REAL-TIME MONOCULAR 6-DOF HEAD POSE ESTIMATION FROM SALIENT 2D POINTS

Jilliam María Díaz Barros[†⋆] *Frederic Garcia*[⋆] *Bruno Mirbach*[⋆] *Didier Stricker*[†]

[⋆] PTU Optical, IEE S.A., Luxembourg
[†] German Research Center for Artificial Intelligence (DFKI), Germany

ABSTRACT

We propose a real-time and robust approach to estimate the full 3D head pose from extreme head poses using a monocular system. To this end, we first model the head using a simple geometric shape initialized using facial landmarks, *i.e*., eye corners, extracted from the face. Next, 2D salient points are detected within the region defined by the projection of the visible surface of the geometric head model onto the image, and projected back to the head model to generate the corresponding 3D features. Optical flow is used to find the respective 2D correspondences in the next video frame. Assuming that the monocular system is calibrated, it is then possible to solve the Perspective-n-Point (PnP) problem of estimating the head pose given a set of 3D features on the geometric model surface and their corresponding 2D correspondences from optical flow in the next frame. The experimental evaluation shows that the performance of the proposed approach achieves, and in some cases improves the state-of-the-art performance with a major advantage of not requiring facial landmarks (except for initialization). As a result, our method also applies to real scenarios in which facial landmarks-based methods fail due to self-occlusions.

Index Terms— Head pose estimation, monocular system, perspective-n-point

1. INTRODUCTION AND RELATED WORK

The extensive literature on head pose estimation (HPE) [1] evidences the importance to determine the position and orientation of the head relative to some coordinate system. Indeed, HPE is an essential step to address well-known research topics such as face recognition [2], facial expression recognition [3] and eye location [4, 5]. Certainly, it has been demonstrated that the combination of head pose and eye location provides a reference framework to estimate the gaze, from which is possible to determine the level of attention or drowsiness, for instance, of a vehicle driver [6, 7]. However, a real-time HPE algorithm that is robust under realistic conditions is still a challenge in computer vision. Such conditions are, *e.g*., varying illumination, (self-)occlusions due to person's accessories or large head rotations.

Among the vast literature on HPE, we herein focus on geometric-based methods that apply to monocular systems. Concerned by illumination changes, La Cascia *et al*. [8] proposed a 3D head tracking approach based on a cylindrical head model (CHM), with registration of texture map images. An illumination correction term was applied to handle light variation among consecutive image frames. In [9], Basu *et al*. presented a framework for rigid motion estimation employing an ellipsoidal head model (EHM). Tracking was performed through motion regularization; optical flow was used to compute the model flow and the motion error was minimized using the simplex gradient descent technique. Motion was estimated similarly in [10], and the head represented with an extended superquadric (ESQ) model. The approach was robust to self-occlusions and non-rigid motion, but not suited for real-time applications. An alternative HPE approach was proposed by Xiao *et al*. in [11], where a CHM was initialized using a reference template of the head image and its 3D pose. Motion was recovered by minimizing an objective function and assuming a perspective projection. A tracking approach based on dynamic templates together with a pixel weighting scheme were included to handle occlusions, non-rigid motion and gradual lighting changes. Jang and Kanade presented in [12, 13] an user-specific approach for head tracking using a CHM. Kalman filter was used to combine motion estimated between consecutive frames and the pose from a database that related SIFT feature points to different head poses. Choi and Kim proposed in [3] an alternative template approach for head motion recovering in which they combined a particle filter with an EHM. Active appearance model (AAM) was implemented to initialize the 3D motion parameters, and a modified version of the online appearance model (OAM) was integrated at the observation model. Similarly, Sung *et al*. [14] combined the AAM with a CHM in order to extent the working range of the head motion. AAM was used to initialize the global motion parameters of the CHM. Later, these parameters were employed to update the AAM during the tracking step. An and Chung proposed in [2] a method for face recognition where the pose was estimated using a parametrized EHM. They modelled a linear system assuming a rigid body motion under perspective projection and estimated the pose using least squares. From this pose, face images were registered into frontal views, facilitating their recognition. In [4, 5], Valenti *et al*. presented a hybrid scheme for gaze estimation that combined HPE and eye tracking. Eyes were located using an isophote-based method that exploited the semi-circular patterns inside the eye and the head was modelled using a CHM. Morency *et al*. introduced in [15] GAVAM (Generalized Adaptive View-based Appearance Model), a probabilistic scheme, where the head was modelled using a half ellipsoid. The pose estimation problem was formulated as a linear system solved by Normal Flow Constraint (NFC). More recently, a pose-adaptive Constrained Local Model (CLM) framework was proposed in [16], which integrates AAM in order to handle large head rotations. Similarly to our work, Ju *et al*. [6] used salient 2D points as 2D features for pose estimation. These features were detected in the face region and later projected onto a EHM using ray tracing. The head pose was computed by solving the

