International Journal of u-and e-Service, Science and Technology
Vol.7, No.4 (2014), pp.359-374
http://dx.doi.org/10.14257/ijunnesst.2014.7.4.33

Head Pose Estimation for Car Drivers

Reda Shbib, Shikun Zhou, David Ndzi and Khalil Alkadhimi

School of Engineering, University of Portsmouth, UK
reda.shbib@port.ac.uk . shikun.zhou@port.ac.uk , david.ndzi@port.ac.uk , khalil.alkadhimi@port.ac.uk

Abstract

Advertising agencies have a growing interest in finding out if roadside advertisements have reached the target audiences. The drivers' head position' would enable us to determine if they have been looking at the billboards. Numerous researches were focusing on the drivers' attention from inside their cars; however, some have tried to determine the interaction between the drivers and outside objects. In this paper, the proposed Head Pose Estimation (HPE) is able to detect drivers' head motion/position and try to determine if they are looking at the billboard. This information is vital for the Advertising companies for the evaluation of the effectiveness of their advertisements.

Keywords: *Head Pose Estimation, Image processing*

1. Introduction

Most of business companies are putting huge efforts in order to promote specific products and influence their costumers by developing attractive advertising plans. Significant research has been carried out to assess the impact of road side advertisement that may distract driver attention and may cause crash, however few research has been carried out to study if these advertisements are effective. The effectiveness of the billboard advertisement can be determined by some measurements such the number of drivers who have showed interactions and other criteria is testing the speed of the cars when passing over road side advertisements. In addition, some business organizations would choose to have their billboard almost seen by precise costumers (females, youths) and thus they would like to count the number of persons, who have seen their advertisements in order to operate correspondingly. [1].

The attractiveness of a varied types of application of which head position and detection are major part have effectively increased the number of research aim at providing more strong approaches of head pose estimation and tracking . It has showed in the last decade that this field has received lots of attention and development from academics and industrial researchers. Captured images through videos of car drivers would be factual data that are stored in the server as a source video analytic. Through Business intelligence and analysis approaches those data and analysis results will be used to evaluate the effectiveness of the advertisement, as well as indicating the impact on the enterprise. This information is indeed one of the vital capabilities of an enterprise which to be aligned with its advertising strategy and objective. In addition, video analysis has become one of the benefit realization factors for organization and this state can be achieved only by clear understanding and collaborative effort of engineers and the Subject matter experts of the business community.

In this paper we propose an integrated face and head position estimation technique in order to assess the effectiveness of the billboard .This would be helpful for businesses to optimize

ISSN: 2005-4246 IJUNNEST

commercial billboard advertisements and decision making. The proposed system involves face detection and tracking module for drivers who are watching the billboard detecting.

2. Related Work

Online pedestrian counting system for electronic board advertisements has been developed by [2] using a fixed camera that can count the number of people who are really looking at a TV board advertisement. The proposed approach makes use of Support Vector Machine (SVM) in order to extract facial features and detect the frontal faces regions of people who are really watching the billboard advertisement. In order to achieve more accurate results, Fisher's Linear Discriminant (FLD) has been applied as a classifier for more robust pedestrian recognition. For the accuracy levels, results have shown that the proposed algorithms provide a count accuracy level up to 80 % on average for the crowded people. However the errors might occur if some people's faces covered with the masks or hair. Moreover, the system is not efficient when use in crowded situations and it has been tested just in small crowded scenes.

Face detection approach for smart electronic billboard has been developed by [3].The proposed system counts the number people who are actually watching the billboard , and it can identify their gender as well in order to compile demographic information about the viewer's enabling companies to have more valuable information for trading . The proposed system has focused on counting number of male and females looking at electronic billboard advertisement by extraction of a set of discriminant features from the torso of the pedestrians. Adaboost Algorithms has been used and an online classifier which would conduct to more strong viewer recognition. However, this system has a number of limitations. First, the image algorithms used in the system are not robustness as the background model is not sufficient to address the changing conditions and relies significantly on the camera position.

Awareness monitoring system for drivers accurate estimates of head's position has been proposed by Murphy-Chutorian [4] which integrating approaches to estimate head position has been developed based on visual 3-D tracking. The two proposed approaches are used in real-time application to detect the pose and orientation of the driver's head. This technique make use of three connected components to achieve head detection that provide a primary estimations of the head's position, and then a continuous tracking its position .The first head detection module involves set of arrays using Haar-wavelet Ad boost cascades . Support vector regression (SVR) has been used as an initial pose's estimate module taken as input with localized gradient orientation (LGO) histograms. The 3-D head orientation and movement of the head has been provided by another tracking estimation module using appearance-based particle filter.

Head movement detection technique has been proposed by Liu et al[5]. The developed approach aim at detect and recognize the activity attention of car drivers. Pose and head movement was estimated by analyzing the relative pose between contiguous views in the subsequent frames. Scale-Invariant Feature detections algorithm has been used in order to match and extract the interesting points and features over two consecutive views. The head position angle is then estimated using two views geometrical approximation so the three dimensional Cartesian coordinate of the head pose x, y and z can then be determined. However, this paper only focus on the fundamental problems of head motion with a suggestion a basic algorithm but without any potential formula and no implementable result has been found in this work; Moreover, the accuracy level and the performance of this approach were not clearly described, consequently, much amount of work need to be done in this work in order to proof the applicability and the effectiveness of the proposed algorithm.

