A ROBUST EYE GAZE ESTIMATION USING GEOMETRIC EYE FEATURES

Krupa Jariwala
Assistant Professor,
Computer Engineering,
SVNIT, Surat, INDIA
knj@coed.svnit.ac.in

Dr. Upena Dalal
Associate Professor,
Electronics Engineering,
SVNIT, Surat, INDIA
udd@coed.svnit.ac.in

Amal Vincent
Computer Engineering,
SVNIT, Surat, INDIA
amal.vincent1993@gmail.com

***Abstract*— Gaze estimation is the process of determining the point of gaze in the space, or the visual axis of an eye. It plays an important role in representing human attention; therefore, it can be most appropriately used in Human Computer Interaction as a means of an advance computer input. Here, the focus is to develop a gaze estimation method for Human Computer Interaction using an ordinary webcam mounted on the top of the computer screen without any additional or specialized hardware. The eye center coordinates are obtained with the geometrical eye model and edge gradients. To improve the reliability, the estimates from two eye centers are combined to reduce the noise and improve the accuracy. Facial land marking is done to identify a precise reference point on the face between the nose. The ellipse fitting and RANSAC method is used to estimate the gaze coordinates and to reject the outliers. This approach can estimate the gaze coordinates with high degree of accuracy even when significant numbers of outliers are present in the data set. Several refinements such as feedback and masking, queuing and averaging are proposed to make the system more stable and useful practically. The results show that the proposed method can be successfully applied to commercial gaze tracking systems using ordinary webcams.**

Keywords—eye tracking, computer vision, image processing

I. INTRODUCTION

Gaze estimation intuitively plays an essential role in representing human attention, feeling and desire, therefore, it can most appropriately be used in Human Computer Interaction as a means of advance computer input and in the analysis of visual scanning pattern to obtain person's attentional focus. Figure 1 shows the eye structure and the optical and visual axis of an eyeball to determine the gaze. The optical/pupillary axis of the eye is a line connecting the center of the pupil and the center of the cornea. The visual axis is the line connecting the corneal center and the center of the fovea, which is the highest acuity region of the retina. The optical axis is the most direct line through the center of the cornea to the pupil, the lens, and the retina, however, this line intersects the retina below the fovea and is not the most light and color sensitive. The visual axis is the line connecting the fixation point with the object nodal point of the eye, which is the line from the center of the pupil to the fovea. The visual axis gives the best color vision. Since the gaze point is defined as the intersection of the visual axis rather than the optical axis with the scene, the angle between the optical axis and the visual axis is called kappa (κ) as shown in Figure 1. The kappa (κ) is a constant value for each person and can be best modelled through a personal calibration.

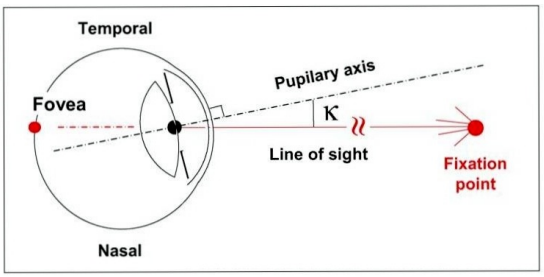

Fig. 1. The visual axis, Kappa and fixation point[1].

II. PROBLEM STATEMENT

There are limited studies focusing on the challenges in webcam-based applications and much of the existing eye tracking research [2][3[4] employed a high resolution camera with a wide angle lenses, which is often set up on the desk below the screen of the computer. Thus, the camera is below the face, and its orientation is an angle of elevation as shown in Figure 2(a). This setup method can avoid the influence of the eyelid movements and handles the high resolution images for better accuracy[18]. However, general users often set up their webcam on the top of the monitor as shown in Figure 2(b), therefore, instead of focusing a specific environment employing a high-quality camera or expensive eye tracking equipment, emphasis is on the eye tracking problem in a general computer use environment for common users with a desktop computer and a low-resolution webcam attached to the monitor and the major light source is the ceiling light. Since, a general low resolution is used and the distance between the webcam and the user is limited to 40–60 cm, which is a usual computer operating distance to guarantee the minimum resolution of an extracted eye.