This work is funded by the National Research Fund, Luxembourg. Project ID 9235599.

transformation between these 3D features and their corresponding 2D features in the next frame, obtained through the KLT tracker [17]. Two different approaches were evaluated, namely the Perspective-n-Point problem and POSIT. The downside of this method was related to the need of detecting the face at each frame in order to define the region of interest (ROI) for 2D feature extraction.

In this work we estimate the head pose and provide its position and orientation across 6-DoF for each video frame. We have modelled the human head as a disembodied rigid object [1] being free to move left/right, up/down and forward/backward, *i.e.*, translation in x, y, and z perpendicular axes, respectively, combined with the rotation about the same axes, often termed as *pitch*, *yaw*, and *roll*, respectively. Similarly to [3, 11] and due to its ease of initialization in comparison to precise and flexible head models [1] or parametrized face models [18, 19], we use a simple geometric shape to model the head. Indeed, a CHM or EHM only require the estimation of fewer facial landmarks to be initialized, as we describe in Section 2.

The main contributions of our method are:

- In general, alternative HPE approaches require face detection to define the ROI for feature detection. We, instead, define the ROI by projecting the visible surface of our geometric model. This enables for a real-time performance.

- The use of salient 2D points as 2D features within the ROI instead of specific facial landmarks, *e.g.*, nose tip, eyes' and/or lips' corners. By doing so, our system is robust in situations where facial landmarks are not available due to *e.g.*, (self-)occlusions.

- We introduce a weighting scheme based on the distance transform to identify those 2D features that are lying close to the border of ROI and thus prone to self-occlusion or likely to belong to the background.

The remainder of this paper is organized as follows: Section 2 describes the proposed approach; Section 3 discuses the experimental evaluation and obtained results. Concluding remarks and future work are presented in Section 4.

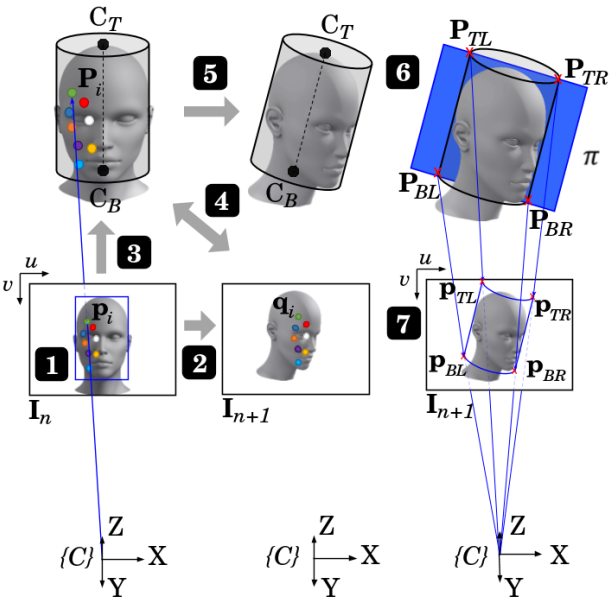

Fig. 1. Proposed HPE pipeline.

