Gaze Tracking by Joint Head and Eye Pose Estimation Under Free Head Movement

1st Stefania Cristina
Department of Systems and Control Engineering
University of Malta
Msida, Malta
stefania.cristina@um.edu.mt

2nd Kenneth P. Camilleri
Department of Systems and Control Engineering
University of Malta
Msida, Malta
kenneth.camilleri@um.edu.mt

***Abstract*—Recent trends in the field of eye-gaze tracking have been shifting towards the estimation of gaze direction in everyday life settings, hence calling for methods that alleviate the constraints typically associated with existing methods, which limit their applicability in less controlled conditions. In this paper, we propose a method for eye-gaze estimation as a function of both eye and head pose components, without requiring prolonged user-cooperation prior to gaze estimation. Our method exploits the trajectories of salient feature trackers spread randomly over the face region for the estimation of the head rotation angles, which are subsequently used to drive a spherical eye-in-head rotation model that compensates for the changes in eye region appearance under head rotation. We investigate the validity of the proposed method on a publicly available data set.**

***Index Terms*—Eye-gaze tracking, pervasive, passive**

I. Introduction

Recent trends in the field of eye-gaze tracking have been shifting towards the estimation of gaze direction in everyday life settings, often referred to as *in the wild* [1]. This is being motivated by the profusion of mobile devices and a growing interest in capturing the natural user behaviour in less constrained scenarios outside the research laboratory. This emergence of *pervasive eye-gaze tracking*, as originally coined by Bulling et al. [1], calls for methods that allow for tracking under less controlled conditions, such as by allowing natural head and face movement during tracking, reduced calibration to allow for situations that do not permit prolonged user cooperation, and the estimation of eye-gaze on mobile devices comprising integrated imaging hardware without requiring further hardware modification [2].

Most research effort over past years has been dedicated towards the development of eye-gaze tracking methods that operate under controlled conditions [3], whereby the illumination is controlled by the projection of infra-red illumination [4], the head movement is constrained to a small volume [5]–[7], and calibration [5], [8], [9] or the collection of a training data set [6], [10] may be carried out as required. Such conditions limit the applicability of these methods within the less controlled settings associated with pervasive eye-gaze tracking [2].

In light of these limitations, we aim to address the challenges of pervasive eye-gaze tracking related to the estimation of gaze under natural head and face movement, reduced calibration prior to gaze estimation, and the use of consumer-grade cameras having a wider field-of-view, such as webcams, in which the eye regions appear at lower resolution. Our method is based upon our definition of the Spherical Eye-in-head Rotation (SphERo) model [11] to compensate for changes in appearance under head rotation, driven by the head rotation angles estimated earlier. The estimation of head pose exploits the trajectories of salient feature points spread randomly over the face region, in comparison with other methods in the literature [12] that require prior training or accurate initialisation of specific facial features for model-fitting. We combine shape and motion factorisation and particle filtering to estimate the head pose, with Kalman filtering in order to handle non-rigid face movement during tracking. The final gaze estimate is defined as a function of the eye and head pose components.

This paper is organised as follows. Section II outlines the overarching idea of the proposed algorithm, followed by the specific details in Section III and subsequently the implementation details in Section IV. The experimental results are presented and discussed in Section V while Section VI draws the final remarks and concludes the paper.

II. Overview of the Algorithm

We propose a method for the estimation of eye-gaze that combines our work for head [13] and eye [11] pose estimation. We estimate the head pose under non-rigid face movement in real-time, by exploiting the sparse 3-dimensional shape of the face recovered by shape and motion factorisation [13]. The 3-dimensional shape is recovered from the trajectories of salient feature points randomly distributed over the face region, rather than making use of specific face landmarks that may be susceptible to surface distortion and self-occlusion [14]. The contribution of feature points undergoing non-rigid face movement to the estimation of head pose is reduced, based on residual values computed by a Kalman filter that capture the discrepancy between the recovered 3-dimensional shape and a rigid shape estimate produced by the Kalman filter. Driven by the estimated head rotation angles, our SphERo model [11] subsequently compensates for changes in eye region appearance due to head rotation, resulting in a transformed image with known head and eye pose. This permits the estimation of