In order to correctly estimate the visual gaze, it would be reasonable to consider the head position and orientation to give a rough initialization of the visual gaze, and then use the information about the eye centers to fine tune the information.

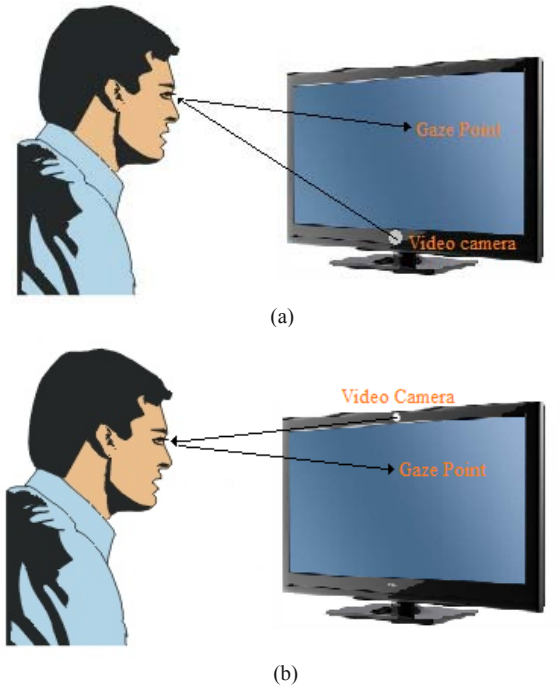

Fig. 2. Environmental setup (a) Existing research (b) Proposed research.

However, the problem of head pose estimation is high dimensionality and involve 6 degree of freedom for yaw, pitch and roll, in order to achieve a realistic modelling. Moreover, it requires high computational processing and excessively large number of training samples, which are not in line with the lightweight requirements of our system. Additionally, small mistakes in pose estimation might introduce additional errors in the final visual gaze estimation.

Therefore, to effectively solve this problem while significantly reduce the training cost, the gaze direction under fixed head pose with a limited head movement is estimated, where the face is always frontal to the camera so the head pose information can be discarded. However no restriction is applied in terms of chin rest and unnaturally fixed head pose.

In all, the aim is to achieve a robust, real-time eye-gaze estimation scheme, which is computationally efficient and have high performance for low-resolution webcam mounted on a consumer grade computer which can be used practically for any eye gaze based computer application such as human computer interaction, learner behaviour analyses for educational psychology, marketing and research etc.

III. PROPOSED METHOD : ROBUST EYE GAZE ESTIMATION USING GEOMETRIC EYE FEATURES

The eye centre coordinates (X_l,Y_l) and (X_r,Y_r) are obtained using eye centre localization method[5] based on geometric eye model and edge gradients. These coordinates are further used to estimate gaze using proposed sequences of steps as per following:

A. Eye Center Stability Improvement

Since, both eye pupils move together, their location remain the same in both the images. The quality of the pupil centre estimate can therefore be improved by combining the estimates from the two eyes. To do this, the probability estimates from the two eyes are overlaid and the two probabilities are multiplied. Hence, (X_r,Y_r) and (X_l,Y_l) of both the eyes are combined and the average eye centre coordinates X_{cavg} and Y_{cavg} are estimated. This will reject the noise and the wrong estimates for the average eye centre coordinates and improve the accuracy[6].

B. Reference point selection

Knowing the coordinates of the obtained eye centres are not enough by itself to find the gaze direction. In order to attach the meaning to the pupil coordinates (X_l,Y_l) and (X_r,Y_r), the pupil coordinates needs to be expressed in terms of an offset from some reference point on the face. This reference point needs to be steady to be able to reliably map the eye gaze direction relative to the reference point. Various approaches exists that include marking the centre of an eye with the dot[6], obtaining a distinctive image features[7][8] and detecting limbus boundary points[9]. These approaches are either unviable, computationally heavy or require eyes to be extremely wide open.

