Hybrid LASSO and Neural Network Estimator for Gaze Estimation

S Deepthi Iyer
Department of Electronic and
Communication Engineering
MSRUAS, Bengaluru
Karnataka 560058
Email: iyer.deepthi123@gmail.com

Hariharan Ramasangu
Department of Electronic and
Communication Engineering
MSRUAS, Bengaluru
Karnataka 560058
Email: hariharan.ec.et@msruas.ac.in

***Abstract*—Gaze estimation has wide applications in drowsiness detection, security, and biomedical domains. The challenges in estimating the gaze angle include varying light conditions and subtle movements of the gaze. The Convolutional Neural Network (CNN) has recently been suggested as a potential method for gaze estimation. In this present work, we have proposed a gaze estimator combining a neural network and Least Absolute Shrinkage and Selection Operator (LASSO). The features considered are both eye and head features. The combined estimator neural network-LASSO (NN-LASSO) outperforms the individual performance of neural network and LASSO estimator. The results are validated using MPII Gaze dataset and it has been shown that the proposed NN-LASSO estimator outperforms CNN in mean error sense.**

I. Introduction

Gaze estimation is the method of estimating the direction and the angle at which a person is gazing at a particular object. The detection of what a person is looking at by detecting the visual attention of human is used in estimating the gaze. Appearance based gaze estimation is used in many domains, for instance in the visual behavior analysis and computer interactions. There are various methods used in appearance based estimation. The eye and head features are extracted for the gaze estimation. There are algorithms that use only eye features, only head features, and both eye and head features. The gaze estimation problem is also suited for sparse estimation as information content in eye and head features may share similar characteristics. In this respect, regression that gives a sparse relation is a potential approach. This paper explores one such approach i.e. LASSO and investigated the performance of an estimator based on a neural network and LASSO (NN-LASSO). The dataset used for validation is MPII Gaze dataset.

The rest of the paper is organized as follows: The second section gives a brief survey of estimation algorithms for gaze estimation. The proposed method has been outlined in the third section and results are discussed in the fourth section.

II. Related Work

The appearance based gaze estimation deals with the extraction of the eye features and mapping it to estimate the gaze angle. The early work of gaze estimation involves the extraction of gaze angle using information obtained from single eye [1]. The experimentation was carried out using the synthesized image and also the head pose was assumed to be fixed. The early techniques of estimation involved using only the eye features. The evaluation of the eye features and the gaze estimation were carried out using probability distribution of eye parameter to estimate the gaze [2]. The random forest regression technique is another method to estimate the gaze for fixed head poses [3]. Spectral imaging has also been used for gaze tracking [4].

The estimation of gaze using a normal camera by capturing the real time videos was the next stage of development in the gaze estimation. The web-cam for the capturing of face, from which eye features were extracted has also been used [5]. The accuracy in these cases relied majorly on the resolution of the camera used. Gaze was estimated without any calibration where gaze is estimated using the human gaze patterns [6]. The slight head motion has been addressed using the adaptive linear regression technique to estimate gaze angle [7].

Head movement correction was addressed by considering the visual axis of eye through which the gaze was estimated [8]. Transformation matrix has been used to consider head features while estimating the gaze [9]. The application of the gaze was extended to the gaze estimation through mouse operation [10] where gaze angle is estimated by using the location of the mouse.

Convolutional neural network has been used for gaze estimation and has been evaluated based on the MPII gaze dataset. The dataset has been created by considering various lighting conditions. The appearance based gaze estimation carried out by considering both head and eye features is fed onto the CNN to carry out the gaze estimation on MPII dataset where the gaze angle is obtained over different lighting conditions [11] [12]. The estimation of gaze using eye shape registration has also been validated using MPII dataset [13][14].

The summary of performance of a few gaze techniques for the MPII gaze dataset has shown mean errors upto 3.5-4 degrees [11]. The proposed NN-LASSO algorithm gives lesser mean error for MPII Gaze dataset.

978-1-5090-2597-8/16/$31.00 ©2016 IEEE

III. Proposed Method

The overview of the gaze estimation using CNN is shown in Figure 1. The inputs to the estimator are the eye and head features, and the output is the corresponding 2D gaze angle.

