Model-Free Head Pose Estimation Based on Shape Factorisation and Particle Filtering

Conference Paper · September 2015

DOI: 10.1007/978-3-319-23117-4_54

2 authors:

Stefania Cristina
University of Malta
11 PUBLICATIONS **11** CITATIONS
SEE PROFILE

Kenneth P Camilleri
University of Malta
90 PUBLICATIONS **774** CITATIONS
SEE PROFILE

Model-Free Head Pose Estimation Based on Shape Factorisation and Particle Filtering

Stefania Cristina(✉) and Kenneth P. Camilleri

Department of Systems and Control Engineering, University of Malta,
Msida MSD2080, Malta
{stefania.cristina,kenneth.camilleri}@um.edu.mt

Abstract. Head pose estimation is essential for several applications and is particularly required for head pose-free eye-gaze tracking where estimation of head rotation permits free head movement during tracking. While the literature is broad, the accuracy of recent vision-based head pose estimation methods is contingent upon the availability of training data or accurate initialisation and tracking of specific facial landmarks. In this paper, we propose a method to estimate the head pose in real-time from the trajectories of a set of feature points spread randomly over the face region, without requiring a training phase or model-fitting of specific facial features. Conversely, without seeking specific facial landmarks, our method exploits the sparse 3-dimensional shape of the surface of interest, recovered via shape and motion factorisation, in combination with particle filtering to correct mistracked feature points and improve upon an initial estimation of the 3-dimensional shape during tracking. In comparison with two additional methods, quantitative results obtained through our model- and landmark-free method yield a reduction in the head pose estimation error for a wide range of head rotation angles.

1 Introduction

Head pose estimation plays an important role in the process of estimating the eye-gaze [8], providing an initial coarse indication of the gaze direction which may then be refined according to the eyeball rotation to define the gaze at a finer level. Information relating to the head pose is relevant to a host of applications, such as in human-computer interaction (HCI) where, in conjunction with eye tracking, the estimation of head rotation permits the calculation of a point-of-regard on a monitor screen at different eye and head configurations. This is especially desirable in unconstrained eye-gaze tracking scenarios where the estimation of head pose permits free head movement during tracking, hence eliminating the need for a chin-rest which would otherwise be required to maintain the head stationary.

The problem of head pose estimation has been receiving increasing interest over the years, leading to the development of various methods that seek to estimate the head pose reliably [13]. Existing methods may be broadly classified into two major categories based on their approach in exploiting either the holistic appearance [3,9,12,14] or distinct features [4,10,11,15,21] of the face

G. Azzopardi and N. Petkov (Eds.): CAIP 2015, Part II, LNCS 9257, pp. 628–639, 2015.
DOI: 10.1007/978-3-319-23117-4_54

for head pose estimation. Appearance-based methods generally exploit the face image information entirely to estimate the head orientation. Typical variants of appearance-based methods search for the best matching head pose from a collection of pose-annotated templates [3], register a flexible model of the facial shape to target colour [14] or texture maps [12], or seek low-dimensional manifolds which model the variations in head pose robustly [9]. Feature-based methods, on the other hand, rely on a sparse set of feature points sampled at specific feature positions within the face region. The chosen features often serve as landmarks for non-rigid [4,10,11,21] or geometrical face models that infer the head orientation from the relative configuration of the facial features [15]. In general, the main challenges associated with existing appearance and feature-based methods relate to the necessity for training data prior to head pose estimation and the capability to estimate the head pose accurately especially in the presence of large head rotation angles. In this regard, the achievable estimation accuracy of methods that rely on a training stage is often contingent upon the size of the training set and the conditions under which the training data was captured [3,9,12,21]. Furthermore, the estimation accuracy of methods that rely on model-fitting generally depends upon accurate initialisation and tracking of specific facial features. Face feature detection is, however, an open problem in itself [21], prompting several model-based methods to resort to manual initialisation of the facial features [4,11], while the accuracy of feature tracking is typically susceptible to distortion and self-occlusions which may hamper the range of achievable head rotations [15].