Therefore, a facial land marking technique[10] is proposed that ensemble regression trees to regress the location of facial landmarks from a sparse subset of intensity values extracted from an input image. This framework is much faster in reducing error compare to other methods[11][12] and can also handle partial or uncertain labels. This method treat different target dimensions as independent variables to take advantage of the correlation of shape parameters for more efficient training and better use of partial labels. Thus, the topmost point of the nose is derived efficiently based on the facial land marking method[10] and a steady reference point (X_{ref}, Y_{ref}) on the nose, between the eye is identified in an accurate and computationally efficient way. The offset eye coordinates (X_{offset}, Y_{offset}) are then computed as per eq (1) relative to this reference point, which can be used to map the precise gaze direction.

$$X_{offset} = Xc_{avg} - X_{ref} \quad \text{and} \quad Y_{offset} = Yc_{avg} - Y_{ref} \quad (1)$$

C. Quadratic Curve Fitting

Here, the pupil is present on a spherical surface but the camera perceives it as a projection of the pupil movement on the flat plane, therefore, the movement of an eyeball around the corner of the eyes can be correlated with the large change in the gaze coordinates as compared to the movements at the center position. The same correlation can be applied to the vertical movements as well. Therefore, the feature based interpolation method [6][13] based on the offset eye coordinates, can be appropriately applied to map the X and Y coordinates corresponding to the gaze on the computer screen by means of second degree polynomial curve fitting. The quadratic equation for this method is formulated as per eq(2).

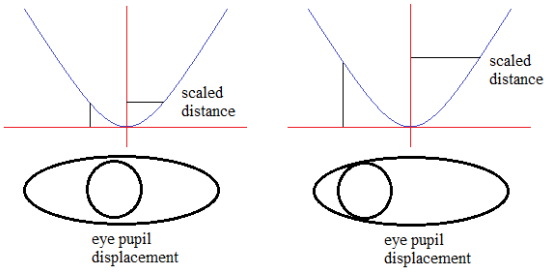

Fig. 3. Eye pupil displacement according to the quadratic equation..

$$g(x,y) = Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F \quad (2)$$

A least squares solver can suitably map the gaze point to the screen according to the quadratic equation as per eq (2). Figure 3 indicates the mapping of gaze point according to curve fitting algorithm to obtain the pupil displacement.

D. Gaze estimation

The aim of the proposed system is to provide a feasible methodologies for the applications which only need general computer and low-resolution webcams where the goal of those applications is only to know which rough "area" the user is paying attention to and not which accurate location the user is looking at. For this setup, since the camera is at the top of the computer, the shape of the extracted eye area when looking at the upper side is different from that when looking at the lower side. The shapes of the iris area also are different while looking at different positions. Moreover, since the eye image is extracted through a low-resolution camera, linear interpolation is not suitable in our scenario. Therefore an eye-gaze model is constructed based on the training mechanism. To map the real-world coordinate system to the 2D screen, the eye-gaze position is calibrated by asking the user to gaze at 9 screen locations (3x3 grid), as illustrated in Figure 4.

A correlation needs to be established with a user click and a crosshair location of the mouse cursor to generate a training data set. A data gathering tool (Pygame library in python) is used for this purpose where, the algorithm records the crosshair location of mouse cursor generated using a click and the corresponding current estimate of pupil to produce one data point for the training set. After generating several training data points, least squares fit method (Numpy library in Python) is used to produce a transformation matrix *H,* which relates pupil offset coordinates (X_{offset}, Y_{offset}) to screen gaze position (X_g,Y_g) using the quadratic features. Therefore the input feature vector is *H* is derived as per eq (3).

$$H = [X_{\text{offset}}, Y_{offset}, 1, X_{\text{offset}}^2, Y_{offset}^2, X_{\text{offset}}Y_{offset}] \quad (3)$$