Extracting Facial Feature and head pose estimation and tracking has been developed using Feature Based tracker (FBT). A 3-D estimation tracking of the head position and deformities of facial features in a sequence of images has been developed by[6] involving an approach formed by two stage A stabilized face tracking is defined by learning different 3-D face deformities from sequence of data also, make use an optic flow representation of combined with features that have been tracked. The Feature-Based Trackers is to somehow more accurate for simultaneously head and tracking of face features however FBT takes over the problems of the computation of stereo vision and optical flow specifically, the proposed approach is limited to the variation in illumination and among these, it is very sensitive to the head's pose variation which can lead geometrical deformation in the presented image of face.

Another approach has been presented by that make use of Active shape models (ASMs) as an alternate technique to FBT [7]. Generally in ASM, the variation of the shape normally can be captured by using Point Distribution model (PDM) which represents the geometrical mean of the shape. The modelling of local appearances for a landmarks combination can be the archived using intensity gradient Distribution. The parameter of the shape are updated frequently by discovering the closely match for every point. However, Variation in illumination and occlusion level can reduce the effectiveness of ASM because they would reduce the textures extracted information. In addition to these factors, ASM would need a big amount of training data set for learning.

Head pose tracking technique has been presented by Song et al. [8] in this work, detection of the face is achieved using Adaboost approach .Then; Head orientations are being detected by analyzing face's location. The Author has introduced five head motions as initial basis of head orientations. In order to find the central coordinate a person head, it is supposed to calculate the geometric center of any detected face. Then, in order to trace head's motion, the coordinates are analyzed over the period of time. The camera that has been used in this application is not a head mounted. The proposed technique was developed to be fast in order to makes it available in some application for disable people. However, the accuracy level and the performance of this approach were not been described in this paper.

The literature also has pointed out the necessity to undertake further developments in areas such as tracking, detection and head pose estimation techniques. In addition the increasing number of commercial and safety application has resulted in more investigation with respect to the development of more efficient and robust head pose estimation system that require algorithms that are able to report high performance with little reconfiguration that have to be sufficiently extensible and adaptable in order to automatically adjusted and be able to cope up with any changes in the environment such as the geometry of the scene activities taking place in the scene, and lighting, location Some of the discussed algorithms has showed a high setup cost, complex deployment procedure limited field of view and occlusion which may affect their accuracy level .

The rest of this article is organized as follows. State of arts has been reviewed is the next section. The proposed technique is described in section 3. Face detection and Kalman filter tracking integrated with AMM is presented in section 4. Some of Experimental results which provide a validation of the proposed approach is demonstrated in section 5 followed by a conclusion and future work.

3. Proposed System

The proposed technique would help in the detection of the current state of the driver's head position if they are looking at the billboard in every frame when approaching the billboard. Several head pose estimation techniques have been presented. Stereo camera has been used in order to extract depth information. However these types of cameras are expensive. Thus a lot

of proposed works still rely on monocular techniques which have been used to estimate head pose movement. Among these methods, tracking approach has shown very good results .Basically the position of the head is estimated by recovering the orientation of head over successive images frames. However lots of tracking approaches try to calculate the parameters of the head movement utilizing the information of grey or color two images frames

In this work we have realized that is very useful to consider the prior state while iterating the motion of the head. For instance, if the head is rotating right in the (k) frame, the probability of that head will rotate right in the (k+1) image frame Thus, this would be as prior information which can raise the accuracy level and effectiveness of the estimation. Consequently we have introduced a new approach which makes use of Kalman filter and facial feature tracking extracted using Active Appearance Model (AMM).

The proposed tracking approach has been applied in order to find the position of the head in the k frame. Then the position of the head in the (k+1) frame will be estimated using Kalman filter which increase the posterior probability of the head's position to the maximum based the earlier estimation . The outcomes of the prediction phase are utilized to enhance the execution of the tracking technique.

The video is acquired by using 700 TVL Weatherproof IR Camera that works well in the areas such as rain, dust, or wet environment with a clear resolution image ,then Matlab is used to order to extract the recorded video and counting the number of people based on the parameters mentioned before .

3.1 Face Detection and Head Pose Estimation

3.1.1. Driver Face Detection

In order to detect if the driver is really watching the billboard, it is important to detect his face positions since that would be important features showing that they are really looking at the billboard advertisement. Therefore, face detection it is an important part to be developed. The proposed technique for object and face detection is extremely rapid in term of processing images and has achieved high rate detection.

The proposed method of face detection involves three key components. Frame segmentation of the moving object, Facial features extractions and face detection.

As a first step the video stream is pre-processed using Gaussian mixture modelling (GMM) for image segmentation and background subtraction in order to extract the region that contain human face. Then a multi-layer classification has been adopted using Adaboost algorithm in order to obtain the precise position of candidate's faces. Figure 1 show the proposed system below.