In light of these challenges, we propose a method to estimate the head pose in real-time based on the trajectories of salient feature points spread randomly over the face region, in order to allow larger head rotation angles without requiring prior training or accurate initialisation of specific facial features. In the absence of specific facial landmarks that fit the face models typically proposed in the literature [10,11,21], we propose to apply shape and motion factorisation to the problem of head pose estimation to recover a sparse 3-dimensional representation of the surface of interest [18]. Factorisation theory is well-known in the domain of structure from motion (SfM), for the purpose of recovering the 3-dimensional shape from the trajectories of a sparse set of feature points, however, to the best of our knowledge, it has never been considered within the context of head pose estimation. Nonetheless, despite its effectiveness, the factorisation method is susceptible to the presence of noise and outliers in the feature trajectories due to drifting feature trackers, which in turn reduce the accuracy of the recovered shape and motion information [20]. Hence, we propose to combine factorisation with particle filtering in order to correct mistracked feature points in real-time, preventing the feature trackers from drifting off the features of interest due to distortion or self-occlusion, while permitting correctly tracked feature points to contribute to the factorisation result and improve upon an initial estimation of the sparse 3-dimensional shape. In comparison with other methods which employ particle filtering to estimate the head rotation [3,4,11], we base our estimation upon the 3-dimensional shape of the face rather than the photometric proper-

ties, hence reducing the susceptibility of the method to intensity variations and repetitive skin texture. Furthermore, we exploit the 3-dimensional information of the surface of interest without necessitating the use of depth sensors [5] or stereo-vision [7], which may reduce the portability of the setup especially in unconstrained scenarios.

This paper is organised as follows. Section 2 describes the details of the proposed method for head pose estimation. Section 3 presents and discusses the experimental results, while Section 4 draws the final remarks and concludes the paper.

2 Method

The following sections describe the stages of the proposed method, by first outlining the overarching idea of the proposed algorithm in Section 2.1 and subsequently presenting the implementation details in Section 2.2.

2.1 Outline of the Algorithm

Our method estimates the head pose angles in real-time by exploiting the sparse 3-dimensional shape of salient feature points randomly distributed over the surface of interest, in combination with particle filtering to generate hypotheses and estimate the head pose at every image frame. In the absence of a specific face model, we employ shape and motion factorisation theory to recover the sparse 3-dimensional surface from feature trajectories initially collected over a sequence of image frames [18]. At every time step, the image frame is first rotated according to the roll angle recovered via factorisation at the previous time step, in order to compensate for the head roll by aligning the horizontal and vertical head axes with the corresponding image axes. Subsequently, the 3-dimensional shape is rotated according to a set of N particles, where each particle defines a hypothesis of the head yaw and pitch angles, and re-projected to the image space such that the image space distance between the x and y-coordinates of the re-projected and the tracked feature positions is calculated separately. This distance permits the particles to be weighted accordingly such that the head yaw and pitch angles are then defined by a weighted average of the particle set. In order to improve the initially estimated 3-dimensional shape and correct the x and y-coordinates of mistracked feature points during tracking, a weighted average between the image coordinates of the tracked features and the feature positions corresponding to the re-projected 3-dimensional shape defined by the particle filter is also computed at every image frame. An updated 3-dimensional shape is finally recovered from this information via factorisation to be used at the next time step.

Specifically, therefore, the proposed algorithm initially tracks P salient feature points through K time steps, where each time step corresponds with the acquisition of a new image frame. The coordinates, $(u_{k,p}, v_{k,p}) \mid k = 1, \ldots, K$

and $p = 1, \ldots, P$, of the feature trajectories are subsequently collected inside a measurement matrix W of size $2K \times P$ as follows,

$$\mathbf{W} = \begin{bmatrix} U \\ \overline{V} \end{bmatrix} \tag{1}$$

According to the factorisation theory [18], in the absence of noise, matrix $\mathbf{W}$ is at most of rank three and may be decomposed into motion and shape, denoted by matrices $\mathbf{M}$ and $\mathbf{S}$ respectively, as follows,

$$\mathbf{W} = \mathbf{MS} \tag{2}$$

In the presence of noise matrix $\mathbf{W}$ is not of rank three and this decomposition may instead be approximated by singular value decomposition (SVD), which results in unitary matrices $\mathbf{U}$ and $\mathbf{V}$, and a diagonal matrix Σ,

$$\mathbf{W} = \mathbf{U}\Sigma\mathbf{V}^T \tag{3}$$

hence allowing the estimation of the motion and shape matrices,

$$\mathbf{M} = \mathbf{U}\Sigma^{\frac{1}{2}} \qquad \mathbf{S} = \Sigma^{\frac{1}{2}}\mathbf{V}^T \tag{4}$$

