

# DRIVER HEAD MOVEMENT ESTIMATOR ON ARM big.LITTLE PROCESSOR

Goutham A.H<sup>1</sup>

<sup>1</sup>M.Tech in VLSI Design and Embedded Systems, Dayananda Sagar College of Engineering, Vishveshwaraya Technological University, Kumaraswamy Layout, Bangalore-560078

## Abstract

A driving environment that is changing constantly and rapidly, demands a driver to make large head turns. Drivers head movements are strong indicators of drivers focus on the road. In this paper a vision based algorithm is presented to estimate the driver head pose. Determining the head pose using vision based algorithms is a non-invasive method in intelligent driver assistance systems. Many existing state of the art vision based head pose algorithms have difficulties in monitoring the driver head movements. This is because in single camera perspective spatially large head turns disturb the facial features which are necessary to determine the head pose. A distributed camera framework and the use of Constrained Local Model algorithm for head pose tracking is presented. The proposed approach monitors the driver head over wide range of movements.

**Keywords:** Driver Safety, Vision Based Assistance, Head Pose, Constrained Local Model.

-----\*\*\*-----

## 1. INTRODUCTION

During 2011, a total of 4,97,686 road accidents were reported by all States/UTs of India. The proportion of fatal accidents in the total road accidents has consistently increased since 2002 from 18.1 to 24.4% in 2011 [1]. Driver distraction and inattention which include phone usage, talking, eating, drowsiness, etc. are the notable causes of accidents. Accordingly there is a great demand for intelligent driver assistance systems (IDASs) that alert a driver of arriving dangers or guide them through a critical situation[2]-[5]. Driver behavior monitoring is hence becoming a necessary component of IDASs.

Drivers head and eye dynamics reveal where or at what the driver is focusing. Vision based systems provide a noninvasive and noncontact solution when used for gaze tracking. However gaze tracking systems are susceptible to illumination changes in real-world driving scenarios. For an IDAS to be robust, head dynamics monitoring is suggested. Recent studies show head motion, along with lane position and vehicle dynamics predicts a drivers intent to turn [6]. Research has been carried out towards fatigue and attention monitoring using head dynamics[7],[8].

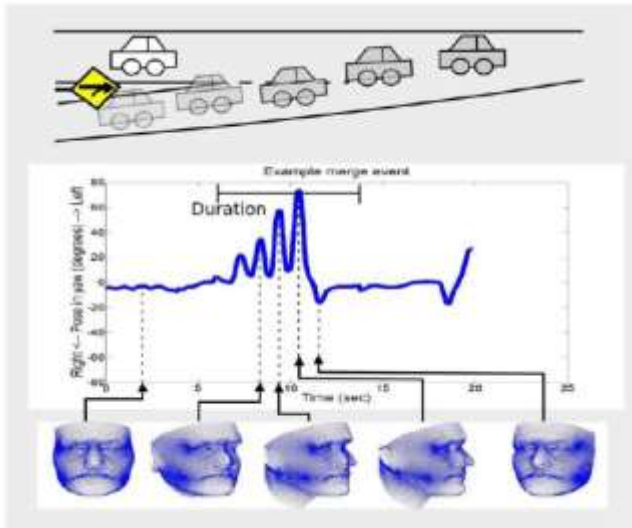
Vision based head dynamics analysis remains a challenging problem. The head dynamics analyzer should be robust to changing driving situations and has to be functional during the entire driving period. Such a system should have the following capabilities.

- *Automatic:* The system should operate without any human intervention and there should be no manual initialization.
- *Fast:* The system must be able to estimate the head pose in real driving environment with real time operation.
- *Wide operational range:* The system should handle large head movements accurately and robustly

- *Lighting invariant:* The system must operate in varying lighting conditions (e.g., sunny and cloudy).
- *Person invariant:* The system must work for different drivers.
- *Occlusion tolerant:* the system should work with partially occluding objects (e.g., sunglasses and hats) or actions (e.g., hand movements).