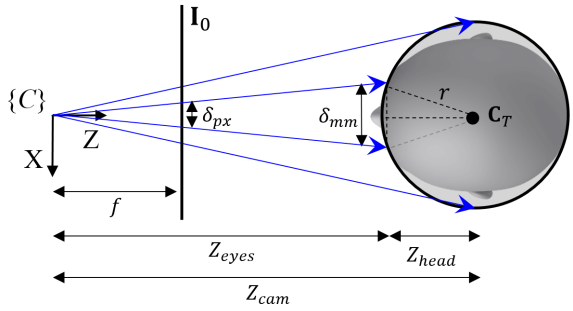

Fig. 2. Location of the CHM in the 3D space.

2. PROPOSED APPROACH

In general, HPE approaches based on simple geometric models and intended for monocular systems require precise detection of facial landmarks, or rely on frame to frame face detection in order to continuously update the CHM/EHM. Contrary, we propose a HPE pipeline that once initialized estimates the head pose with no need of face detection, neither facial landmarks. As a result, the proposed approach is robust to large head rotations regardless of the facial expression and/or self-occlusions. The pipeline is shown in Fig. 1.

In the following and due to space limitations, we present our approach using a CHM. However, EHM has been also implemented and evaluated in the experiments (see Section 3).

2.1. System Initialization

An essential step before starting the HPE process is to adjust the CHM to the user's head. To that end, the user's face dimensions as well as the distance between the eye's centers are needed. Herein, we use the OpenCV [20] implementation of the Viola-Jones detection method [21] which is based on a cascade of boosted classifiers constructed from low-level features. However, alternative techniques such as appearance template methods [1] might also work, as soon as the eyes' centers are also provided.

In the following we assume a calibrated camera with known intrinsic parameters. As shown in Fig. 2, we define the distance between the camera and the baseline of the eyes in the CHM by $Z_{eyes} = f \cdot \frac{\delta_{mm}}{\delta_{px}}$, with f the focal length of the camera, δ_{px} the estimated distance between the eye's centers (both in pixels), and δ_{mm} the mean interpupillary distance, which corresponds to 64.7 mm for male and 62.3 mm for female according to anthropometric databases. Z_{cam} denotes the distance between the center of the CHM and the camera, and results from the sum of Z_{eyes} and Z_{head}, *i.e.*, the distance from the head's center to the midpoint between the eyes' baseline. Z_{cam} is related to the radius r of the CHM as $Z_{head} = \sqrt{r^2 - (\delta_{mm}/2)^2}$, with $r = \frac{1}{2}|\mathbf{p}_{TR} - \mathbf{p}_{TL}| \cdot \frac{\delta_{mm}}{\delta_{px}}$, and $\mathbf{p}_{TR}$ and $\mathbf{p}_{TL}$ the top right and top left corners of the detected face bounding box (see Fig. 3). Similarly, the height of the CHM is

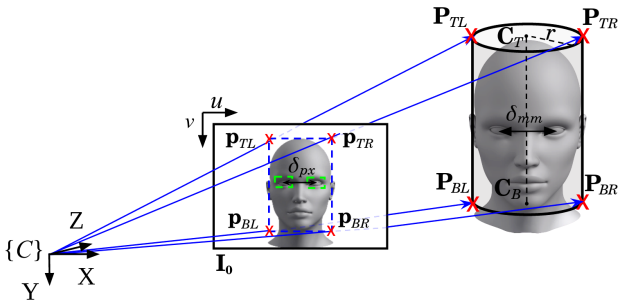

Fig. 3. Initialization of the CHM to fit the user's head.

adjusted to the user's head by $h = |\mathbf{p}_{TR} - \mathbf{p}_{BR}| \cdot \frac{\delta_{mm}}{\delta_{px}}$.