978-9-0827-9703-9/19/$31.00

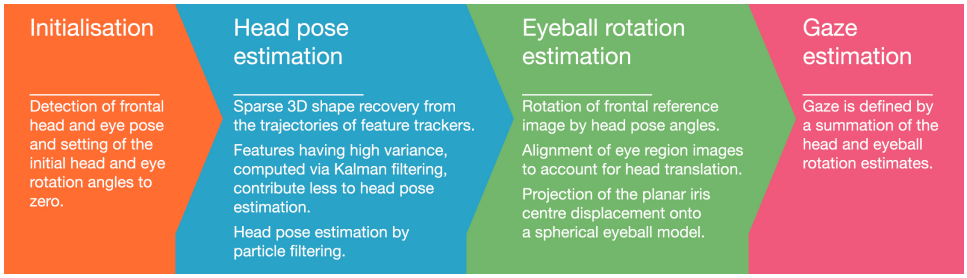


Fig. 1: Main steps of the proposed method carried out on every image frame.

eyeball rotation in a newly observed eye region image, based on the iris displacement in both the transformed and observed eye region images. The initialisation of the SphERo model requires only a brief calibration, whereby the user is requested to hold a frontal and upright head pose and a frontal eye pose. Figure 1 summarises the stages of the proposed method.

III. Details of the Algorithm

Frontal eye and head poses are initially detected from the relative distances between the face features, permitting the extraction of the frontal eye regions, $f_{\mathbf{x}_o}$, where $\mathbf{x}_o$ denotes the iris centre coordinates. At this instance, the head yaw and pitch angles, (θ, ϕ), together with the eyeball yaw and pitch angles, (α, β), are initialised to zero, where the torsional component of the eyeball rotation is assumed not to contribute to gaze estimation [15] and hence removed for simplicity. The pixels $\mathbf{x} = (x, y)$ in the reference images $f_{\mathbf{x}_o}$ are projected orthographically to positions $\mathbf{u} = (u, v)$ onto the surface of the SphERo model. This is illustrated in Figures 2(a) and 2(b). In order to compensate for changes in eye region appearance under head rotation the SphERo model is driven by the head pose, which needs to be estimated in advance.

The estimation of the head rotation angles exploits the sequential factorisation theory for shape and motion recovery of Morita and Kanade [16], in order to recover the 3-dimensional shape, $\mathbf{S}_k$, from the trajectories of P salient feature points initialised over the face region. The image positions of the feature points, $(x_{k,p},\ y_{k,p}) \mid p = 1,\ldots,P$, are updated at every time step k, where each time step corresponds with the acquisition of a new image frame. Our aim is to use this information to detect features points undergoing non-rigid face movement, after it was experimentally found that such feature points tend to be characterised by a larger error in the estimated 3-dimensional shape, $\mathbf{S}_k$. Since the method of Morita and Kanade [16] operates under a rigid-object assumption, Kalman filtering is employed to provide a notion of the difference between the measured shape, $\mathbf{S}_k$, and a static shape estimate, $\boldsymbol{\xi}_k$, produced by the Kalman filter algorithm. In this regard, of particular interest is the measurement residual, $\tilde{\mathbf{y}}_k$ of size $P\times 1$, computed during the update stage of the Kalman filter [17]. The measurement residual provides an indication of the discrepancy between the actual measurement, $\mathbf{z}_k = \mathbf{S}_k$, and the measurement prediction of the Kalman filter [17]. The actual and predicted measurement values of the tracked feature points residing on rigid surface regions are expected to be in agreement, and hence the corresponding residual values remain close to zero. Feature points that undergo non-rigid face movement, on the other hand, are expected to display larger discrepancy between the actual and predicted measurement values and are, therefore, characterised by higher residual values. We capture this discrepancy by computing the variance of the measurement residual values over a time window of length, L, for each feature point as follows,