Many state-of-the-art vision based head pose algorithms are capable of being automatic, fast and person invariant [9]. During a typical ride, it has been shown that the driver spends 95% of the time facing forward [10]. The 5% nonfrontal glances are of special interest where events that are critical to driver safety can occur. Fig.1 illustrates typical dynamics of head pose seen from a fixed camera during a merge event. It is shown that the head pose varies rapidly from forward facing (00 yaw angle). During these times, the performance of monocular based systems degrades due to insufficient facial features caused by self-occlusion. To continuously estimate a drivers head movement, new sensing approaches are required. A good choice is to use a system with multi-cameras [11], [12].

The contributions of this paper are threefold. First a distributed camera framework and different multi-camera configurations are proposed. Second we propose a solution for head pose estimation based using state-of-the-art facial feature tracking algorithm. Finally the success of the system is quantitatively demonstrated on the road. The dataset used for on road evaluation contains spatially large head turns that occur during vehicle maneuvers. We evaluate the proposed system using failure rate, which is the percentage of unreliable output given by the system.



**Fig. 1.** Head movements during a merge event. The 3-D model of a head illustrates the observed facial feature from a fixed camera perspective and the self-occlusion that was induced by large head movements.

## 2. RELATED STUDIES

Vision based head dynamics tracking methods are challenging in naturalistic driving scenario. In a car, constantly changing lighting conditions and illumination changes cause heavy shadows. As a result the techniques which exhibit good proficiency in stable lighting often fail to work. We review past works that have been evaluated in naturalistic driving that have potential to work in real world driving environment. A good overview of head pose estimation in computer vision is given by Murphy-Chutorian and Trivedi [13].

Head pose algorithms are generally classified into the following categories: geometric/shape feature based, appearance/texture feature based and hybrid (shape + texture) feature based methods. Shape feature algorithms analyze geometric configuration of facial features along with face model (e.g. cylindrical [14] or mean 3D face [15]) to recover head pose. Several strategies are proposed using global motion and color statistics to track eyes, lip corners and face bounding box [16]. These facial features are used to estimate the head orientation and gaze direction. However when the driver wears eye glasses or makes conversations this method cannot always find facial features. Lee *et al.* [17] used facial features along with ellipsoidal face model to estimate yaw when the driver rotates head away from head pose. Martin *et al.* [18] tracked non deformable facial features (e.g. eye corners, nose corners and nose tip) on real world driving data and analyzed the geometric configurations to extract the head pose. This is similar to proposed approach but limited to single camera perspective.

The proposed approach uses shape-feature based method. Robust and accurate localization of facial features in shape-feature based algorithms is a challenging task. The use of multi-cameras improve the operational range and helps in improving accuracy. Unlike stereo cameras, the visibility of

faces in both cameras is not assumed and the lengthy calibration process is not required. The cameras used are uncalibrated and the proposed framework utilizes them parallelly and independently.

## 3. CONTINUOUS AND ROBUST HEAD MOVEMENT ANALYSIS: ISSUES AND CHALLENGES

Drivers head dynamics analysis present unique challenges for researchers working in driver monitoring systems. Methods designed and tested in laboratories are not guaranteed to provide robust performance in real world driving environment. Hence a proper evaluation on a real world driving database is needed. The placement of cameras should neither block a driver's view for safe driving and should not be prone to frequent occlusions. The choice of placement is application dependent and a desirable choice would be one that covers a large pose space exhibited by a driver in a given ride. From computer vision perspective, head dynamics properties present a challenge to robustness of many algorithms. Many existing state-of-the-art head pose algorithms rely on portion of face to be visible in the image plane to extract the head pose. This means that the algorithms require visibility of facial features to continuously track the head even during large head movements. With a single perspective approach, the large head movements induce self-occlusion of facial features as shown in Fig. 2. In Fig.2, each row is taken from different camera perspective and each column is time synchronized. Clearly, the multi-perspective approach decreases the issue of self-occlusion at any instant of time helping in increasing the robustness of continuous head tracking.