Once Z_{cam} is estimated, we compute the corners of the face bounding box in the 3D space, *i.e.*, $\{\mathbf{P}_{TL}, \mathbf{P}_{TR}, \mathbf{P}_{BL}, \mathbf{P}_{BR}\}$, together with the centers of the top and bottom bases of the CHM, *i.e.*, $\mathbf{C}_T$ and $\mathbf{C}_B$. We note that during the initialization step we assumed that the user is facing the camera. Therefore, both rotation angles $(\omega_x, \omega_y, \omega_z)$ and initial translation values (t_x, t_y, t_z) are set to zero.

2.2. Robust Non-facial 2D Feature Detection and Matching

Geometric-based HPE approaches either require facial landmarks like nose tip, eye's and/or lips' corners to be frame-to-frame detected, or sample and project onto the image plane the visible surface of the CHM/EHM in a grid-like structure for 2D feature detection [4]. Contrary, we propose to detect salient 2D points as 2D features $\mathbf{p}_i$ (see Step 1 in Fig. 1) within the ROI defined by the projection of the visible surface of the CHM onto the image plane, further explained in Section 2.5.

To detect the 2D features we use FAST [24], a machine learning-based corner-detection algorithm. Aside from being very well suited for real-time applications, it provides accurate corners. Each feature is weighted according to its location within the ROI. That is, we have introduced a feature weighting scheme that enables us to discard outliers and/or non-reliable features and thus, to only consider robust features for HPE. This weighting scheme results from applying the distance transform onto the ROI, *i.e.*, each pixel within the ROI has a normalized weight related to the Euclidean distance to the closest boundary. Therefore, a higher weight will be set to those features located in the center of the ROI than to those near to the ROI boundaries. As can be observed in Section 3, a better performance is achieved when removing low-weighted features within the HPE process.

The iterative Lucas-Kanade feature tracker with pyramids described in [25] has been used to find the feature correspondences in the next frame (see Step 2 in Fig. 1).

2.3. Computation of 3D Features

3D features $\mathbf{P}_i$ (see Step 3 in Fig. 1) result from the intersection between the ray starting at the optical center of the camera and passing through the 2D feature at the image plane, and the visible surface of the CHM. The equation of the line is defined as $\mathbf{P} = \mathbf{C} + k\mathbf{V}$, with $\mathbf{V}$ being a vector parallel to the line that goes from the camera's optical center $\mathbf{C}$ through $\mathbf{P}$. The scalar parameter k is computed by solving the quadratic equation of the geometric model. In the case of the CHM, the equation is defined as $|\mathbf{V}_1|^2 k^2 + 2(\mathbf{V}_1 \cdot \mathbf{X}_1) k + |\mathbf{X}_1|^2 - r^2 = 0$, with $\mathbf{V}_1 = \mathbf{V} - \frac{\mathbf{V}\cdot\mathbf{A}}{|\mathbf{A}|}\mathbf{A}$, and $\mathbf{X}_1 = (\mathbf{C} - \mathbf{C}_B) - \frac{(\mathbf{C}-\mathbf{C}_B)\cdot\mathbf{A}}{|\mathbf{A}|}\mathbf{A}$. $\mathbf{A}$ is a vector parallel to the axis of the cylinder, $\mathbf{C}_B$ is the center of its lower base, and r is the radius.

2.4. Head Pose Estimation

The set of 3D features along with the corresponding 2D features detected in Step 2 (see Fig. 1) from the next frame are used to solve the PnP problem and thus, to estimate the head pose. The pose (see Step 4 in Fig. 1) is computed by minimizing the error between the reprojection of the 3D features onto the image plane and their respective detected 2D features by means of an iterative approach based on the Levenberg-Marquardt algorithm.

From the resulting translation $\mathbf{t}$ and rotation $\mathbf{R}$ we update the CHM (see Step 5 in Fig. 1) as well as the set of 3D features, *i.e.*, $\mathbf{P}'_i = \mathbf{R} \cdot \mathbf{P}_i + \mathbf{t}$. However, we do not update the pose of the corners defining the face bounding box in the 3D space, *i.e.*, $\{\mathbf{P}_{TL}, \mathbf{P}_{TR}, \mathbf{P}_{BL}, \mathbf{P}_{BR}\}$ in a same way, as they need to be recalculated to define the new ROI for feature detection in the next frame.