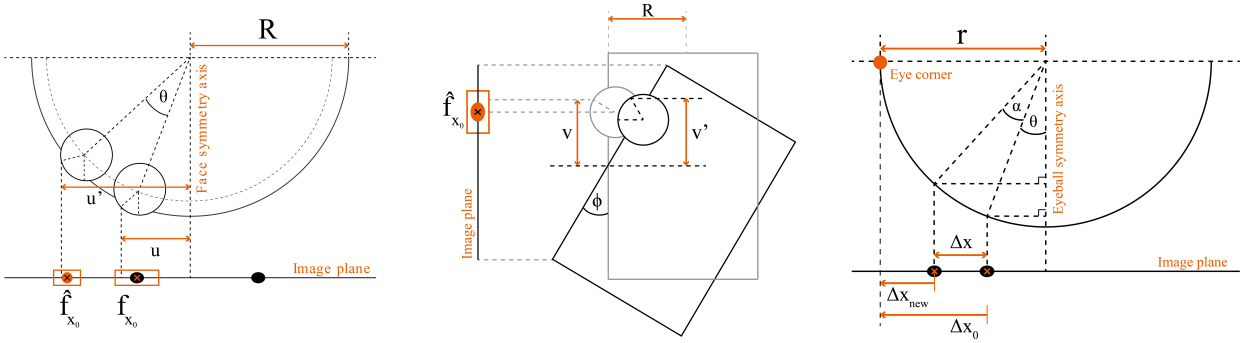
$$\sigma^2_{k,p} = \frac{1}{L}\sum_{j=(k-L)}^{k} \tilde{y}^2_{j,p} \qquad p = 1,\ldots,P \tag{1}$$

which variance will be exploited for the estimation of head pose by particle filtering, as will be explained shortly.

In order to estimate the head yaw and pitch angles by particle filtering, a set of N particles is generated as hypotheses of state $\mathbf{s}_k = (\theta_k, \phi_k)$ with known probability density function $p(\mathbf{s}_k)$. A rotation transformation is subsequently applied to the 3-dimensional rigid shape estimate produced by the Kalman filter, $\boldsymbol{\xi}_k$, according to every particle hypothesis. The re-projection of each rotated shape inside the image space assigns a set of candidate coordinates, $C_k(p) = \{(c^{(n)}_{k,p}, d^{(n)}_{k,p})\} \mid n = 1,\ldots,N$, to each feature of interest, p. This permits the calculation of the horizontal and vertical image distances, $D^{(n)}_k(\mathbf{x})$, between the tracked feature positions, $\mathbf{x}_{k,p}$, and corresponding candidates, $\mathbf{c}^{(n)}_{k,p}$. We define the likelihood model of the particle filter based on these distances, by a multivariate normal distribution having mean, $\boldsymbol{\mu} = 0$, and covariance, $\boldsymbol{\Sigma}_k$, as follows,

$$p(x_{k,p|1,\ldots,P} \mid \mathbf{s}^{(n)}_k) = Ae^{-\frac{1}{2}(\mathbf{D}^{(n)}_k(x))^T\boldsymbol{\Sigma}^{-1}_k(\mathbf{D}^{(n)}_k(x))} \qquad where,\ \ A = (2\pi)^{-\frac{P}{2}}|\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \tag{2}$$

for the head yaw and similarly for the head pitch angles. In Equation 2, P denotes the dimensionality of the measurement vector, while the pre-computed variances, $\sigma^2_{k,p} \mid p = 1,\ldots,P$, populate the diagonal entries of the $P\times P$ covariance matrix, $\boldsymbol{\Sigma}_k$, assuming feature error independence. Based on the likelihood measures in Equation 2, each particle is subsequently assigned weights denoting the likelihood of representing the true head yaw and pitch angles. Normalisation of these weights permits the estimation of state, $\mathbf{s}_k = (\theta_k, \phi_k)$, as a weighted average of the particle set for the yaw and pitch angles respectively. It is worth noting that head pose estimation is



(a) (b) (c)

Fig. 2: The SphERo model rotates face features inside a frontal reference image, $f_{\mathbf{x}_o}$, by a head yaw angle, θ, around a circular trajectory of radius R (a), and a head pitch angle, ϕ, around the nose region (b). The change in appearance due to the spherical shape of the eyeballs is accounted for by the spherical trajectory of radius, r (c).

based upon the shape rather than the photometric information of the object of interest, to reduce the susceptibility of the method to intensity variations and repetitive skin texture.

The reference eye images projected onto the model are then rotated according to the estimated head pose, $G_h = (\theta, \phi)$, and re-projected back onto the image plane to generate the transformed reference images, $\hat{f}_{\mathbf{x}_o}(i,j)$, as shown in Figures 2(a) and 2(b). Upon the acquisition of a newly observed eye region image, g_x, having an iris centre at $\mathbf{x}_{new}$, displacements $\Delta\mathbf{x}_o$ and $\Delta\mathbf{x}_{new}$ are calculated with respect to a reference point inside each of $\hat{f}_{\mathbf{x}_o}(\mathbf{i})$ and g_x. The eyeball rotation angles, α and β, are finally estimated by projecting the displacement,

$$\Delta\mathbf{x} = \Delta\mathbf{x}_o - \Delta\mathbf{x}_{new} \tag{3}$$

onto the spherical surface of the eyeball with known radius, r, as shown in Figure 2(c).