**Fig. 2.** Multi-perspective data collected during naturalistic on-road driving. Each row of images shows that the images are from a fixed single camera location, and each column of images is time synchronized. Camera 1 is near the right A pillar, Camera 2 is close to the dashboard, and Camera 3 is near the rearview mirror.

Occlusion of facial features can occur due to external objects (e.g. hand movements, sunglasses, hats etc.) as shown in Fig. 2. Camera locations also affect the lighting conditions. Therefore multi-perspective approach can mitigate the adverse effect of any one camera perspective being unable to track the head.

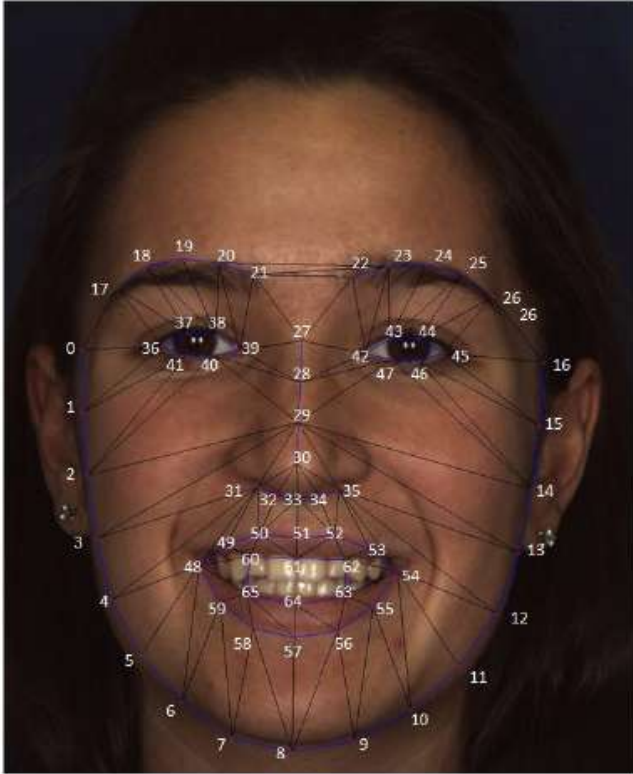


Fig. 3. Tracked facial feature/landmarks and their correspondences in a 3-D face image

#### 4. MULTI-PERSPECTIVE FRAMEWORK

In multi-perspective framework, each perspective is kept independent of one another and are processed in parallel. For estimating the head dynamics, we use shape based algorithms to find the facial features and recover pose from the relative configuration of these facial features.

##### 4.1 Facial Feature Detection and Tracking

Facial features refer to salient landmarks on the face, such as eye corners, nose corners, nose tip, mouth contour and outer face contour as shown in Fig. 3. For automatic facial feature detection and tracking Constrained Local Model (CLM) algorithm is used introduced by Cristinacce and Cootes [19], [20]. CLM is extensively used for facial landmark detection and tracking in a video data. CLM represents faces using local appearance descriptions centered around landmarks of interest, and a parameterized shape model to capture convincing deformation of the landmark locations. The local descriptors are trained from labeled training images for each landmark.

The shape model, also known as point distribution model (PDM), a term coined by Cootes and Taylor [21] is given as follows:

$$\mathbf{x}_i = s \cdot \mathbf{R}_{2D} \cdot (\bar{\mathbf{X}}_i + \Phi_i \mathbf{q}) + \mathbf{t} \quad (1)$$

where  $\bar{\mathbf{X}}_i = [\bar{x}_i, \bar{y}_i, \bar{z}_i]^T$  is the mean value of the  $i$ th feature,  $\Phi_i$  is a  $3 \times m$  principal component matrix and  $\mathbf{q}$  is an  $m$  dimensional vector of parameters that controls non rigid shape.  $s$  is a scaling term that is used to controls the distance from camera ( $s = f/Z$ ). here  $f$  is the focal length and  $Z$  is the depth from camera,  $\mathbf{t} = [t_x, t_y]^T$  is the translation term.  $\mathbf{R}_{2D}$  represents in-plane rotation. Let us define  $\theta_{rg} = \{s, \mathbf{R}_{2D}, \mathbf{t}\}$ . Let  $\mathbf{p} = \{\theta_{rg}, \mathbf{q}\}$ . The deformable model is controlled by parameters  $\mathbf{p}$  and the model instance can be described by locations of respective feature points  $\mathbf{x}_i$  in an image  $I$ . the CLM fitting algorithm look for maximum a posteriori probability (MAP) of the parameters  $\mathbf{p}$ :