2.5. ROI for Non-facial 2D Feature Detection

As introduced in Section 2.2, the ROI for non-facial 2D feature detection is defined by the projection of the visible surface of the CHM onto the image, regardless of the orientation of the head (see Step 6 in Fig. 1). However, this is not a trivial task. Alternative geometric-based methods simply keep track of the 3D face bounding box, *i.e.*, $\{\mathbf{P}_{TL}, \mathbf{P}_{TR}, \mathbf{P}_{BL}, \mathbf{P}_{BR}\}$ and define the ROI by its projection to the image plane. Even so, these methods fail for large head rotations in which the track of the face bounding box results partially visible or even not visible. Other authors used a grid-like or mesh structure along the CHM surface from which the projection of each vertex is used for feature detection. However, a dense grid is needed to ensure robust 2D features detection, which is not practical if real-time is required. Herein, we present an alternative solution in which only the four corners defining the visible surface of the CHM along with the curvature of the CHM's bases are needed to define the ROI. To do so, we first define a plane π with its normal vector resulting from the cross product between a parallel vector to the x-axis of the camera and the vector resulting from the two centers of the CHM bases, *i.e.*, $\mathbf{C}_T$ and $\mathbf{C}_B$. The visible surface of the CHM is given by the furthermost intersected points between the CHM and π, *i.e.*, $\{\mathbf{P}'_{TL}, \mathbf{P}'_{TR}, \mathbf{P}'_{BL}, \mathbf{P}'_{BR}\}$, whereas the new ROI results from filling the polygon defined by their projection along with the curvature of the CHM's bases (see Step 7 in Fig. 1).

3. EXPERIMENTS

The evaluation of the proposed approach was carried out using an Intel Core(TM) i5-4210U processor. The approach has been implemented in C++, with assistance of the OpenCV library. We have evaluated our approach for both the CHM and EHM using the Boston University (BU) head pose dataset [8]. The BU dataset contains 45 video sequences of 5 persons performing different motions in an office under uniform illumination. Ground truth is generated using the Flock of Birds tracker with a nominal accuracy of 1.8 mm in translation and 0.5 degrees in rotation.

Weight thresh.	CHM			EHM		
	Roll	Pitch	Yaw	Roll	Pitch	Yaw
0	4.22	6.37	6.30	3.41	4.36	5.31
5	3.66	5.73	6.16	3.36	4.46	5.09
10	3.23	5.89	6.22	3.29	4.79	5.44
15	3.15	6.19	6.56	3.34	5.04	5.69
25	3.06	6.32	6.46	3.31	5.19	5.66

Table 1. RMSE of the head orientation depending on the chosen weight threshold.

		Roll	Pitch	Yaw
RMSE	Our approach			
	CHM	3.66	5.73	6.16
	EHM	3.36	4.46	5.09
	Valenti *et al*. [4]			
	Fixed template with eye cues	3.00	5.26	6.10
	Fixed template w/o eye cues	3.85	6.00	8.07
	Updated template with eye cues	3.93	5.57	6.45
	Updated template w/o eye cues	4.15	5.97	6.40
	Asteriadis *et al*. [26]*	3.56	4.89	5.72
	Sung *et al*. [14]	3.1	5.6	5.4
	An *et al*. [2]			
	CHM	3.22	7.22	5.33
	EHM	2.83	3.95	3.94
	Plane Head Model (PHM)	2.99	7.32	18.08
STD	Our approach			
	CHM	3.35	4.54	5.42
	EHM	2.98	3.84	4.56
	Valenti *et al*. [4]			
	Fixed template with eye cues	2.82	4.67	5.79
	Fixed template w/o eye cues	3.43	5.21	7.37
	Updated template with eye cues	3.57	4.56	5.72
	Updated template w/o eye cues	3.72	4.87	5.49
MAE	Our approach			
	CHM	2.80	4.58	4.87
	EHM	2.56	3.39	3.99
	Wang *et al*. [18]*	1.86	2.69	3.75
	Prasad *et al*. [19]*	3.6	2.5	3.8
	Jang *et al*. [13]	2.07	3.44	4.22
	Jang *et al*. [12]	2.1	3.7	4.6
	Choi *et al*. [3]			
	CHM	2.45	4.43	5.19
	EHM	2.82	3.92	4.04
	Morency *et al*. [15]	2.91	3.67	4.97

*These methods do not use a simple geometric head model but a 3D parametrized face model.

