ISSN:0975-9646

# ANN Based Head Movement Detection with Eye Tracking

NIMI M R
*Department of Computer Science and Engineering*
*Sarabhai Institute of Science and Technology*
*Vellanad, Trivandrum, India*

RENJI S
*Department of Computer Science and Engineering*
*Sarabhai Institute of Science and Technology*
*Vellanad, Trivandrum, India*

**Abstract: Eye tracking is the process of watching where a person is looking. The significance of eye movements with regards to the perception of and attention to the visual world is certainly acknowledged since it is the means by which the information needed to identify the characteristics of the visual world is gathered for processing in the human brain. Hence, robust eye detection and tracking are considered to play a crucial role in the development of Human-Computer Interaction (HCI), creating attentive user interfaces and analyzing human affective states. The proposed work uses ANN-based head movement detection with eye tracking. This proposed system aim to estimate the head movements and to locate the eye in the image and then use the obtained information about the eye region and the head pose to estimate the eye gaze direction to be used in various applications, such as HCI, virtual reality, driver assistance systems and assistance for people with disabilities.**

**Index Terms— Eye tracking; ANN-based head movement detection; Human Computer Interaction**

## I. INTRODUCTION

Head movement is found to be a natural, simple and effective way of pointing to objects, interaction and communication. Thus, head movement detection has received significant attention in recent research. One of the various purposes for head movement detection and tracking is to allow the user to interact with a computer. It also provides the ability to control many devices by mapping the position of the head into control signals.

Eye-gaze detection and tracking has been an active research field in the past years as it adds convenience to a variety of applications. It is considered a significant untraditional method of human computer interaction. Head movement detection has also received researchers' attention and interest as it has been found to be a simple and effective interaction method.

Eye tracking and head movement detection are widely investigated as alternative interface methods. Using eye tracking or head movement detection as alternative interface, control or communication methods is beneficial for a wide range of severely disabled people who are left with minimal ability to perform voluntary motion. Eye and head movements are the least affected by disabilities because, for example spinal cord injuries do not affect the ability to control them, as they are directly controlled by the brain. Combining eye tracking and head movement detection can provide a larger number for possible control commands to be used with assistive technologies such as a wheelchair. Examples of different fields of applications for both technologies, such as human-computer interaction, driving assistance systems, and assistive technologies are also investigated. The following figure gives an outline of the system.

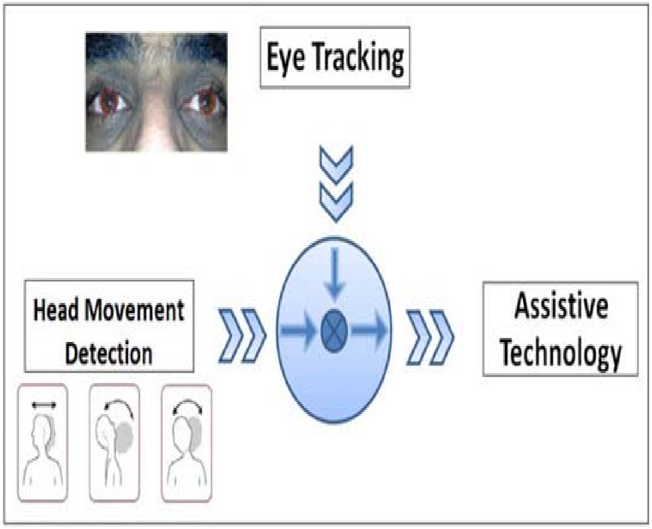

***Fig 1.1:*** *Outline of the system*

In this paper, ANN-based head movement detection with eye tracking. This proposed system aim to estimate the head movements and to locate the eye in the image and then use the obtained information about the eye region and the head pose to estimate the eye gaze direction to be used in various applications, such as HCI, virtual reality, driver assistance systems and assistance for people with disabilities.

The remaining paper is organized as follows: Section 2 discuss about the related works based on eye tracking and head movement. In Section 3, the proposed system has been described which includes system architecture and detailed description of each stage of the proposed system. Section 4 describes the architecture. Section 5 summarizes the contents of this paper.

## II. RELATED WORKS

Real-time Face Detection[1], **p**reprocessing steps will enhance the quality of image. An edge line across the body is found by performing gradient operators and interpolation. Then the projection of the horizontal distance between the edges is calculated for each row. The head and body location, local minima and maxima points are calculated on the projection vector. The projection vector is clustered around head and body using a k-means algorithm. When the head segment is extracted from the image, Bayesian discriminating feature method is used to search the target's face in the region.