$$\rho(\mathbf{p} | \{\mathcal{J}_i = 1\}_{i=1}^n, \mathcal{I}) \propto \rho(\mathbf{p}) \prod_{i=1}^n \rho(\mathcal{J}_i = 1 | \mathbf{x}_i, \mathcal{I}) \quad (2)$$

where  $\mathcal{J}_i \in \{1, -1\}$  is a discrete random variable indicating the alignment or misalignment of the  $i$ th feature point,  $\rho(\mathbf{p})$  represents the prior probability of the model parameters  $\mathbf{p}$  and  $\prod_{i=1}^n \rho(\mathcal{J}_i = 1 | \mathbf{x}_i, \mathcal{I})$  is the joint probability of feature points that are aligned at locations  $\mathbf{x}_i$ , for the image  $I$ . The probability of a certain feature being aligned at location  $\mathbf{x}_i$  is  $\rho(\mathcal{J}_i = 1 | \mathbf{x}_i, \mathcal{I})$ .

To facilitate the optimization the optimization process for efficiency and numerical stability, the true response map  $\rho(\mathcal{J}_i = 1 | \mathbf{x}_i, \mathcal{I})$  of the local detectors are approximated by kernel density estimate (KDE) [22] for its fast convergence property.

##### 4.2 Head Pose Estimation

Given  $n$  ( $n \geq 3$ ) 3D reference points in the object framework and their corresponding 2D projections the perspective- $n$ -point (PnP) problem solving algorithm determines the orientation and position of a camera. A Direct Least Squares (DLS) [23] method is used to solve the PnP and to obtain the head rotation in 3 degrees of freedom. The CLM algorithm outputs 68 landmark points that are considered for PnP problem.

#### 5. RESULTS AND DISCUSSION

The proposed framework is evaluated on naturalistic driving data. The data collection was focused around various maneuvers that cause large head turns as they are critical for driver safety. By evaluating these events, we show the necessity of multi-perspective setup for continuous and reliable head tracking.

## 5.1 Test Bed and Data Set

Three cameras are mounted facing the driver, i.e., one camera near the A pillar, one camera just above the steering wheel, on the dashboard and one camera near the rearview mirror. All the three cameras capture color stream at 30 frames per second at a resolution of 640 x 480 pixels.

Using this test bed, 3 drivers were asked to drive naturally drive on roads near DSCE campus, Kumaraswamy Layout, Bangalore. 30 minutes of data was evaluated for determining the success in facial feature tracking by the CLM algorithm. The traffic was moderate in sunny weather and speed breakers, potholes and irregularities introduced slight unwanted drifts on the drivers head. The drivers passed through stop signs and made multiple turns and lane changes, resulting in a data with wide changes in head pose. Events such as left/right turns, left/right lane changes, merges which are of special interests are shown in Table I. Table 1.

**Table 1:** list of events in the dataset and its respective count

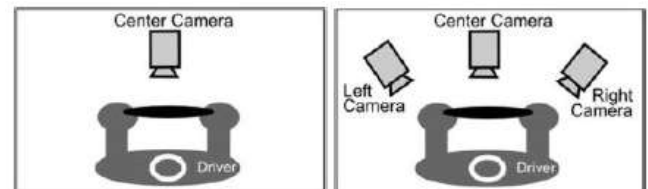
Events	No. of events
Right turns	6
Left turns	7
Stop sign	3
Right lane change	8
Left lane change	4
Merge	3
Total	31

## 5.2 On Road Performance Evaluation

Head tracking is first tested on single camera perspective and it is compared with the multi-perspective tracking. Single perspective refers to the output obtained from front facing camera. The spatial distribution of cameras for single perspective and multi-perspective is shown in Fig. 4. Head tracking is said to be failed for the chosen camera perspective if the estimated pose is unavailable. The number of frames in which the head tracking is lost normalized by the total number of frames over all the events gives the failure rate.