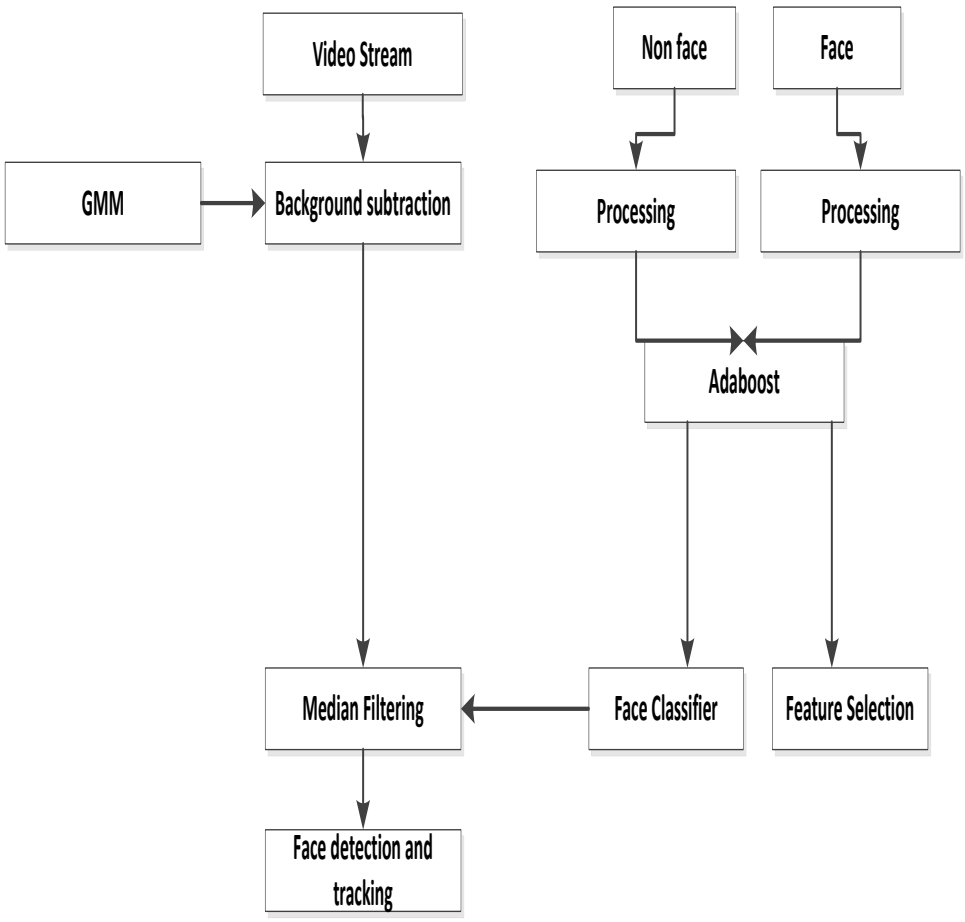

Figure 1. Proposed System Flowchart

3.1.2. Object Segmentation

Single state HMM or a GMM is a robust parametric model for modelling various types of distributions or densities. It has been used for adaptive Gaussian mixture modelling of an image background subtraction[9]. A Gaussian mixture model is a parametric learning model and it assumes the process being modelled has the characteristics of a Gaussian process. A Gaussian process assumes that the parameters do not change over time. So this is an excellent assumption which is true as an image can easily be assumed to be stationary during a single frame.

A GMM tries to capture the underlying probability distribution governing the instances presented during training of the GMM. In our case, we let it learn the background pixels by analyzing the common pixels across a few video frames (first few video frames). Given a test instance it tries to estimate the maximum likelihood of each of the image pixels to test whether each area (comprising of pixels) is background or the Blob (moving objects on the background) given the trained GMM model. It operates on the probability of a pixel lying under in the background or the foreground. This threshold is configurable.

In this model the value of particular pixel over the time is seen as a mesurment X_t of a stochastic variable. At any time along the current measurment of X_t , The history M_t= { X_1 , X_2 , …, X_{t-1} } is known (Stauffer, 1999).

Therefore , the current history of a particular pixel can be moddeled by mixture of K Gaussian distrubutions. Different colours are supposed to denotes as different Gaussian .The probability to detect the current background pixel X_t is the weighted sum of the K distrubution

$$P(X_t) = \sum_{i=1}^{K} w_{i,t} * \mathcal{F}(X_t, \mu_{i,t}, \sum_k) \quad \textbf{(1)}$$

K= number of Gaussian distrubutions

$w_{i,t}$, is the wieght of i^{th} distrubution at time t ,and the $\sum w_i$ =0

Where μ_k is the mean and $\sum_k$ is the covariance matrix of the kth density

Thus , how longer a color is staying in the picture in represented by the probability density function :

$$\mathcal{F}\left(X \middle| \mu_{i,t}, \sum_k\right) = \frac{1}{(2\pi)^{\frac{n}{2}} |\sum_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu_k)^T \sum_k^{-1} (X-\mu_k)} \quad (2)$$