Table 2. Comparison of the RMSE, STD and MAE achieved by the proposed approach to other methods of the state of the art.

Table 1 reports the root mean square errors (RMSE) of the estimated head orientation by both CHM and EHM, and given by the proposed HPE approach depending on the chosen weight threshold described in Section 2.2. As expected, the overall performance improves in both geometrical models as soon as a weight threshold above 0 is chosen. Indeed, outliers that correspond to 2D features from the background of the head are then discarded. However, in the case of the EHM, the weighting scheme improvement is less significant for the pitch rotation, which might be due to the fact that the EHM fits better to the shape of the head and thus, less outliers are detected. This better fitting also explains why the EHM outperforms the CHM, where errors are introduced when computing the 3D features by ray tracing. Table 2 reports the RMSE, the mean absolute error (MAE), and the standard deviation (STD) measures of the estimated head orientation given by our approach and by the most outstanding HPE approaches in the literature. In this case, we have chosen a weight threshold of 5 for robust 2D feature selection. From the table we can observe that our results are comparable or even better than state-of-the-art approaches such as [4], and with a major advantage of not using eye cues. Also, we note that some of the approaches considered for evaluation do not use geometric models, *e.g*., [26], requiring the detection of facial landmarks in each frame. On the other hand, [18] and [19] use 3D parametrized face models, which make them non suitable for real-time applications. We note that the results related to the head translation have not been reported as the calibration data for the ground truth is not available. Finally, Table 3 reports the time consumption of each step of the proposed HPE pipeline (see Fig. 1). As can be observed, the initialization is the most time-demanding step whereas the rest of the process perfectly suits for real-time applications in which a robust HPE is required, *e.g*., drowsiness, distraction/attention, or fatigue detection.

	Time in ms	
	CHM	EHM
System Initialization	657	631
Robust Non-facial 2D Feature Detection	2.29	1.99
2D Feature Matching	9.89	6.97
Computation of 3D Features	1.04	1.68
Head Pose Estimation	3.43	2.49
ROI for Non-facial 2D Feature Detection	6.87	4.70
Time consumption for first frame	680.51	648.83
Time consumption for other frames	23.51	17.83

Table 3. Average runtimes of each step of the proposed HPE pipeline presented in Fig. 1, for 150 launches.

4. CONCLUDING REMARKS

We present a real-time HPE approach that handles extreme head positions under real scenarios. The proposed approach is intended for monocular systems and only requires 2D salient points to accurately provide the 6 DoF of the user's head. The proposed approach is suitable for real-time applications as it does not require to detect the face of the user, neither facial landmarks at each frame. Indeed, and in contrast to the alternative HPE approaches, we simply use 2D salient points that we filter according to an introduced 2D feature weighting scheme based on the distance transform. We model the head using simple geometric models such as a CHM or an EHM and we generate 3D features by projecting the previously computed 2D features onto the model surface. 2D feature correspondences are also detected in the following frame and used to solve the PnP problem, from which results the head pose. By doing so, our approach can handle extreme head rotations in which facial landmarks are not necessarily available and it is robust to self-occlusions.

Occlusion handling and pose recovering are already topics in which we are working on, together with eye tracking and gaze estimation. Furthermore, we are also working on a novel HPE and gaze estimation dataset that will address large head rotations, illumination changes, long sequences, multiracial subjects, and accessories such as glasses, hat or scarf. This new dataset will be publicly available for the research community.

5. REFERENCES

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

[2] K. H. An and M. J. Chung, "3d head tracking and pose-robust 2d texture map-based face recognition using a simple ellipsoid model," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*. IEEE, 2008, pp. 307–312.