The CLM algorithm was implemented using OpenCV library and the platform used for on-road evaluation is ODROID-XU3 single board computer (SBC) containing ARM processor which is highly portable. All the 8 cores present in the processor were efficiently utilized by parallelization to improve the computation time. The single camera perspective runs at 6 frames per second whereas the multi-perspective approach runs at 4 frames per second for each one of the cameras. The maximum power dissipation by this SBC is less than 20 watts making it power efficient

reducing fuel consumption required to charge the car battery. As shown in Table II failure rate of the single perspective approach is 19.28% and that of multi-perspective approach is 14.07% showing considerable improvement. These percentages are important for applications such as lane change intent prediction, where large percentage of loss in head tracking leads to an unreliable system. Hence smaller loss percentage in head tracking is preferable in critical situations during driving.



**Fig. 4** Setup of the one-camera perspective and the three-camera perspective the performance evaluation of the multi-view framework

**Table 2:** Multi-perspective evaluation

Perspective	Failure Rate
Single	19.28
Multi	14.07

## 6. CONCLUSION AND FUTURE WORK

In real-world driving environment during maneuvers, a driver has to make large head turns to make left/right turns, change lanes. At these critical moments, continuous and reliable tracking of driver head becomes a very important task. Since a single perspective is unreliable during large head movements, a multi-perspective framework is proposed. The system was evaluated on 30 mins naturalistic driving data and the failure rate was calculated.

For future work, we consider perspective selection procedure [24] wherein the cameras are automatically switched when the driver is facing those particular cameras. For switching the cameras, angle constraints are defined for every camera. We also consider using CUDA technology [25] by NVIDIA for performing the computations and expect head pose tracking in real-time leading to improved accuracy as compared to the current result.

## REFERENCES

- [1]. Manisha Ruikar, "National statistics of road traffic accidents in India" Journal of orthopaedics, traumatology and rehabilitation. 2013, p. 1-6
- [2]. M. M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," IEEE Trans. Intell. Transp. Syst., vol. 8, no. 1, pp. 108–120, Mar. 2007.
- [3]. M. M. Trivedi and S. Y. Cheng, "Holistic sensing and active displays for intelligent driver support systems," Computer, vol. 40, no. 5, pp. 60–68, May 2007.