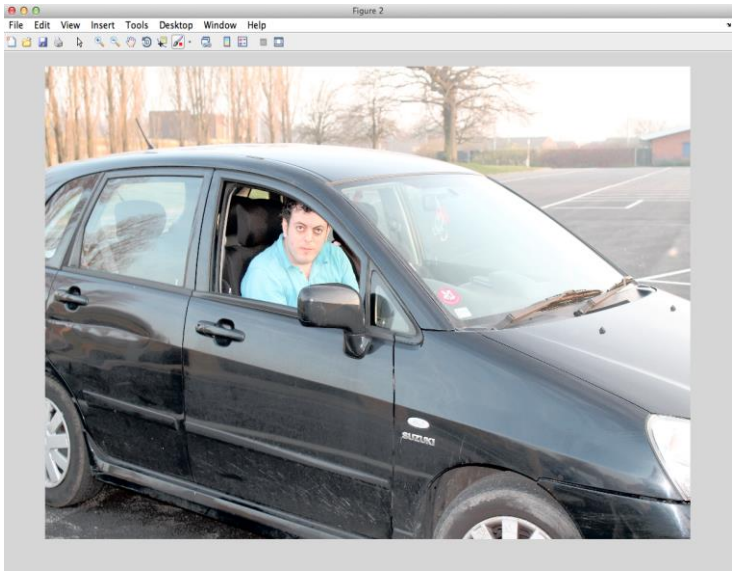

Figure 2. Object Segmentation. (a) Original Image

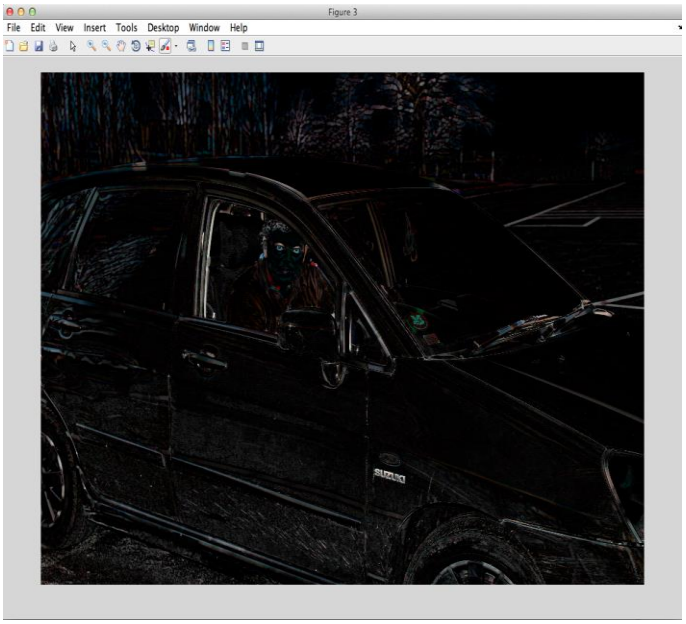

Figure 3. Object Segmentation. (b) Background Subtraction

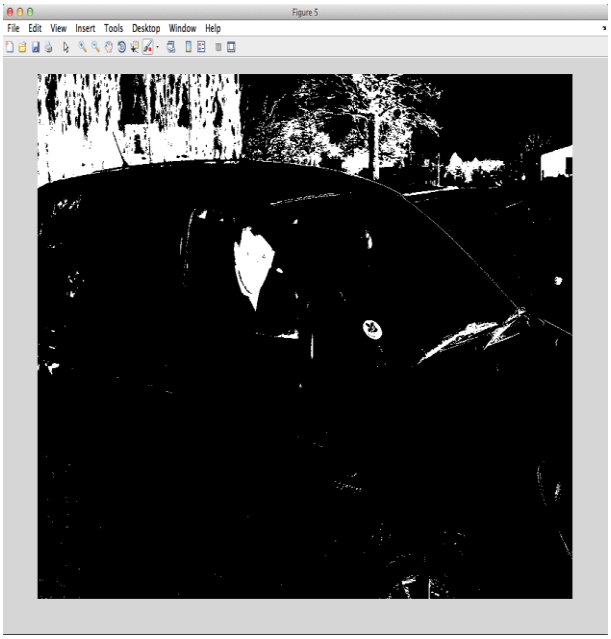

Figure 4. Object Segmentation. (c) Image after Applying the GMM Method

After applying GMM for pixel edge extract of subsequent frames. Dilation and median filter approach has been applied in order to reduce the noise. The equation 3 shows the Dilation calculation:

$$I_{Image} = A \oplus B = (i | \quad B_i \cap A) \neq \emptyset \qquad (3)$$

Where A and B are the binary images

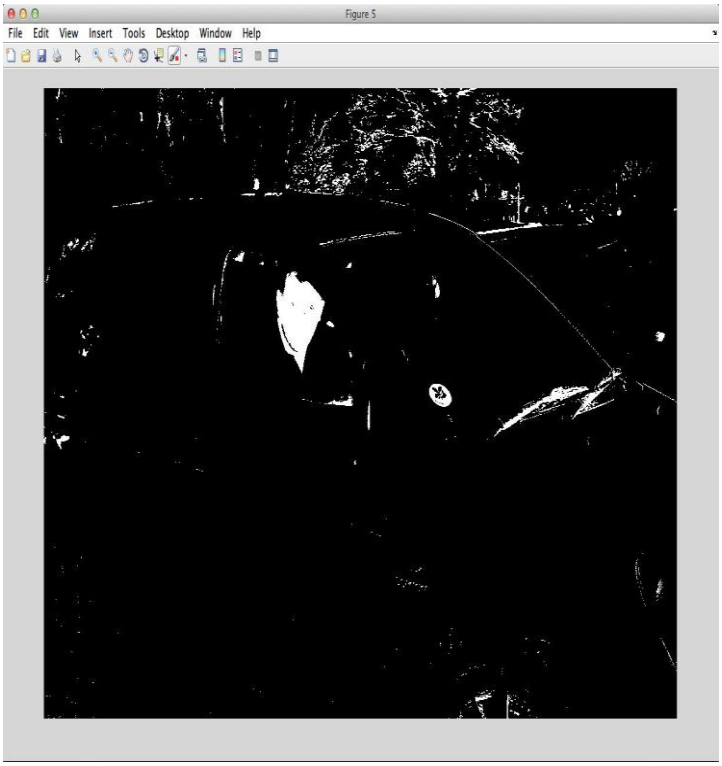

Figure 5. Noise Filtration

4. Features Extractions

4.1. Haar Features

After image processing segmentation and background subtraction we have to extract facial features in order to detect the face candidates. In this part and for very fast features evaluation we adopt face detector developed by Paul Viola *et al.*, [10]. Viola has developed effective Haar features for face detection in 2001 which would be and effective method to differentiate between face and non-face, one of the main advantages of using Haar-like features is it's the very fast speed of its calculation. Haar features are a composition of rectangular features. As definition, it is the difference between the summations of the value of pixels of areas within the rectangle. The detection process used by Paul Viola et al. is based on the features instead of directly pixel detections. Three types of features have been used.

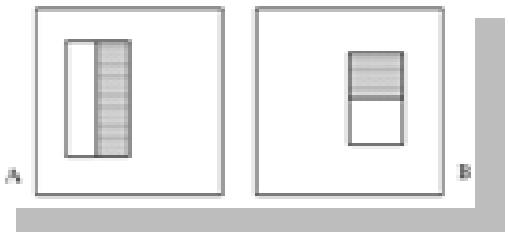

Figure 6. Two-rectangle Feature

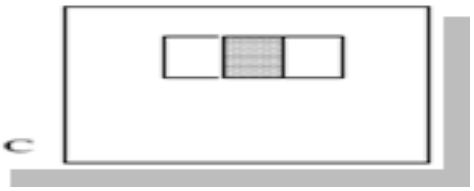

Figure 7. Three-rectangle Feature

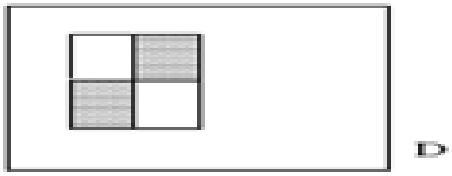

Figure 8. Four- rectangle Feature

In this work, Haar features that have been selected are shown in Figure 9.

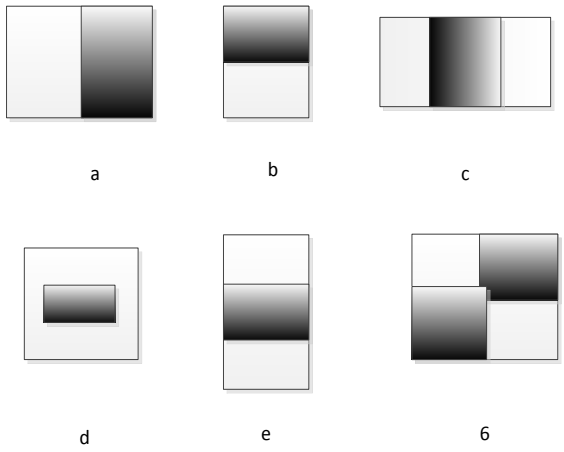

Figure 9. Selected Haar Features

We can calculate easily Haar features using the approach of the integral image that can be computed as follows.

$$ii(x,y) = \sum_{x' \le x, y' \le y} i(x',y') \qquad (4)$$

Where $i(x',y')$ represent the original image and has (x',y') points.

Each point within the obtained integral image represents the summation of the values of pixel in the rectangle which use the origin of the image and the point in order to construct the diagonal endpoint.

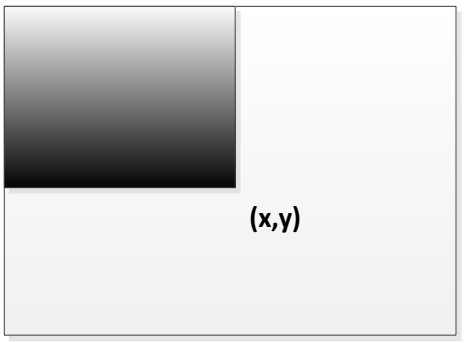

Figure 10. Integral Value of (x, y)

Using a recurrence process the integral image can be then calculated

$$ii(x,y) = ii(x-1,y) + s(x,y) \quad (5)$$

$$s(x,y) = s(x,y-1) + i(x,y) \quad (6)$$

Given a set of features and tanning set of "face" and "non-face" images, there variety of machine learning technique that can be effectively used in order to learn a classification function. Adaboost techniques have been used in order to select the features set and for the classification training. Full explanation of Adaboost algorithm can be found in [10].

5. Experimental Results

Using the proposed approach in this work for face detection multi-face video images, we have got a good detection which has good level of accuracy. Some experimental results are shown in Figure 11.