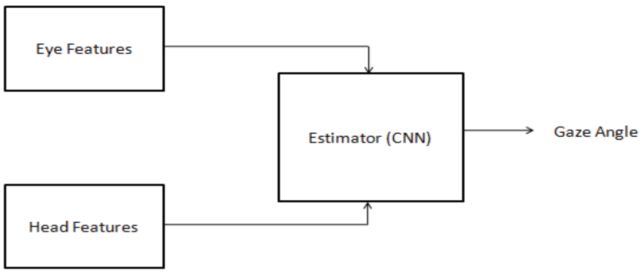

Fig. 1. Existing Gaze estimation block diagram

The steps involved in the proposed NN-LASSO based gaze estimation is shown in Figure 2. The proposed gaze estimation algorithm is validated using the MPII Gaze dataset.

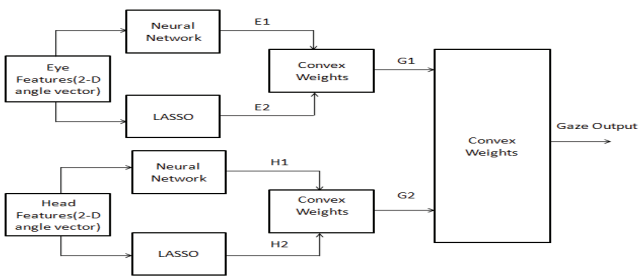

Fig. 2. Proposed Gaze estimation block diagram

The CNN estimation is replaced by the hybrid combination of a neural network and LASSO estimators. The way in which the features are used in the existing CNN method are modified in NN-LASSO method by using convex weights. In this work, we use the extracted eye and head features as reported in MPII Gaze dataset directly. The method in which these features are extracted is briefly discussed.

A. Face Alignment and 3D Head Pose Estimation

The face detection is carried out by applying the SURF cascade technique. The 3D position of facial landmarks are used to detect the head pose. The facial landmarks include the two eye corners and the corners of the mouth region. The head coordinate system is defined according to the triangle connecting three midpoints of the eyes and mouth, and further refining of the pose is carried out by using nonlinear optimization technique. Once the head pose is obtained, the image and the head pose is normalized to the polar coordinate angle space to carry out further evaluations. The head and the eye features in polar coordinates are used to evaluate the gaze estimation algorithms.

B. Estimators

The estimator used in the proposed technique is the hybrid weighted combination of Neural Network (NN) and LASSO estimators. The neural network used has 20 neurons and the neural network is trained by using the head and the eye features.

The other estimator used in the system is the LASSO estimator. The gaze estimation using LASSO estimator is carried out by the following equations.

$$min_\beta h(\beta) = \frac{1}{2}||E2 - X\beta||_2^2 + \lambda||\beta||_1 \tag{1}$$

$$E2 = X * min_\beta h(\beta) \tag{2}$$

where
β is a px1 vector
X is a nxp matrix that represents the input eye features
h is the LASSO criterion
λ is the penalizing factor that encourages sparsity
$||.||_2$ represents 2-norm
$||.||_1$ represents 1-norm
E2 is a nx1 neural network output vector that represents gaze angle estimated using eye features

The eye features are given as input to both the estimators and the corresponding output gaze angle values are combined by using the following optimization equation

$$G1 = w1 * E1 + w2 * E2 \tag{3}$$

where
w1 and w2 are convex weights (w1+w2=1)
G1 is the estimated gaze angle using eye features
E1 is the gaze estimation using neural network estimator for eye features
E2 is the gaze estimation using LASSO estimator for eye features

The evaluation of the gaze angle using the head features is same as the gaze angle estimation carried out using the eye features. The head features are given to the estimators and the corresponding gaze angle values obtained are combined by using the following optimization equation

$$G2 = w3 * H1 + w4 * H2 \tag{4}$$

where
w3 and w4 are convex weights (w3+w4=1)
G2 is the estimated gaze angle using head features
H1 is the gaze estimation using neural network estimator for head features
H2 is the gaze estimation using LASSO estimator for head features

The gaze angle is obtained by combining G1 and G2.

$$G_{output} = w5 * G1 + w6 * G2 \tag{5}$$

where

$w5$ and $w6$ are convex weights (w5+w6=1)
$G1$ is the estimated gaze angle using eye features
$G2$ is the estimated gaze angle using head features
G_{output} is the final estimated gaze angle using head and features

The actual gaze is an angle vector represented by θ and ϕ. The estimated value corresponds to the estimated angle vector represented by $\hat{\theta}$ and $\hat{\phi}$. The error evaluation is carried out by determing the mean error between the estimated value and the actual value. The neural network estimator gives a good performance for gaze estimation and LASSO eliminates the over-fitting problem, hence the combination of the two estimator using the convex weight provides a good performance for gaze estimation.