Following the computation of the 3-dimensional shape $\mathbf{S}$, a set of N particles $\mathbf{x}_k^{(n)} \sim p(\mathbf{x}_k), n = 1, \ldots, N$ is generated at time step $k = (K+1)$, where each particle denotes a hypothesis of state $\mathbf{x}_k = (\alpha_k, \beta_k)$ with known probability density function $p(\mathbf{x}_k)$. The feature coordinates $(u_{k,p}, v_{k,p})$ are also updated at time step $k = (K+1)$ by tracking the feature positions inside a newly acquired image frame following image rotation to compensate for the head roll angle recovered in matrix, $\mathbf{M}$. The 3-dimensional shape is then transformed by a rotation matrix $\mathbf{R}_k^{(n)}$ according to every particle,

$$\mathbf{S}_k^{(n)} = \mathbf{R}_k^{(n)}\mathbf{S} \tag{5}$$

and re-projected back inside the image space such that each feature of interest p is assigned a set of candidate coordinates, $C_k(p) = \{(c_{k,p}^{(n)}, d_{k,p}^{(n)})\} \mid n = 1, \ldots, N$. We define distance measurements, $D_k^{(n)}(\alpha)$ and $D_k^{(n)}(\beta)$, for the yaw and pitch angles respectively as follows,

$$D_k^{(n)}(\alpha) = \sum_{p=1}^{P} \left| u_{k,p} - c_{k,p}^{(n)} \right| \qquad D_k^{(n)}(\beta) = \sum_{p=1}^{P} \left| v_{k,p} - d_{k,p}^{(n)} \right| \tag{6}$$

which permit the calculation of the horizontal and vertical distances between the coordinates of the tracked feature positions, $\mathbf{u}_{k,p}$, and the candidate coordinates of the re-projected shape inside the image space, $\mathbf{c}_{k,p}^{(n)}$, according to each particle n. Based on these distances, we define the likelihood model of the particle filter corresponding to the head yaw by a normal distribution having mean, $\mu = 0$, and standard deviation, σ, as follows,

$$p(u_{k,p|1,\ldots,P} \mid \mathbf{x}_k^{(n)}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{D_k^{(n)}(\alpha)^2}{2\sigma^2}} \tag{7}$$

and similarly for the head pitch angles. The likelihood model allows each particle to be assigned a weight, $w_k^{(n)}(\alpha)$, according to the likelihood $p(u_{k,p|1,\ldots,P} \mid \mathbf{x}_k^{(n)})$ of representing the actual measurement $u_{k,p|1,\ldots,P}$ [1],

$$w_k^{(n)}(\alpha) = w_{k-1}^{(n)}(\alpha) \frac{p(u_{k,p|1,\ldots,P} \mid \mathbf{x}_k^{(n)}) p(\mathbf{x}_k^{(n)} \mid \mathbf{x}_{k-1}^{(n)})}{q(\mathbf{x}_k^{(n)} \mid \mathbf{x}_{k-1}^{(n)}, u_{k,p|1,\ldots,P})} \tag{8}$$

where, $p(\mathbf{x}_k^{(n)} \mid \mathbf{x}_{k-1}^{(n)})$ and $q(\mathbf{x}_k^{(n)} \mid \mathbf{x}_{k-1}^{(n)}, u_{k,p|1,\ldots,P})$, denote the prior probability distribution and importance function respectively. Since the distance measurements in Equation 6 consider the horizontal and vertical components of the feature positions separately, each particle is assigned weights $w_k^{(n)}(\alpha)$ and $w_k^{(n)}(\beta)$ denoting the likelihood of representing the true head yaw and pitch angles respectively. These weights are subsequently normalised such that the state, $\mathbf{x}_k = (\alpha_k, \beta_k)$, is estimated as a weighted average of the particle set for the yaw and pitch angles respectively. It is worth noting that we base the process of weighting the particles upon the shape information of the object of interest rather than its photometric properties, in order to reduce the susceptibility of the method to intensity variations and repetitive skin texture.