IV. Algorithm Implementation

The next sections outline the implementation of methods to extract the image information required for gaze estimation.

A. Face Detection, Feature Initialisation and Tracking

The face region was initially detected by the Viola-Jones algorithm [18] in real-time, trained on a wide variety of face images. Feature trackers were randomly latched upon salient facial features within the face bounding box detected earlier, by employing the feature detection method of Shi and Tomasi [20] to identify *good features to track*. These feature points were then tracked across consecutive image frames via the Kanade-Lucas-Tomasi (KLT) feature tracker [21].

B. Kalman and Particle Filtering

Kalman filter: The initial state estimate, $\hat{\boldsymbol{\xi}}_0$, of the Kalman filter is set to the 3-dimensional shape, $\hat{\mathbf{S}}$, computed via the shape and motion factorisation algorithm of Tomasi and Kanade [22], which recovers the required 3-dimensional shape information from the trajectories of salient feature points tracked across a sequence of image frames.

Particle filter: Implemented a Bootstrap filter [23], approximating its state evolution by a Gaussian random walk model.

C. Gaze Estimation

The iris centre coordinates used in the computation of the eyeball rotation angles were localised by a Bayes' classifier pre-trained on iris and non-iris pixels [24] to segment the iris regions by exploiting their photometric appearance.

In order to compute the displacements, $\Delta\mathbf{x}_o$ and $\Delta\mathbf{x}_{new}$, as explained in Section III, we investigated the use of the inner eye corners as the designated reference points, in comparison to the weighted median of sets of salient feature trackers positioned above the eyes, or a weighted average of the displacements from both features simultaneously. The weighting parameter that permits computation of the weighted average assumes a value between 0 and 1, and corresponds to a measure of tracking confidence obtained from a template-matching technique that tracks the inner eye corners.

V. Experimental Results and Discussion

In light of our interest in pervasive eye-gaze tracking, the proposed method was evaluated on video clips from the EYEDIAP Database [25] captured by a visual Kinect camera at 30 frames per second and a spatial resolution of 640×480 pixels. The videos feature different participants performing free head and face movements while gazing at discrete visual targets displayed on a monitor screen. For evaluation purposes, we selected subjects with suitable eye region contrast that permits the iris region to be visually discerned from the sclera. The proposed method has been implemented in MATLAB and runs at 12fps on a 2.5GHz Intel Core i7 processor.

Table I presents the mean absolute error (MAE) and standard deviation (SD) values of the gaze angles for different subjects in the EYEDIAP Database, calculated for the left and right eyeballs separately, and for the combined gaze estimates. Knowing the 3-dimensional positions of the eyeball centres with respect to the visual targets displayed on the monitor screen, as provided by the ground truth information in the

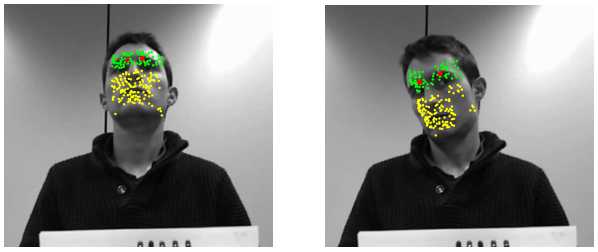


Fig. 3: Weighted median (red) of salient feature trackers positioned above the eyes (green), selected from the set of feature points used for head pose estimation (green, yellow).

EYEDIAP Database, the ground truth combined gaze was calculated from the vector that joins the mid-point between the eyes together with the visual targets on the monitor screen. Similarly, the estimated combined gaze was calculated from the vector that joins the mid-point between the eyes to the mid-point between the intersections of the individual gaze vectors with the screen. This permitted the computation of error for the estimated gaze with respect to ground truth, as presented in Table I. This table also presents the results obtained in utilising the inner eye corners as the designated reference points for the estimation of eyeball rotation (A), in comparison to the weighted median of sets of salient feature trackers positioned above the eyes (B), as shown in Figure 3, or a combination of both calculated as a weighted average (C).