IV. Results and Discussions

A. MPII Gaze Dataset

MPII gaze dataset is the dataset created for estimating the gaze angle. The dataset was implemented by using software running at the background of the participants laptop where the images were captured every 10 minutes. The image capturing was followed by the procedure where the participant is asked to look at random sequence of 20 onscreen positions. The participants concentration was confirmed by asking the participants to press the space bar once circle onscreen was about to disappear. The features included in the MPII dataset were that the dataset is ideally used to calculate gaze angle using the head and the eye feature information. The dataset comes along with a normalized folder that holds the normalized values of the data being used. The dataset also holds a data folder that holds the cropped eye images and an annotation file that has the eyes and head feature and the target (gaze angle) values for every participant, every day wise. The annotation file for each image has 35 features and the features. Features 1-24 represents the eye features, 25-26 represents the onscreen target values, 27-29 represents the 3D target values in camera space and the features 30-35 represents the facial landmarks that represents the head pose.

The dataset is created by considering various eye images oriented at different angles from 15 participants over a period of several days per participant. There are a total of 2,13,659 images collected from 15 participants with the number of images collected per participant varying from 34,745 to 1,498. The dataset has participants wearing spectacles and participants without wearing spectacles. The capturing procedure is also carried out with various illumination conditions, thereby estimation of gaze angle using the MPII gaze dataset accounts for the illumination conditions. The input features are eye and the head features taken separately which is fed into proposed estimators block and the output gives the gaze angle. Leave one out cross validation has been performed repeatedly and the average results have been given. The performance of the system is measured by estimating the mean error. The labels or the desired outputs available are in 3-D vector format represented by (x,y,z), which is converted into 2-D angle represented by (θ,ϕ) by following the formulas as shown in equation 6 and 7 [11].

$$\theta = sin^{-1}(-y) \tag{6}$$

$$\phi = tan^{-1}\frac{-x}{-y} \tag{7}$$

B. Result Analysis

The performance analysis parameter used to evaluate the system performance is by evaluating the mean error of the system and the formula used to evaluate the mean error is as follows:

$$Er_{mean} = mean(Er) \tag{8}$$

$$Er = \sqrt{(m_1 - m_2)^2 + (a_1 - a_2)^2} \tag{9}$$

where
m_1 is actual magnitude value
m_2 is the estimated magnitude value
a_1 is the actual direction value
$a_2$is the estimated direction value
Er_{mean} is the error obtained by averaging over all samples
Er is the error

The values of mean error for various features using various estimators is shown in Table 1. The estimation is carried out by passing the head features alone as input into the estimator and checking the corresponding mean error, the similar method is carried out in testing the eye features alone as input and checking the corresponding mean error. The next case is carried out by combining the head and the eye features as single input and passing it through the estimator. The final case is carried out by feeding the eye and the head features separately and combining it through the convex weights. It is observed that the proposed estimator method performs better with the mean error of 3.4 degrees by combining the head and the eye features through convex weights. The performance of the proposed system is compared with the existing systems and the mean errors of various techniques is shown in Figure 2.

TABLE I
MEAN ERROR OBTAINED FOR VARIOUS FEATURES USING DIFFERENT ESTIMATORS

Features	Only NN	Only LMS	Only LASSO	Proposed method
Head	8	13.8	9	6.5
Eye	5	12	6	4.5
Eye and Head (without convex weights)	4.8	11.4	5.8	4
Eye and Head (with convex weights)	4.2	10.9	5.2	3.4

It is observed from Table 2 that the proposed method outperforms the existing methods available which are applied on the MPII Gaze dataset. The proposed method has mean

TABLE II
MEAN ERROR OBTAINED COMPARING VARIOUS TECHNIQUES

Techniques	Mean Error
CNN on MPII [11]	3.8
Open Face [15]	9.9
Proposed method	3.4

error of 3.8 degrees whereas the CNN method has 3.8 and Open face method has 9.9 degrees mean error. The mean error obtained using the eye shape registration method on the MPII Gaze dataset varies between 7 to 16 degrees [13].