[3] S. Choi and D. Kim, "Robust head tracking using 3d ellipsoidal head model in particle filter," *Pattern Recognition*, vol. 41, no. 9, pp. 2901–2915, 2008.

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

[5] R. Valenti, Z. Yucel, and T. Gevers, "Robustifying eye center localization by head pose cues," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 612–618.

[6] J. Kun, S. Bok-Suk, and K. Reinhard, "Novel backprojection method for monocular head pose estimation," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 13, no. 1, pp. 50–58, 2013.

[7] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, no. 2, pp. 300–311, 2010.

[8] M. La Cascia, S. Sclaroff, and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 4, pp. 322–336, 2000.

[9] S. Basu, I. Essa, and A. Pentland, "Motion regularization for model-based head tracking," in *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*. IEEE, 1996, vol. 3, pp. 611–616.

[10] Y. Zhang and C. Kambhamettu, "3d head tracking under partial occlusion," *Pattern Recognition*, vol. 35, no. 7, pp. 1545–1557, 2002.

[11] J. Xiao, T. Moriyama, T. Kanade, and J. F. Cohn, "Robust full-motion recovery of head by dynamic templates and re-registration techniques," *International Journal of Imaging Systems and Technology*, vol. 13, no. 1, pp. 85–94, 2003.

[12] J. S. Jang and T. Kanade, "Robust 3d head tracking by online feature registration," in *8th IEEE Int. Conf. on Automatic Face and Gesture Recognition*. Citeseer, 2008.

[13] J. S. Jang and T. Kanade, "Robust 3d head tracking by view-based feature point registration," 2010.

[14] J. Sung, T. Kanade, and D. Kim, "Pose robust face tracking by combining active appearance models and cylinder head models," *International Journal of Computer Vision*, vol. 80, no. 2, pp. 260–274, 2008.

[15] L. Morency, J. Whitehill, and J. Movellan, "Generalized adaptive view-based appearance model: Integrated framework for monocular head pose estimation," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–8.

[16] L. Zamuner, K. Bailly, and E. Bigorgne, "A pose-adaptive constrained local model for accurate head pose tracking," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014, pp. 2525–2530.

[17] B. D. Lucas, T. Kanade, et al., "An iterative image registration technique with an application to stereo vision.," in *IJCAI*, 1981, vol. 81, pp. 674–679.

[18] H. Wang, F. Davoine, V. Lepetit, C. Chaillou, and C. Pan, "3-d head tracking via invariant keypoint learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 8, pp. 1113–1126, 2012.

[19] B. H. Prasad and R. Aravind, "A robust head pose estimation system for uncalibrated monocular videos," in *Proceedings of the Seventh Indian Conference on Computer Vision, Graphics and Image Processing*. ACM, 2010, pp. 162–169.

[20] G. Bradski, "The opencv library," *Dr. Dobb's Journal of Software Tools*, 2000.

[21] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. IEEE, 2001, vol. 1, pp. I–511.

[22] Neil A. D., "Variation and extrema of human interpupillary distance," in *Proceedings of SPIE: Stereoscopic Displays and Virtual Reality Systems XI*, 2004, vol. 5291, pp. 36–46.

[23] C. C. Gordon, B. Bradtmiller, C. E. Clauser, T. Churchill, J. T. McConville, I. Tebbetts, and R. A. Walker, "Anthropometric survey of u.s. army personnel: Methods and summary statistics," in *Technical report 89-044. Natick MA: U.S. Army Natick Research, Development and Engineering Center.*, 1989.

[24] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, pp. 105–119, 2010.

[25] J. Y. Bouguet, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel Corporation*, vol. 5, pp. 1–10, 2001.

[26] S. Asteriadis, K. Karpouzis, and S. Kollias, "Head pose estimation with one camera, in uncalibrated environments," in *Proceedings of the 2010 workshop on Eye gaze in intelligent human machine interaction*. ACM, 2010, pp. 55–62.