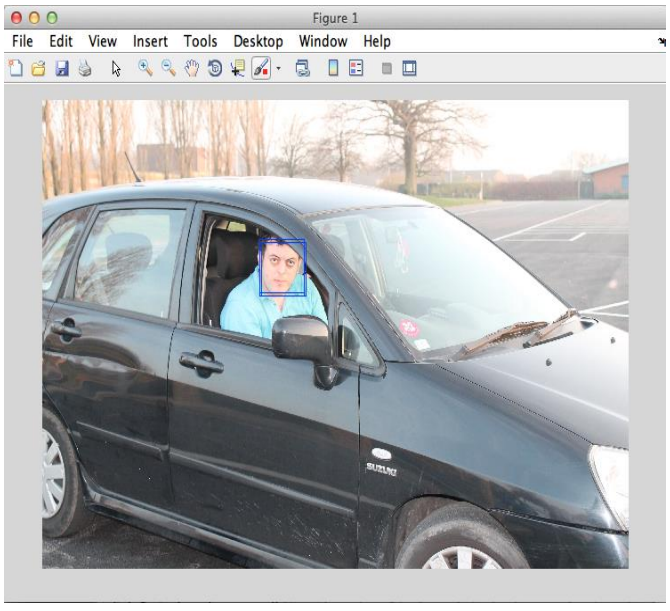

Figure 11. .Diver Face Detection

6. Head Position

6.1. Drivers Head Position Movements

At second stage, it's important to detect the head movement of driver. Tracking head's driver would enable us to know if there is any interaction with the billboard. We define the interaction when the driver rotates his head to look at the advertisement. Because, it is difficult to achieve an accurate measurement of the head's rotation due to the speed of the car and the need of high resolution camera we assume that a measurement is a little amount of head rotation when passing over the billboard.

One of the problems of tracking technique is to obtain a precise estimation and initialization of the position of head. In general, head's frontal view is considered as the initial position. In this paper we have used Adaboost technique in order to detect the frontal drivers' faces. This technique would assume that all detection as frontal even if the face has rotated for an small angle, so in order to avoid this , for driver we have integrate this technique with Active Appearance Model for extracting more facial features.

In this work we have used in this paper texture based tracking approach to achieve head pose estimation. Let μ denote the motion vector $i = [\varepsilon_x, \varepsilon_y, \varepsilon_z, t_x, t_y, t_z]$, where $\varepsilon_x, \varepsilon_y, \varepsilon_z$ denote the rotations and t_x, t_y, t_z denote the 3D transformations relative x, y, z axes

In addition Haar features have been used for frontal face detection and find a first head position S_0 (Figure 12). It is possible to recover The relative movement among two consecutive nth and (n+1)th by applying the tracking technique .So the position of the head at nth frame is given by this equation

$$S_k = S_0 + \sum_{n=1}^{k} i_k \qquad (7)$$

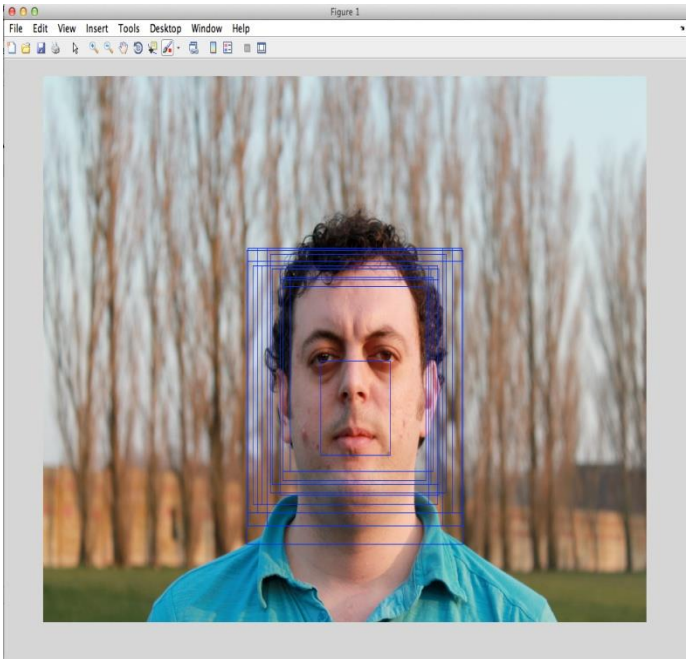

Figure 12. Frontal Face Detection using Haar Features and AAMs at P_0

Applying Kalman filter which has two stages. The initial prediction stage and the correction stage. The position of the head in the new image frame can be predicted by the first stage which would help the tracking process to estimate the position of the head in the new frame. Then the covariance error and the new estimation is updated at the correction stage.Fig.13 shows the flowchart of the proposed system.

After the front face is detected, the position of the head is allocated its first value P_0. So that the head pose estimation can now be achieved by recovering the relative head's movements between nth frames. Tacking head using texture based would find small shift movement even if the head is slightly shifting. The proposed technique apply a cylindrical model in order to approximate the real representation of the head's candidate .Head area the is extracted in order to generate a texture which is can be mapped onto the surface of the 3d cylindrical model that involves n number of triangles .Then, Based on the cylinder's position and its orientation the head texture will be rendered on the image.

In order to implement the proposed image registration technique it is required to perform a reiterative approach for image registration. Head pose and turning in the new image frame are presumed to be identical as the pose of the rendered image frame which would lead to minor and small amount of mead error among the image and the new frame . However, if the mean error is high than a specific threshold (the changing in head position) in this case the approach would apply a searching loop for motion's parameters. The proposed algorithms involves several number of reiteration, in which each one the approach utilise gradient information and estimates the motion vector μ

$$\mu = \frac{-\sum_n (A_t (A_u F_\mu)^T)}{\sum_n (A_u F_\mu)^T (A_u F_\mu)} \quad (8)$$

Where A_t and A_u indicate temporally and spatially gradients, n number of reiterations

- F_μ Denote the deferential F
- A_u can be calculated by applying a sobel operator
- A_t is calculated by differentiate the image that has been rendered and the new image frame