V. Conclusion

The hybrid estimator model comprising weighted combination of LASSO and Neural Network estimators to estimate the gaze angle using the head and eye features has been proposed. The gaze angle was also estimated using the neural network, LASSO and the LMS estimators separately and their corresponding mean errors are analyzed. It has been shown that the performance of the proposed system is better in comparison to the other estimators, and the proposed estimator also outperforms the CNN based gaze estimation algorithm. The performance of the proposed NN-LASSO for other datasets have to be investigated, and also extending to a larger datasets, where CNN may perform better, is planned for future work.

References

[1] J, Wang, E, Sung, and R, Venkateswarlu. Eye gaze estimation from a single image of one eye. *In Computer Vision, 2003. Proceedings. Ninth IEEE International Conference* , (pp. 136-143). 2003,October.

[2] J, Chen and Q, Ji. Probabilistic gaze estimation without active personal calibration. *In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference* , (pp. 609-616). 2011, June.

[3] Y, Sugano, Y, Matsushita and Y, Sato. Learning-by-synthesis for appearance-based 3d gaze estimation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 1821-1828). 2014.

[4] Gebejes, Ana and Fält, Pauli and Bednarik, Roman and Hauta-Kasari, Markku. SPEED: SPectral eye vidEo database *In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (pp. 662-671) 2016.

[5] R, Valenti, N, Sebe, T, Gevers and J, Staiano. Webcam-based visual gaze estimation. *In Image Analysis and ProcessingICIAP 2009*, (pp. 662-671). 2009.

[6] F. Alnajar, T. Gevers, R. Valenti, and S. Ghebreab. Calibration-free gaze estimation using human gaze patterns. *In Proceedings of the IEEE International Conference on Computer Vision* , (pp. 137-144). 2013.

[7] F, Lu, Y, Sugano, T, Okabe, and Y, Sato. Adaptive linear regression for appearance-based gaze estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 36(10)* , (pp.2033-2046.). 2014.

[8] Z, Zhu and Q, Ji. Novel eye gaze tracking techniques under natural head movement. *Biomedical Engineering, IEEE Transactions on, 54(12)* , (pp.2246-2260). 2007.

[9] R, Valenti, N, Sebe, and T, Gevers. Combining head pose and eye location information for gaze estimation. *Image Processing, IEEE Transactions on, 21(2)*, (pp.802-815.). 2012.

[10] Y, Sugano, Y, Matsushita, Y, Sato, and H, Koike. Appearance-Based Gaze Estimation With Online Calibration From Mouse Operations. *Human-Machine Systems, IEEE Transactions on, 45(6)* , (pp.750-760.). 2015.

[11] X, Zhang, Y, Sugano, M, Fritz and A, Bulling. Appearance-based gaze estimation in the wild.*In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* , (pp. 4511-4520). 2015.

[12] F, Lu, Y, Sugano, T, Okabe, and Y, Sato. Gaze Estimation From Eye Appearance: A Head Pose-Free Method via Eye Image Synthesis. *Image Processing, IEEE Transactions on, 24(11)* , (pp.3680-3693.). 2015.

[13] E, Wood, T, Baltruaitis, X, Zhang, Y, Sugano, P, Robinson and A, Bulling. Rendering of eyes for eye-shape registration and gaze estimation. *2015 IEEE International Conference on Computer Vision (ICCV)*, (pp. 3756–3764). 2015,March.

[14] E, Wood, T, Baltruaitis, P, Robinson and A, Bulling. Learning an Appearance based gaze estimator from one million synthesized images.*In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications* 2016.

[15] T, Baltruaitis, P, Robinson and L, P,Morency. OpenFace: an open source facial behavior analysis toolkit. *In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV).* 2016.