A comparison between the results achieved by different reference points utilised in the estimation of eyeball rotation shows that in general the methods that exploited the inner end of each eyebrow as reference points, that is (B) and (C) in Table I, produced the lowest gaze estimation errors for the gaze yaw and pitch angles, as marked in green. It has been noted that the tracking of the inner eye corners as the designated reference points (A) suffered from occlusion during head rotations, where leftward or rightward rotations tended to occlude the corresponding inner eye corner resulting in increased gaze estimation error for the respective eyeball. This was not the case for (B) since this depended upon features that remained mostly visible during head rotations. However, due to increased distortion in the appearance of the face region underneath the features of interest during head rotation, these reference points tended to drift away from their initial image positions and hence give rise to increased gaze estimation error. In this regard, it was observed that the feature trackers tended to exhibit less drift in the vertical direction, mainly since the forehead is less curved in the vertical direction and hence its appearance suffers less distortion during vertical head rotations. It may also be observed from the results in Table I that the majority of the lowest gaze estimation errors for the gaze yaw, marked in green, were obtained by computing a weighted average of the image positions of the inner eye corners (A) together with the image coordinates of the weighted median of salient feature trackers (B), to produce new reference points denoted by (C) in Table I. The combination of both methods allowed for an improvement in the gaze yaw estimation by relying on the weighted median of the salient feature trackers when the inner eye corner tracking became unreliable or failed entirely.

The results in Table I for every participant indicate an overall improvement in the separate left and right gaze estimates if these are combined into a single gaze estimate. This is congruent with other methods in the literature, which have reported that the combination of the two gaze vectors into a single estimate tends to compensate for the effect of noise [26], [27]. In comparison with the state-of-the-art, as per Table II, our method achieves comparable or better performance than the methods of [26], [28], [29], albeit the wide range of head rotation angles of the considered subjects from the EYEDIAP Database, spanning [-25°, 18°] in yaw and [-20°, 19°] in pitch. The method of Haiyuan et al. [30] that performs better than ours, does not consider free head movement.

VI. Conclusion

In this paper, we have proposed a method for eye-gaze estimation that brings together our work for head [13] and eye [11] pose estimation. The combined method is based on our definition of the SphERo model [11], that compensates for the change in eye region appearance under head rotation. The estimation of the head rotation angles, which drive the SphERo model, exploits the trajectories of salient feature points spread randomly over the face region, and combines shape and motion factorisation together with particle and Kalman filtering to estimate the head pose under non-rigid face movement. Our method has been evaluated on the EYEDIAP Database and its performance was found to be comparable or better than the state-of-the-art.

Future work aims to improve the robustness of the salient feature tracking especially under large head rotations, in order to improve upon the achieved gaze estimation accuracy.

Acknowledgment

This work forms part of the project R&I-2016-010-T Wild-Eye financed by the Malta Council for Science and Technology through FUSION: The R&I Technology Development Programme 2016.

References

[1] A. Bulling, A. Duchowski, and P. Majaranta, "The 1st international workshop on pervasive eye tracking and mobile eye-based interaction (petmei 2011)," in *Proceedings of the 13th International Conference on Ubiquitous Computing*, 2011, pp. 627–628.

[2] S. Cristina and K. P. Camilleri, "Unobtrusive and pervasive video-based eye-gaze tracking," *Image Vis. Comput.*, vol. 74, pp. 21–40, 2018.

[3] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478–500, 2010.

[4] S. Wibirama, V. Mahasitthiwat, S. Tungjitkusolmun, C. Pintavirooj, and K. Vichienchom, "Eye-motion tracking using motion gradient," in *3rd International Symposium on Biomedical Engineering*, 2008, pp. 84–87.

[5] A. Al-Rahayfeh and M. Faezipour, "Classifiers comparison for a new eye gaze direction classification system," in *IEEE Long Island Systems, Applications and Technology Conference (LISAT)*, 2014, pp. 1–6.

[6] C. Holland and O. Komogortsev, "Eye tracking on unmodified common tablets: Challenges and solutions," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, 2012, pp. 277–280.