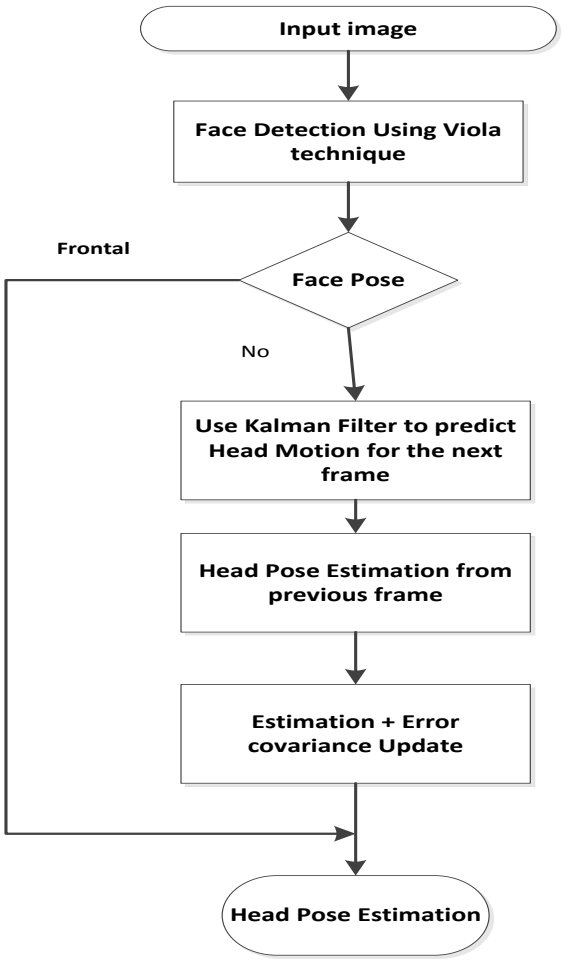

Figure 13. Head Position Flowchart

If the error between the frame and the image is slightly minor the reiteration in this case would converge ,Meanwhile, the conversion is being consider as the motion of the head candidate and the head zone will be extracted and considered as a texture for the subsequent frame else the iterative process will stay running Equation 10 showed that all pixel within the head zone are engaged in the calculation however it seems that the level of confidence for the 3d cylinder is different between pixels . We have utilized a confidence level (CL) which specify pixel's confidence and can be obtained over the execution procedure of the 3D model in which it is equal to zero at the first stage.

$$CL = 255^{\cos^2 \alpha}$$

Where α represent the angle between camera direction and the normal line perpendicular to triangle surface.

Due to its effectiveness and feasibility to reduce noise Kalman filter has been used in order to predict face position. It involves two processes. The first one is the processing model can detect the head movements and the second is the measuring model. The relation between states of two consecutives phases is ruled by the processing model

The following equation show this relationship

$$X_{n+1} = HX_n + Y_n \qquad (9)$$

Where the X_n represent the motion vector state of the recent frame .Head's rotation is represented by six different variables (velocities +acceleration) of the head motion

$$X_n = \begin{bmatrix} (\varphi_x \varphi_y \varphi_z)(t_x t_y t_z) \\ (\varphi 1_x \varphi 1_y \varphi 1_z)(t1_x t1_y t1_z) \\ (\varphi 2_x \varphi 2_y \varphi 2_z)(t2_x t2_y t2_z) \end{bmatrix} \qquad (10)$$

Where H_n is the matrix

$$H_n = A_{6.6} \oplus \begin{vmatrix} 1 & \Delta & \frac{\Delta^2}{2} \\ 0 & 1 & \Delta \\ 0 & 0 & 1 \end{vmatrix} \qquad (11)$$

The sample interval over two successive measurements is denoted by Δ and it is possible to be obtained from frame rate of incoming video clips. The operation between the two matrices is achieved using Kronecker production $\oplus$

The system noise is expressed by Y_n . All element of Y_n are normally distributed $f(0, \sigma_n)$ Where σ_n is the covariance matrix and can be calculated by:

$$\sigma_n = a.A_{6.6} \oplus \begin{vmatrix} \frac{\Delta^5}{20} & \frac{\Delta^4}{8} & \frac{\Delta^3}{6} \\ \frac{\Delta^4}{8} & \frac{\Delta^3}{3} & \frac{\Delta^2}{2} \\ \frac{\Delta^3}{6} & \frac{\Delta^2}{2} & \Delta \end{vmatrix} \qquad (12)$$

The acceleration of the movement of the head is represented by a

The position of head is estimated by the measurement vector P_n

$$P_n = C_n X_n + v_n \qquad (13)$$

Noise Matrix C_n

$$C_n = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad (14)$$

Kalman filter offer a good approximation of the recent state of $X_n$taking the measurement vector P_n as an input, and it will provide the estimate stage of the X_{n+1}