A video based head detection and tracking surveillance system[2] is used for retrieving clear human head or face images from real surveillance systems. Human heads are first detected using an adaptive head detector based on Histogram of Gradients (HoG) feature. The motion and appearance information are extracted from the video sequence. Based Bayesian theory, they used two likelihood to evaluate the probability of a detected region represents an actual human head. The false positives are eliminated and the true positives are tracked by the EMD tracking algorithm with SURF points.

A robust eye detection[3] system that uses face detection for finding the eye region. The Circular Hough Transform (CHT) is used for locating the center of the iris. A new method for eye gaze direction classification using Support Vector Machine (SVM) is introduced and combined with Circular Hough Transform to complete the task required.

## III. Proposed System

In addition to increasing system effectiveness in classifying eye gaze direction, determining the head movements is useful and canbe used as a control method. The first stage is face detection. After preprocessing, Viola-Jones detection algorithm is used to obtain data about the face region. This information is used for detecting the head movements and defining the eyes region and calculating the approximate eye radius range to be used. Then the eyes region is extracted and grayscale converted and PCA is performed to extract the feature vector.The benefit for face detection is that the extracted face information can be used in defining the eye region which will be processed. This reduces the number of potential eye candidates detected due to the elimination of the noisy image background. This is fed into ANN to obtain weighted value for a frame. Obtained value is compared with values in storage. In this work a Resilient Back Propagation algorithm (RBP) has been chosen. A set of inputs has been provided together with the desired outputs.

In this work, PCA and MLP are used. Feature extraction and dimension reduction can be combined in one step using principal component analysis (PCA). Artificial Neural Network can be used to identify the most likely class for input video by finding common features between samples of known classes. ANN is robust to errors in training data and has been successfully applied to problem such as face recognition, speech recognition and interpreting visual scenes. Multi-layer perceptron consists of multiple layers of computational units, usually interconnected in a feed-forward way. Each neuron in one layer has directed connections to the neurons of the subsequent layer. In many applications the units of these networks apply a sigmoid function as an activation function.

The proposed system is divided into modules as follows: -

### A. *Video capturing & preprocessing:*

Firstly the video stream from a live webcam is captured. The video can also be uploaded. For the video analysis, the frames first undergo a number of pre-processing steps. The noise in the video should be eliminated for better recognition results.

### B. *Real time face detection:*

The Viola–Jones object detection framework is the first object detection framework to provide competitive object detection rates in real-time proposed in 2001 by Paul Viola and Michael Jones. The features employed by the detection framework universally involve the sums of image pixels within rectangular areas. As such, they bear some resemblance to Haar basis functions, which have been used previously in the realm of image-based object detection.

The combination of three concepts: the Integral Image, AdaBoost classifier and Cascading classifiers leads to effective real-time face detection. The integral image is an intermediate presentation for the image that is used to allow fast evaluation of features. The integral image value at a location x, y is defined by the sum of pixels above and to the left of that specific location:

$$ii(x,y) = \sum_{x1\leq x\, y1\leq y} i(x1,y1)$$

### C. *Head movement detection:*

One of the many purposes for the detection and tracking of head movements is to allow the user to interact with a computer. In such a situation, the user often does not have to rotate her/his head by a large amount to keep eye contact with the relatively small surface of the monitor screen, and the distance between the user and the monitor can be assumed to be approximately constant. It is therefore sufficient for the head tracking system dealing with such a situation to simply track two-dimensional head positions.

To achieve the goal of tracking head movements, it is performed head pose detection for every captured frame in a sequence. By assuming slow, contiguous, and continuous head movements, it can predict the position and orientation of the head pose in the next frame of the video using the detected position and orientation of the head pose in the current image. By comparing consecutive head poses, it can infer, using our head movement model, the occurrence and the type of head movement in the each frame.

### D. *Eye tracking:*

Eye tracking data is collected using either a remote or head-mounted 'eye tracker' connected to a computer. While there are many different types of non-intrusive eye trackers, they generally include two common components: a light source and a camera. The light source (usually infrared) is directed toward the eye.

The camera tracks the reflection of the light source along with visible ocular features such as the pupil. This data is used to extrapolate the rotation of the eye and ultimately the direction of gaze. Also each frame must be converted into gray scale image. A grayscale digital image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information. Images of this sort, also known as black-and-white, are composed exclusively of shades of gray, varying from black at the weakest intensity to white at the strongest. Then this gray scale image is given as input for feature extraction.