	Yaw (MAE/°(SD/°))			Pitch (MAE/°(SD/°))			Head rotation angles
Subject #	Left eye	Right eye	Combined	Left eye	Right eye	Combined	[Min/°, Max/°] (Yaw, Pitch)
	A: Inner eye corner						
1	10.23 (6.40)	8.35 (6.02)	8.92 (6.56)	10.05 (8.38)	11.23 (7.13)	10.47 (7.16)	[-24.95, 18.25], [-20.03, 18.95]
4	5.05 (3.24)	6.17 (3.92)	4.76 (3.19)	8.10 (5.22)	8.86 (6.38)	6.63 (4.16)	[-5.34, 7.43], [-11.40, 10.41]
8	11.97 (13.94)	8.92 (7.05)	6.83 (5.14)	15.13 (12.04)	14.61 (11.20)	11.76 (8.57)	[-21.59, 17.13], [-7.81, 15.28]
10	4.98 (3.07)	4.35 (1.90)	**3.75 (2.25)**	13.46 (10.35)	9.48 (5.29)	8.46 (6.02)	[-15.85, 6.32], [0.63, 15.32]
15	7.01 (2.66)	10.10 (6.15)	7.88 (4.30)	6.74 (3.41)	11.59 (5.33)	8.79 (3.79)	[-8.84, 5.35], [6.06, 16.08]
Mean	7.85 (5.86)	7.58 (5.01)	6.43 (4.29)	10.70 (7.88)	11.15 (7.07)	9.22 (5.94)	
	B: Inner end of eyebrow by a weighted median of salient feature trackers						
1	9.16 (6.11)	10.99 (6.83)	9.36 (6.35)	15.37 (11.64)	12.19 (8.51)	**6.62 (2.91)**	[-24.95, 18.25], [-20.03, 18.95]
4	10.38 (8.08)	7.63 (5.44)	8.67 (6.21)	9.84 (5.94)	8.48 (6.17)	**6.34 (4.10)**	[-5.34, 7.43], [-11.40, 10.41]
8	11.74 (8.82)	10.67 (8.67)	7.01 (4.29)	12.47 (13.39)	8.68 (7.43)	**4.41 (4.39)**	[-21.59, 17.13], [-7.81, 15.28]
10	7.84 (8.06)	10.22 (5.57)	6.31 (4.53)	10.64 (5.43)	9.49 (5.05)	8.35 (4.51)	[-15.85, 6.32], [0.63, 15.32]
15	11.06 (5.43)	9.80 (6.46)	9.56 (4.05)	9.72 (5.96)	6.11 (5.70)	**7.27 (5.99)**	[-8.84, 5.35], [6.06, 16.08]
Mean	10.04 (7.30)	9.86 (6.59)	8.18 (5.09)	11.61 (8.47)	8.99 (6.57)	6.60 (4.38)	
	C: Combined inner eye corner and weighted median						
1	8.80 (5.23)	7.46 (5.04)	**7.39 (4.66)**	11.02 (8.37)	10.50 (5.57)	9.62 (6.03)	[-24.95, 18.25], [-20.03, 18.95]
4	3.79 (1.88)	6.17 (4.99)	**3.78 (3.10)**	7.44 (4.64)	8.90 (6.55)	7.50 (4.55)	[-5.34, 7.43], [-11.40, 10.41]
8	7.57 (5.14)	8.47 (4.61)	**4.73 (1.60)**	9.70 (8.27)	10.43 (10.80)	6.84 (6.14)	[-21.59, 17.13], [-7.81, 15.28]
10	7.79 (8.12)	4.66 (2.43)	5.29 (4.67)	9.91 (5.54)	9.69 (4.89)	**7.75 (4.62)**	[-15.85, 6.32], [0.63, 15.32]
15	6.47 (2.54)	7.37 (4.57)	**5.99 (3.45)**	6.54 (3.60)	10.43 (5.45)	8.09 (4.02)	[-8.84, 5.35], [6.06, 16.08]
Mean	6.88 (4.58)	6.83 (4.33)	5.44 (3.50)	8.92 (6.08)	9.99 (6.65)	7.96 (5.07)	
Overall Mean			**5.13 (3.01)**			**6.48 (4.40)**	

TABLE I: Mean absolute error (MAE) and standard deviation (SD) values of the gaze estimates in visual angle, computed using different reference points as delineated (A-C). The overall mean is worked out on the values in green.