The least mean squares error solver (Numpy) is used to train the matrix *H* such that,

$$(X_g, Y_g) = \text{feature} * H \quad (4)$$

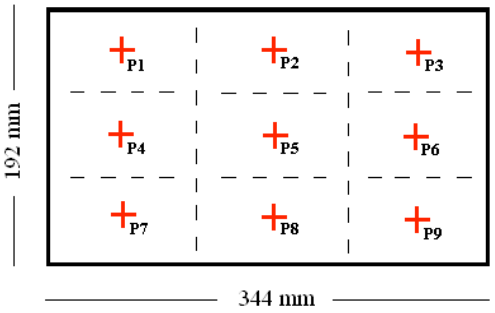

Fig. 4. 9 screen locations (3x3 grid).

However, the least squares fit is strongly degraded if there are outliers, as it will generally optimally fitted to all points, including the outliers. These outliers can come from extreme values of the noise or from erroneous measurements or incorrect hypotheses about the interpretation of data, such as when the user's eyes were closed or the pupil position is incorrect in training data in our case. RANSAC, on the other hand, can produce a model, which is only computed from the inliers, provided that the probability of choosing only inliers in the selection of data is sufficiently high. Therefore, RANSAC (Random Sample Consensus) [14] [15] algorithm is proposed to be used to reject outliers. A random small subset of the training data is chosen for training an H matrix, and using only the training data that is well described by H to refine H. Many such H candidates are computed and the best-performing one is chosen.

E. Proposed Refinements : Gaze Estimate Stability Improvement

In a constant lighting condition, the gaze output using the proposed method is estimated correctly most of the time. However, sometimes the gaze estimate is unstable and practically not useful due to the fact that the slight change in the eye center localization due to changing background and light intensity causes a large change in the output screen coordinate. To overcome this problem, several refinements are proposed to produce stable gaze output as following:

Feedback and Averaging :

Though the eye center localization is independent of the previous frame, it is observed that the change in the eye center coordinate is gradual in successive frames. Only under significant unfavorable lighting condition, the next predicted eye center is far away from its previous successive frames. To avoid this problem feedback and average mechanism is used. The mask region (pixels having maximum votes) from previous frame is resized and applied to the current frame as a feedback so that gaze estimation is stable even in case of changing lighting condition. Further if every estimated eye center offset is directly mapped to the screen, the gaze

becomes unstable. Therefore the output of the estimated offset is a weighted average of the previous offset value and current offset value as per eq (5) where W indicates the weight, which is empirically set to 0.4. Similar feedback is also applied while making the combined eye center estimate to further enhance the accuracy and stability.

$$offset_{new}=(1-W)*offset_{Prev}+W*offset_{current} \quad (5)$$

IV. RESULTS AND ANALYSIS

The gaze should be accurate to be useful in analysis and interaction. For pervasive use, the accuracy should be stable over the device with varying usage conditions. Therefore, an experiment is carried out to measure the proposed system's accuracy over varied participants and realistic usage and lighting conditions.

A. Evaluation measures

The experimental setup as per Figure 5, similar to Ferhat[20] is prepared to test the performance of the system such that the display of the system mounted on a support enables the subject to face the center of the display directly. The camera is on the top of the display at location A, on the center of the screen. The approximate distance D between C and E is 40 cm, which is kept as a normal operating distance of a person with a computer. The estimation accuracy is measured in degrees, which is computed as per equation (6).

$$Error = \left| \arctan\left(\frac{D_{XC}}{D_{EC}}\right) - \arctan\left(\frac{D_{X'C}}{D_{EC}}\right) \right| \quad (6)$$

Where x is the correct and x' is the estimated gaze point, C is the screen center and E is the face center point on the nose, which is a reference point between the eyes on the nose derived based on facial land marking[10]. The variables D_{xC}, $D_{x'C}$ and D_{EC} denote the corresponding distances between the specified points. Similar calculation for vertical error is performed along the y axis. For the set of N points, the root-mean-squared error RMSE is used to measure the error between the true and the estimated gaze points.