### *E.Feature extraction using PCA:*

In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be very redundant, then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

Feature extraction involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

The PCA is performed by computing the eigenvectors and eigenvalues of the covariance matrix. The covariance is determined by the tendency to two random variables that vary together. Covariance, cov(X, Y) = E[E[X] - X] . E[E[Y ] - Y ] where E[X] and E[Y ] denote the expected value of X and Y respectively

Algorithm:

STEP 1: Prepare the Data

$S=\{\Gamma_1,\Gamma_2...\Gamma_n\}$

STEP 2: Obtain the Mean, φ as

$$\varphi = \frac{1}{M}\sum_{n=1}^{M} \Gamma_n$$

STEP 3: Mean is subtracted from original image as

$\phi_i = \Gamma_i - \varphi$

STEP 4: Calculate the Covariance Matrix, C

$$C = \frac{1}{M}\sum_{n=1}^{M} \phi_n\phi_n^T$$

$$= AA^T$$

$A = \{\phi_1, \phi_2 \ldots \phi_n\}$

STEP 5: Eigenvectors and Eigen values of the Covariance Matrix are calculated and principal components are selected.

In this step, the eigenvectors and the corresponding eigen values will be calculated. From M eigenvectors only M' will be chosen, which have the highest eigen values.

The best results are achieved when an expert constructs a set of application-dependent features. Feature extraction and dimension reduction can be combined in one step using principal component analysis (PCA).

### *F.Eye gaze direction using ANN:*

Gaze direction and gaze point is used in interaction with computers and other interfaces and in behavioral research/human response testing to better understand what attracts people’s attraction.

In this module, the eye gaze direction is classified using Artificial Neural Networks. A multilayer perceptron (MLP) is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. A MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training the network.MLP is a modification of the standard linear perceptron and can distinguish data that are not linearly separable.

## IV. System Architectural Design

The real time video of a human is extracted using a webcam. The video can be captured using OpenCV. OpenCV (Open Source Computer Vision Library) is a library of programming functions mainly aimed at real-time computer vision. The captured video must undergo some preprocessing steps. The extracted video will contain some noise as it is extracted in real time. The noise can be eliminated by blurring the video.

Preprocessing, involves the correction of distortion, degradation, and noise introduced during the imaging process. This process produces a corrected image that is as close as possible, both geometrically and radiometrically, to the radiant energy characteristics of the original scene. After face detection and head movement detection, eye movements are tracked using Viola-Jones detector. The video must be converted to grayscale. Grayscale conversion is necessary to reduce the computational complexity. This method combines the use of an ANN both with an image processing algorithm and PCA for feature extraction. Such a choice represents an original approach to the gaze tracking problem. There are some reasons to justify the use of an ANN in a feature-based context. Among them, the most important is the ability of an Artificial Neural Network to generalize, once the right input set has been chosen.

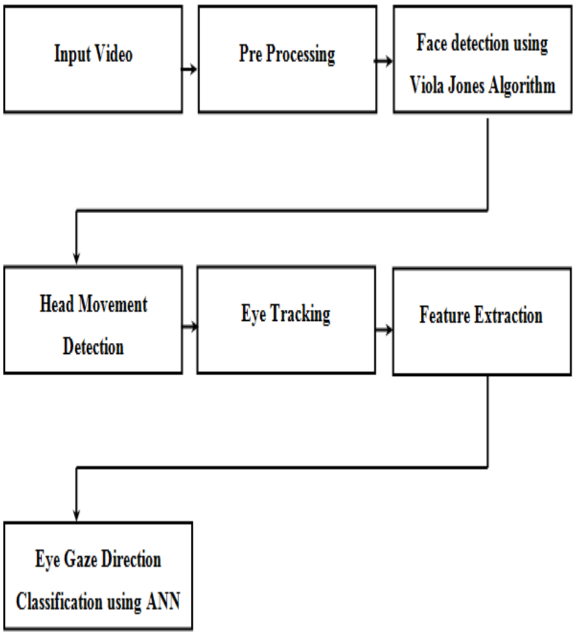
Input Video
Pre Processing
Face detection using Viola Jones Algorithm
Head Movement Detection
Eye Tracking
Feature Extraction
Eye Gaze Direction Classification using ANN

***Figure 4.1: Architecture Diagram of ANN-based head movement detection with eye tracking***

After extracting the motion of head and eye movements both the features must be concatenated. The thus obtained feature vector is of high dimensionality. An effective feature vector should contain only the significant information which carries higher discriminating capacity to formulate the classification task easily. Though inadequate features normally lead to a failure with a good classifier, having too many features may again increase time and space complexities with no guaranteed advantage in the classification process. Therefore, dimensionality reduction is an important step in solving the problem of dimensionality in an efficient manner. When the input data to an algorithm is too large to be processed and it is suspected to be very redundant (e.g. the same measurement in both feet and meters, or the repetitiveness of images presented as pixels), then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. Here the feature extraction is performed using the technique Principal Component Analysis(PCA). It is not possible to maintain the head perfectly still, an appropriate filtering is required to eliminate low-frequency movements and enhance the movements with high velocity, such as the eyelids ones.