Method	Yaw (MAE(°))	Pitch (MAE(°))	Head rotation angles
[26]	7		N/A
[28]	5.3	7.7	N/A
[29]	5.953		N/A
[30]	2.36	2.11	None

TABLE II: Results reported by relevant state-of-the-art 3-dimensional gaze estimation methods.

[7] K. Kunze, S. Ishimaru, Y. Utsumi, and K. Kise, "My reading life - towards utilizing eyetracking on unmodified tablets and phones," in *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing*, 2013, pp. 283–286.

[8] R. Valenti, N. Sebe, and T. Gevers, "Combining head pose and eye location information for gaze estimation," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 802–815, 2012.

[9] N. H. Cuong and H. T. Hoang, "Eye-gaze detection with a single webcam based on geometry features extraction," in *2010 11th International Conference on Control, Automation, Robotics and Vision*, 2010, pp. 2507–2512.

[10] T. Okabe, F. Lu, Y. Sugano, and Y. Sato, "Adaptive linear regression for appearance-based gaze estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 10, pp. 2033–2046, 2014.

[11] S. Cristina and K. P. Camilleri, "Model-based head pose-free gaze estimation for assistive communication," *Computer Vision and Image Understanding - Special Issue on Assistive Computer Vision and Robotics*, vol. 149, pp. 157–170, 2016.

[12] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607–626, 2009.

[13] S. Cristina and K. P. Camilleri, "Model-free non-rigid head pose tracking by joint shape and pose estimation," *Machine Vision and Applications*, vol. 27, pp. 1229–1242, 2016.

[14] M. Sapienza and K. P. Camilleri, "Fasthpe: A recipe for quick head pose estimation," University of Malta, Tech. Rep. TR-SCE-2014, 2014.

[15] J. Edinger, D. Pai, and M. Spering, "Rolling motion makes the eyes roll: torsion during smooth pursuit eye movements," *Journal of Vision*, vol. 13, no. 9, 2013.

[16] T. Morita and T. Kanade, "A sequential factorization method for recovering shape and motion from image streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 858–867, 1997.

[17] G. Welch and G. Bishop, "An introduction to the kalman filter," University of North Carolina at Chapel Hill, Tech. Rep., 1995.

[18] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, pp. 1231–1238.

[19] "Opencv," URL: http://http://opencv.org/, 2014.

[20] J. Shi and C. Tomasi, "Good features to track," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun 1994, pp. 593–600.

[21] C. Tomasi and T. Kanade, "Detection and tracking of point features," Carnegie Mellon University, Tech. Rep. CMU-CS-91-132, 1991.

[22] ——, "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.

[23] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Tran. Signal Process.*, vol. 50, no. 2, pp. 174–188, 2002.

[24] S. Cristina and K. P. Camilleri, "Cursor control by point-of-regard estimation for a computer with integrated webcam," in *The 8th International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP)*, 2014, pp. 126–131.

[25] K. Funes Mora, F. Monay, and J. Odobez, "Eyediap database: Data description and gaze tracking evaluation benchmarks," Idiap, Tech. Rep. Idiap-RR-08-2014, 2014.

[26] E. Pogalin, A. Redert, I. Patras, and E. A. Hendriks, "Gaze tracking by using factorized likelihoods particle filtering and stereo vision," in *Third International Symposium on 3D Data Processing, Visualization, and Transmission*, 2006, pp. 57–64.

[27] Y. Matsumoto and A. Zelinsky, "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement," in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 499–504.

[28] H. Yamazoe, A. Utsumi, T. Yonezawa, and S. Abe, "Remote gaze estimation with a single camera based on facial-feature tracking without special calibration actions," in *Proceedings of the 2008 Symposium on Eye Tracking Research and Applications*, 2008, pp. 245–250.

[29] M. J. Reale, S. Canavan, L. Yin, K. Hu, and T. Hung, "A multi-gesture interaction system using a 3-d iris disk model for gaze estimation and an active appearance model for 3-d hand pointing," *IEEE Transactions on Multimedia*, vol. 13, no. 3, pp. 474–486, 2011.

[30] W. Haiyuan, Y. Kitagawaa, and T. Wada, "Tracking iris contour with a 3d eye-model for gaze estimation," in *Proceedings of the 8th Asian Conference on Computer Vision*, 2007, pp. 688–697.