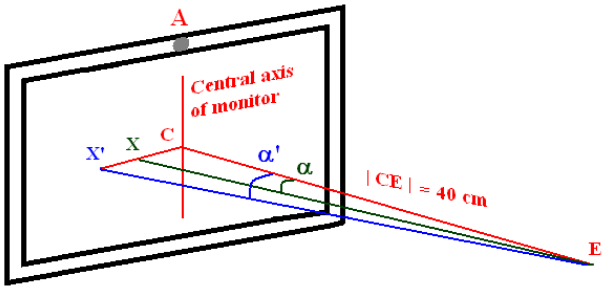

Fig. 5. Experimental setup for the proposed gaze estimation method.

B. Experimental results

One time system calibration with adequate lighting conditions and a stationary head is performed for the system having 1366*768 of screen resolution and 640 x 480 of webcam resolution by clicking on the screen at 9 screen locations (Fig. 4) and looking at the mouse pointer.

With the proposed approach the system gets quicker stability with approximately 13 clicks. The training is carried out with approximately 15 clicks to get a fix on the gaze estimate. Table 1 shows the effect of training duration on the system accuracy. Table 2 shows the obtained results for the ground truth gaze (X_i,Y_i), estimated gaze (X_g, Y_g) and the error along X axis and Y axis in terms of degree(X^o,Y^o) and pixels(X_p,Y_p) using the proposed gaze estimation method.

Figure 6 shows the true click position (Red) and the system's predicted eye gaze location (green) at 9 points as per the results of table 2. The results indicate that the system performs fairly well along the X axis however most of the error is in the vertical direction is much evident which is due to the positioning of the camera over the top of the computer screen and the system has trouble locating the exact center of the pupil vertically while looking down. Figure 7 shows the estimation error distribution for 120 gaze points, along the X axis and Y axis. The distribution is highly centered at zero indicating high accuracy of the proposed method.

With the proposed test setup the horizontal mean square error of 56.8 pixels for the x estimate and 91.8 pixels for the y estimate without any headrest and slight head movement is achieved. This corresponds to 1.8° horizontally and 2.7° vertically which is highly competitive with the commercial systems available in the market that claim the accuracy of 2° to 4°.

Table 3 shows the comparison of the proposed system with some existing eye gaze estimation methods. It is evident that the proposed method has low computational complexity and high accuracy that makes it competent enough for commercial application of the technology.

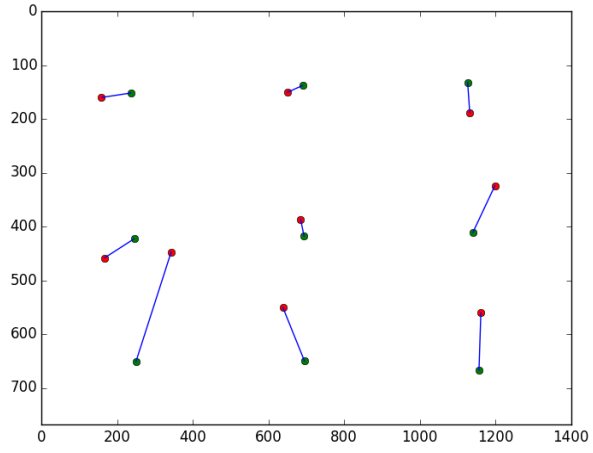

Fig. 6. Eye gaze estimation using proposed method at 9 points with 30 clicks for true click position (Red) and predicted(green)