The design of an ANN comprises the assessment of the following variables: - typology of net; typology and number of the inputs; typology and number of the outputs; number of hidden layers; and number of neurons for each hidden layer.

A multilayer perception with 2 hidden layers, 45 neurons each, has been chosen. This choice is the result of several trials in different conditions and it is strictly correlated to the typology of the input. To allow the ANN to accomplish the task, an appropriate set of input parameters is needed. These parameters correspond to the geometrical relations between specific features of the eyes: the irises and the eyelids. Such features are detected by a preliminary image processing. Information on the relative position of the iris in the eye socket is given by two distance vectors chosen for each eye, i.e. the distances between the centre of the iris and the corners of the eyelids. Each vector has amplitude and a phase: the amplitude is important mainly for the horizontal movements of the eyes, while the phase for the vertical shifts. In order to have information on the absolute position, orientation and distance of the head, the 2D coordinates of the external corners of the eyes have been added. Considering both eyes, the total amount of the inputs is 12 (4 magnitudes, 4 phases, 2 couples of x-y co-ordinates)

Artificial Neural Network can be used to identify the most likely class for input video by finding common features between samples of known classes. ANN is robust to errors in training data and has been successfully applied to problem such as face recognition, speech recognition and interpreting visual scenes. It is well suited to problem in which the training data corresponds to noise and complex data. The algorithm used for training samples is Back Propagation algorithm. It is a supervised learning method, and is a generalization of the delta rule. It requires a dataset of the desired output for many inputs, making up the training set. It is most useful for feed-forward networks (networks that have no feedback, or simply, that have no connections that loop). Back propagation requires that the activation function used by the artificial neurons (or "nodes") be differentiable.

## V. Conclusion

The proposed system mainly involves face detection, eye and iris detection, classification of eye gaze direction and head flexion detection. This system is used for detecting the head movements and defining the eyes region and calculating the approximate eye radius range to be used. The benefit for face detection is that the extracted face information can be used in defining the eye region which will be processed. This reduces the number of potential eye candidates detected due to the elimination of the noisy image background. This is fed into ANN to obtain weighted value for a frame.

## Acknowledgment

I would like to show my gratitude to Dr. C G Sukumaran Nair (HOD), Associate Professor, Sudha S K and Assistant Professor, Renji S, Department of Computer Science & Engineering, Sarabhai Institute of Science and Technology, for their valuable guidance.

## References

[1] Sung-Kwan Kang, Kyung-Yong Chung, Jung-Hyun Lee [May 2013], Development of head detection and tracking systems for visual surveillance”

[2] D. Xie, L. Dang, and R. Tong, ``Video based head detection and tracking surveillance system,'' in *Proc. 9th Int. Conf. FSKD*, 2012,pp. 2832_2836.

[3] Amer Al-Rahayfeh, Miad Faezipour[December 2013] “Enhanced Eye Gaze Direction Classification Using a Combination of Face Detection, CHT and SVM”

[4] Amer Al-Rahayfeh, Miad Faezipour[November 2013] “Eye Tracking and Head Movement Detection: A State-of-Art Survey”

[5] Z. Zhu and Q. Ji, ``Novel eye gaze tracking techniques under natural head movement,'' *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12,pp. 2246_2260, Dec. 2007.

[6] D. W. Hansen and Q. Ji, ``In the eye of the beholder: A survey of models for eyes and gaze,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3,pp. 478_500, Mar. 2010.

[7] M. Mehrubeoglu, L. M. Pham, H. T. Le, R. Muddu, and D. Ryu, ``Realtime eye tracking using a smart camera,'' in *Proc. AIPR Workshop*, 2011,pp. 1_7.

[8] H. Liu and Q. Liu, ``Robust real-time eye detection and tracking for rotated facial images under complex conditions,'' in *Proc. 6th ICNC*, vol. 4. 2010,pp. 2028_2034.

[9] D. Beymer and M. Flickner, “Eye Gaze Tracking Using An Active Stereo Head,” Proceedings of the IEEE Computer SocietyConference on Computer Vision and Pattern Recog nition,vol 2,pp.451-458, 2003.

[10] G. Y. Zhang, B. Cheng, R.J. Feng: Real-time Driver Eye Detection Method Using Support Vector Machine with Hu Invariant Moments. In:Proceeding of the seventh International on Machine Learning and Cybernetics .pp. 2999-3004, 2008.