- [4]. K. S. Huang, M. M. Trivedi, and T. Gandhi, "Driver's view and vehicle surround estimation using omnidirectional video stream," in Proc. IEEE Intell. Veh. Symp., 2003, pp. 444–449.
- [5]. A. Doshi, S. Y. Cheng, and M. M. Trivedi, "A novel active heads-up display for driver assistance," IEEE Trans. Syst., Man, Cybern., B, Cybern., vol. 39, no. 1, pp. 85–93, Feb. 2009.
- [6]. S. Y. Cheng and M. M. Trivedi, "Turn-intent analysis using body pose for intelligent driver assistance," IEEE Pervasive Comput., vol. 5, no. 4, pp. 28–37, Oct.–Dec. 2006.
- [7]. S. Baker, I. Matthews, J. Xiao, R. Gross, T. Kanade, and T. Ishikawa, "Real-time non-rigid driver head tracking for driver mental state estimation," Carnegie Mellon Univ., Tech. Rep. CMU-RI-TR-04-10, Robot. Inst., Feb. 2004.
- [8]. J. P. Batista, "A real-time driver visual attention monitoring system," in Proceedings of the Second Iberian conference on Pattern Recognition and Image Analysis—Volume Part I. Berlin, Germany: Springer-Verlag, 2005, pp. 200–208.
- [9]. E. Murphy-Chutorian and M. Trivedi, "Head pose estimation in computer vision: A survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 4, pp. 607–626, Apr. 2009.
- [10]. S. Martin, A. Tawari, E.-M. Chutorian, S. Y. Cheng, and M. M. Trivedi, "On the design and evaluation of robust head pose for visual user interfaces: Algorithms, databases, and comparisons," in Proc. 4th ACM SIGCHI Int. Conf. AUTO-UI, 2012, pp. 149–154.
- [11]. S. Martin, A. Tawari, and M. M. Trivedi, "Monitoring head dynamics for driver assistance systems: A multi-perspective approach," in Proc. IEEE Int. Conf. Intell. Transp. Syst., 2013, pp. 2286–2291.
- [12]. A. Tawari and M. M. Trivedi, "Head dynamic analysis: A multi-view framework," in Proc. Workshop Soc. Behaviour Anal., Int. Conf. Image Anal. Process., 2013, pp. 536–544.
- [13]. E. Murphy-Chutorian and M. Trivedi, "Head pose estimation in computer vision: A survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 4, pp. 607–626, Apr. 2009.
- [14]. R. Valenti, T. Gevers, and T. Gevers, "Combining head pose and eye location information for gaze estimation," IEEE Trans. Image Process., vol. 21, no. 2, pp. 802–815, Feb. 2012.
- [15]. S. Martin, A. Tawari, E.-M. Chutorian, S. Y. Cheng, and M. M. Trivedi, "On the design and evaluation of robust head pose for visual user interfaces: Algorithms, databases, and comparisons," in Proc. 4th ACM SIGCHI Int. Conf. AUTO-UI, 2012, pp. 149–154.
- [16]. P. Smith, M. Shah, and N. da Vitoria Lobo, "Determining driver visual attention with one camera," IEEE Trans. Intell. Transp. Syst., vol. 4, no. 4, pp. 205–218, Dec. 2003.
- [17]. S. J. Lee, J. Jo, H. G. Jung, K. R. Park, and J. Kim, "Real-time gaze estimator based on driver's head orientation for forward collision warning system," IEEE Trans. Intell. Transp. Syst., vol. 12, no. 1, pp. 254–267, Mar. 2011.
- [18]. S. Martin, A. Tawari, E.-M. Chutorian, S. Y. Cheng, and M. M. Trivedi, "On the design and evaluation of robust head pose for visual user interfaces: Algorithms, databases, and comparisons," in Proc. 4th ACM SIGCHI Int. Conf. AUTO-UI, 2012, pp. 149–154.
- [19]. D. Cristinacce and T. F. Cootes, "A comparison of shape constrained facial feature detectors," in Proc. 6th Int. Conf. Autom. Face Gesture Recog., 2004, pp. 375–380.
- [20]. D. Cristinacce and T. F. Cootes, "Feature detection and tracking with constrained local models," in Proc. British Mach. Vis. Conf., 2006, pp. 929–938.
- [21]. T. F. Cootes and C. J. Taylor, "Active shape models—smart snakes," in BMVC92. New York, NY, USA: Springer-Verlag, 1992, pp. 266–275.
- [22]. J. Saragih, S. Lucey, and J. Cohn, "Face alignment through subspace constrained mean-shifts," in Proc. IEEE 12th Int. Conf. Comput. Vis., Oct. 2, 2009, pp. 1034–1041.
- [23]. J. Hesch and S. Roumeliotis, "A direct least-squares (DLS) method for PnP," Proc. ICCV, pages 383–390, 2011.
- [24]. S. Martin, A. Tawari, and M. M. Trivedi, "Monitoring head dynamics for driver assistance systems: A multi-perspective approach," in Proc. IEEE Int. Conf. Intell. Transp. Syst., 2013, pp. 2286–2291.
- [25]. J. Nickolls et al., "Scalable Parallel Programming with CUDA," ACM Queue, vol. 6, no. 2, 2008, pp. 40–53.