I. HEAD ROTATION RESULTS

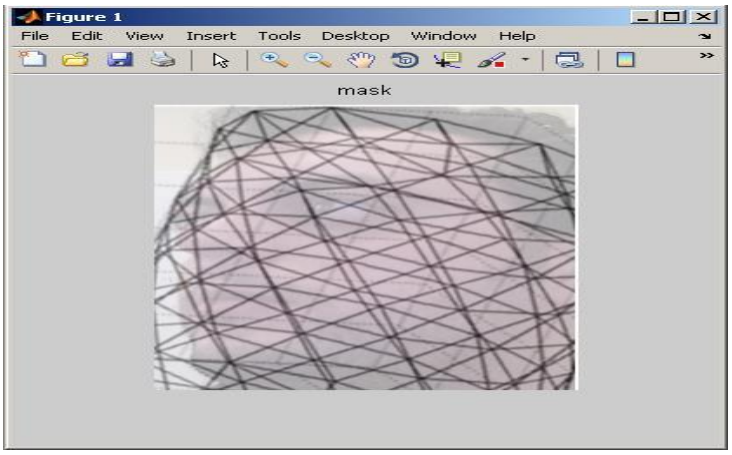

Figure 14. Head Position

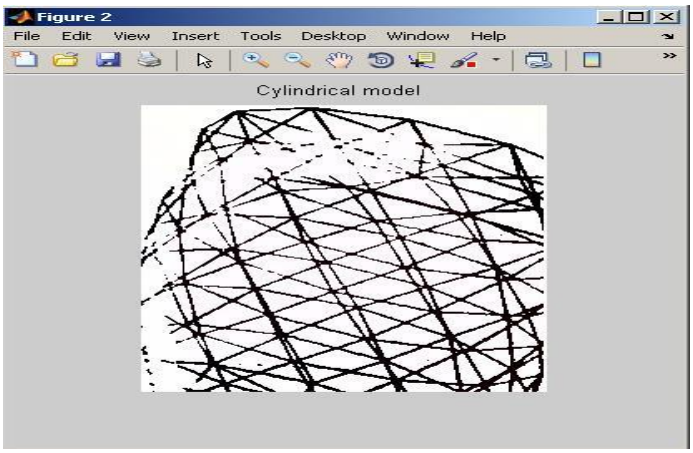

Figure 15. Cylindrical Model

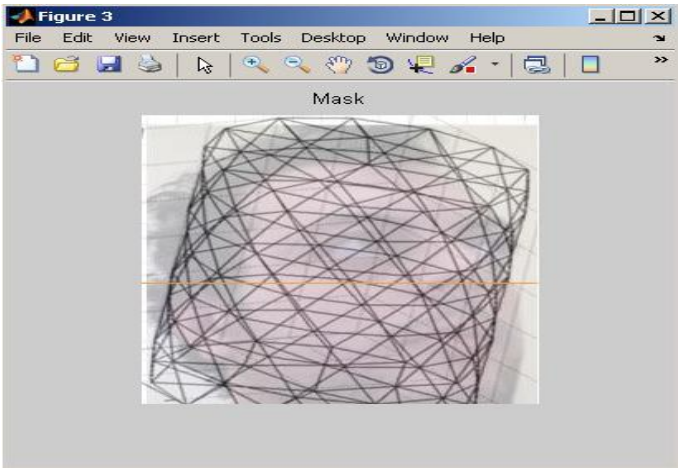

Figure 16. Head Left Position

7. Conclusion

The video analysis system in context of demand of today's upcoming technology has originated a new discipline where business intelligence has taken the role of utilizing the analysis result for successful operation of organizational main stream. At present the video solution and its storage can provide increasing levels of business intelligence for successful advertisement operation to remain competitive in context of the today's global economy. In this paper have explored the video capabilities as an intrinsic part of advertising business operation by considering the technical features so far evolved with the interpretation of some video algorithms. The Video analytics with the support of Business intelligence is now taken the position of one of the capabilities of advertise management which should have the alignment with the objective of the organizations for the organizational operation and its success. For roadside billboard, we have proposed a head pose estimation technique can detect people who are really seen and have interaction with the billboard in order to obtain useful information so that the existing billboard advertisement can be updated according to the collected data over a period of time which lead businesses to effective and attractive advertisement management system and efficient decision making

References

[1] H. Fradi and J.-L. Dugelay, "People counting system in crowded scenes based on feature regression," in Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European, 2012, pp. 136-140.
[2] C. Duan-Yu, "An online people counting system for electronic advertising machines," in Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on, 2009, pp. 1262-1265.
[3] D.-Y. Chen and K.-Y. Lin, "Face-based multiple instance analysis for smart electronics billboard," Multimedia Tools and Applications, vol. 59, pp. 221-240, 2012/07/01 2012.
[4] E. Murphy-Chutorian and M. M. Trivedi, "Head Pose Estimation and Augmented Reality Tracking: An Integrated System and Evaluation for Monitoring Driver Awareness," Intelligent Transportation Systems, IEEE Transactions on, vol. 11, pp. 300-311, 2010.
[5] K. Liu, et al., "Attention recognition of drivers based on head pose estimation," in Vehicle Power and Propulsion Conference, 2008. VPPC '08. IEEE, 2008, pp. 1-5.
[6] S. B. Gokturk., "A data-driven model for monocular face tracking," in Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, 2001, pp. 701-708 vol.2.
[7] T. F. Cootes, "Active Shape Models-Their Training and Application," Computer Vision and Image Understanding, vol. 61, pp. 38-59, 1995.
[8] S. You, "Detection of Movements of Head and Mouth to Provide Computer Access for Disabled," in Technologies and Applications of Artificial Intelligence (TAAI), 2011 International Conference on, 2011, pp. 223-226.
[9] C. a. G. Stauffer, W.E.L, "Adaptive Background Mixture Models for Real-Time Tracking, Computer," August 1999, pp. pp. 2246-252.
[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, 2001, pp. I-511-I-518 vol.1.

