human head tracking using hypothesized polygon model*, IEEE International Conference on Systems, Man and Cybernetics, Volume 2, Pages:1396 - 1401, Hawaii, Oct. 2005.

[3] G. Loy, E. J. Holden, and R. Owens, *A 3D Head Tracker for an Automatic Lipreading System*, Proc. of Australian Conference on Robotics and Automation (ACRA), Melbourne, Aug. 2000.

[4] M. L. Cascia, S. Sclaroff and V. Athitsos, *Fast Reliable Head Tracking under Varying Illumination: An Approach Based on Registration of Texture-Mapped 3D Models*, IEEE PAMI, vol. 21, no.6, June 1999.

[5] D. B. Russakoff and M. Herman, *Head Tracking Using Stereo*, Machine Vision and Applications,vol. 13, pages 164-173, 2002.

[6] H. Nanda, K. Fujimura, A robust elliptical head tracker, Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, p. 469 - 474, Seoul, Korea, 17-19 May 2004.

[7] I. Ravyse, V. Enescu, H. Sahli, *Kernel-based head tracker for videophony*, IEEE International Conference on Image Processing, ICIP, vol. 3, p. 1068-1071, Genova, Italy, 11-14 Sept. 2005.

[8] M. Kourogi, T. Kurata, *A Wearable Augmented Reality Sytem With Personal Positioning based on Walking Locomotion Analysis*, Proc. 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality, Page 342, Tokyo, Oct. 2003.

[9] J. D. Mulder, J. Jansen, A. van RhijnAn, *Affordable Optical Head Tracking System for Desktop VR/AR Systems*, Proceedings of the workshop on Virtual environments, Zurich, 2003.

[10] W.D. McCarty, S. Sheasby, P. Amburn, M.R. Stytz, C. Switzer, *A virtual cockpit for a distributed interactive simulation*, Computer Graphics and Applications, IEEE Volume 14, Issue 1, Pages 49 - 54, Jan. 1994.

[11] F.-E Ababsa, M. Mallem, *Inertial and vision head tracker sensor fusion using a particle filter for augmented reality systems*, Proc., International Symposium on Circuits and Systems, ISCAS 04, p. 156-162, Vancouver, 23-26 May 2004.

[12] R. Huber, C. Nowak and B. Spatzek, *Image acquisition using aperture control adapted to spatio-temporal properties*, Journal Machine Vision and Applications Volume15, Issue 4, Pages 204-215, October 2004.

[13] K Levenberg, *A Method for the Solution of Certain Problems in Least Squares*, Quart. Appl. Math. 2, 164-168, 1944.

[14] D. Marquardt, *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*, SIAM J. Appl. Math. 11, 431-441, 1963.

[15] P. E. Gill and W. Murray, *Algorithms for the solution of the nonlinear least-squares problem*, SIAM J. Numer. Anal. 15 [5] 977-992, 1978.

[16] S. Gurbuz, K. Kinoshita, and S. Kawato, *Real-time human nose bridge tracking in presence of geometry and illumination changes*, in Second International Workshop on Man-Machine Symbiotic Systems, Kyoto, Japan, 2004.