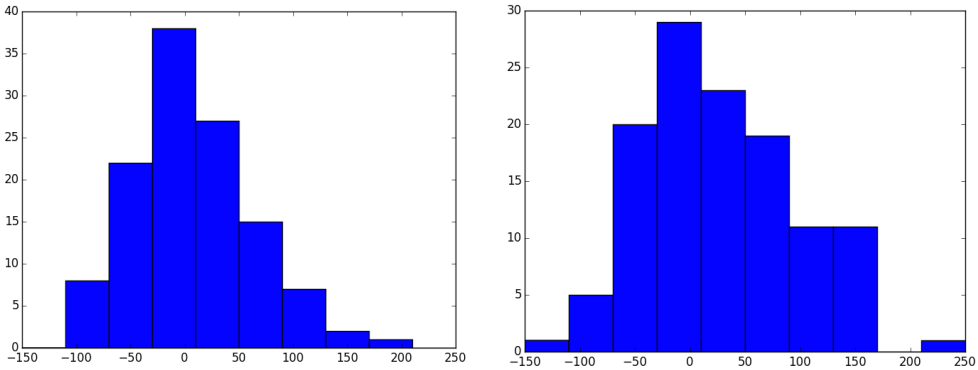

Fig. 7. Eye gaze estimation error distribution along the (a) x axis and (b) y axis.

TABLE I. THE GAZE-ESTIMATION PERFORMANCES FOR VARIOUS TRAINING DURATIONS.

Training duration (no of clicks)	MSE X^o	MSE Y^o	MSE X-pixel	MSE Y-pixel
10	-	-	-	-
15	2.8^o	5.2^o	90.6	168.7
20	2.5^o	3.7^o	75.5	124.8
25	1.8^o	3.81^o	64.8	126.2
30	1.8^o	2.7^o	56.8	91.8
35	1.7^o	3.1^o	52.8	105.0

TABLE II. THE ACCURACY OF EYE-GAZE ESTIMATION FOR 9 GAZE POINTS WITH 30 CLICKS.

Gaze point	Ground Truth X_i	Ground Truth Y_i	Estimated Gaze X_g	Estimated Gaze Y_g	Error in degrees X^o	Error in degrees Y^o	Error in pixels X_p	Error in pixels Y_p
1	236	152	159.2	160.3	2.5^o	0.3^o	76.8	-8.3
2	690	138	649.6	150.9	0.9^o	0.4^o	40.4	-12.9
3	1126	132	1131.6	188.3	0.2^o	1.9^o	-5.6	-56.3
4	1140	411	1198.1	324.1	1.9^o	2.8^o	-58.1	86.9
5	694	418	684.9	387.2	0.3^o	1.0^o	9.1	30.8
6	245	423	166.6	458.6	2.6^o	1.1^o	78.4	-35.6
7	250	651	343.0	447.4	3.1^o	6.0^o	-93.0	203.6
8	695	650	638.1	550.7	1.2^o	2.8^o	56.9	99.3
9	1156	667	1161.0	560.5	0.2^o	3.0^o	-5.0	106.5
				RMSE	1.8^o	2.7^o	56.8	91.8

TABLE III. THE COMPARISON OF THE PROPOSED METHOD WITH SOME EXISTING WORK.

Method	**Video resolution**	**Mean Error (Degree)**	**Environment**	**Computational load**	**Head Pose**	**Distance**
Piratla[16]	640x480	N/A	Lab	Very High	Free	30-60 cm
Valenti[17]	N/A	N/A	Lab	High	Constrained	N/A
Lin[18]	640x480	N/A	Lab	High	5 Poses	50-70 cm
Allen[6]	640x480	2.8^o	Lab	High	Highly Constrained	20-60 cm
Woods[19]	1280x720	6.8^o	Lab	Low	Free	20 cm
Proposed method	1280x720	2.25^o	Lab	Low	Constrained	40-60 cm

ISBN: 978-1-4673-9379-9 ©2016 IEEE

V. CONCLUSION

The problem of gaze estimation is addressed using an ordinary webcam mounted on the top of the computer screen without any additional or specialized hardware. The eye center coordinates are obtained with the proposed geometrical eye model and the edge gradients. The Quadratic curve Fitting method and RANSAC is used to estimate the gaze coordinates and to reject the outliers.

Identification of a stable reference point on the face using facial land marking resulted in substantial improvement in the eye gaze estimation complexity and relaxation of head movement. The number of required calibration points are also reduced considerably and the system responds almost instantaneously. With further proposed improvements such as combining the two eye center estimates, feedback and averaging, the gaze output is stable even in varying lighting condition and estimated correctly most of the time.

The proposed system is non-intrusive, user independent, and works in visible spectrum with inbuilt webcam mounted on the top of the screen. It neither requires any external high resolution camera, nor does the camera require to be mounted at the eye level or at the bottom of the screen to improve the accuracy as other methods. It also does not require very high end technical specification, which may include graphic card, a large RAM or fast processor.

The proposed method clearly outperforms the existing methodologies in terms of speed, efficiency and accuracy that make the practical deployment of the proposed gaze estimation system realistic.

ACKNOWLEDGMENT

We thank SVNIT for the whole hearted support in this project.

REFERENCES

1. http://pabloartal.blogspot.in/2008/08/on-definition-of-angle-kappa.html
2. S. Asteriadis, D. Soufleros, K. Karpouzis, S. Kollias, "A Natural Head Pose and Eye Gaze Dataset", International Conference on Multimodal Interfaces (ICMI 2009)
3. Hansen D.W, Hansen PJ, Nielsen M, Johansen AS, and Stegmann, "Eye typing using Markov and active appearance models." Sixth IEEE Workshop Appl Comput Vis pp. 132–136, 2002.
4. Magee JJ, Betke M, Gips J, Scott MR, Waber BN, "A human-computer interface using symmetry between eyes to detect gaze direction", IEEE Trans Syst Man Cybern 38(6):1248–1261, 2008.
5. K N Jariwala , A Nandi , U D Dalal, "A Real Time Robust Eye Center Localization using Geometric Eye Model and Edge Gradients in Unconstrained Visual Environment", *International Journal of Computer Applications* 128(1):22-27, October 2015.
6. L. Allen, A.Jenson, CS221 Project Report: "Webcam-based Gaze Estimation", Stanford University, 2013.
7. D. G. Lowe, " Distinctive Image Features from Scale-Invariant Key Points, 2004.
8. Herbert Bay," SURF : Speeded Up Robust Features, " Computer Vision and Image Understanding, Vol 110, No-3, 2008.
9. E. Wood and A. Bulling, "EyeTab: model-based gaze estimation on unmodified tablet computers, in proceeding of the Symposium on Eye Tracking Research and Applications - ETRA 2014.
10. V. Kazemi, J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression trees", Proceedings of CVPR - 2014.
11. P. Doll´ar, P. Welinder, and P. Perona, "Cascaded pose regression", In CVPR, pp 1078–1085, 2010.
12. X. Cao, Y. Wei, F. Wen, and J. Sun, "Face alignment by explicit shape regression", In CVPR, pp 2887–2894, 2012.
13. R. Urano, R. Suzuki, and T. Sasaki, "Eye Gaze Estimation Based on Ellipse Fitting and Three-Dimensional Model of Eye for Intelligent Poster", IEEE International Conference on Advanced Intelligent Mechatronics (AIM), July 2014.
14. Fischler, M. and Bolles, R., "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography", *Communications of the ACM*, vol. 24, pp. 381-395, 1981.
15. http://scikit-learn.org/stable/modules/linear_model.html
16. N. Piratla and A. Jayasumana, "A neural network based real-time gaze tracker, Journal of Network and Computer Applications 25, 179-196, 2002.
17. Roberto Valenti1, Jacopo Staiano1, Nicu Sebe2, and Theo Gevers, "Webcam-Based Visual Gaze Estimation, Image Analysis and Processing – ICIAP 2009.
18. Y. Lin, R. Lin, Yu-Chih Lin & G. Lee, "Real-time eye-gaze estimation using a low-resolution webcam", Multimed Tools Appl, 65:543–568, 2013.
19. E. Wood, A. Bulling, "EyeTab: Model-based gaze estimation on unmodified tablet computers", In Proceedings of the Symposium on Eye Tracking Research and Applications, ETRA 2014.
20. O.Ferhat, F. Vilarino, F. J. Sanchez. "A cheap portable eye-tracker solution for common setups", In Journal of eye movement research 7(3):2, 1-10, 2014.