Following the estimation of the state, $\mathbf{x}_k = (\alpha_k, \beta_k)$, the re-projection of shape $\mathbf{S}$ inside the image space corresponding to the estimated head rotation angles permits correction of mistracked feature points, hence preventing the feature trackers from drifting off the features of interest during tracking. In turn, the correctly tracked feature points permit the estimated 3-dimensional shape to be updated at every time step via factorisation, in order to improve upon the initial estimation of the shape information. To this end, a weighted average between the re-projected shape coordinates, $\hat{\mathbf{c}}_{k,p}$, according to the state estimate, $\mathbf{x}_k = (\alpha_k, \beta_k)$, and the tracked feature positions at time step k is calculated as follows,

$$\hat{\mathbf{u}}_{k,p} = a_{k,p}\mathbf{u}_{k,p} + (1 - a_{k,p})\hat{\mathbf{c}}_{k,p} \qquad p = 1, \ldots, P \tag{9}$$

The value of the weighting parameter, $a_{k,p}$, corresponds to a measure of tracking confidence for every feature point and assumes a value between 0 and 1, with 1 denoting the highest tracking confidence. If a feature point is lost during tracking, denoted by a tracking confidence of 0, the corrected feature coordinates $\hat{\mathbf{u}}_{k,p}$ are defined entirely by the corresponding re-projected shape coordinates for that particular feature. This addresses one of the issues that is commonly associated with factorisation relating to the occurrence of missing entries inside the measurement matrix $\mathbf{W}$, hence ensuring reliable factorisation results by correcting the measurement information in real-time. The averaged coordinates, $\hat{\mathbf{u}}_{k,p}$, are finally included in the measurement matrix such that the shape information of the surface of interest is updated at every image frame by factorisation.

2.2 Implementation Details

Following an overview of the proposed algorithm in Section 2.1, the next sections describe the implementation details to extract the required information from the image frames.

Face Region Detection. The first stage in the implementation of the method detects the bounding box enclosing the face region such that this constrains the initialisation of the salient features to track, as explained in the next section. We chose the Viola-Jones algorithm for rapid detection of the face region given the real-time requirements of our application. The Viola-Jones framework combines several weak classifiers of increasing complexity into a cascade structure, where each classifier is trained by a technique called boosting to search for specific image features by classifying between positive and negative candidate image samples [19]. In our work, we employed the trained cascade classifier available in MATLAB since its detection capabilities were found to generalise well across different subjects.

Initialisation and Tracking of Feature Points. In order to track the object of interest and hence generate the feature trajectories to populate the measurement matrix $\mathbf{W}$, several feature trackers were latched upon salient facial features within the boundaries of the face region detected earlier. The chosen feature points were randomly distributed over the surface of interest and selected according to the method proposed by Shi and Tomasi [16], who define the good features to track as points characterised by a steep brightness gradient along at least two directions. The initialised salient features are subsequently tracked between successive image frames via the Kanade-Lucas-Tomasi (KLT) feature tracker, which matches search windows between consecutive image frames to identify correspondences based on a measure of similarity [17].

Particle Filter. Following the estimation of the 3-dimensional shape of the surface of interest by factorising the trajectories of salient feature points, the implemented particle filter algorithm generates hypotheses of state $\mathbf{x}_k = (\alpha_k, \beta_k)$ at every time step. To this end, we chose to implement the Bootstrap filter [1] due to its simplicity in applying the prior probability distribution, $p(\mathbf{x}_k^{(n)} \mid \mathbf{x}_{k-1}^{(n)})$, as the importance function, $q(\mathbf{x}_k^{(n)} \mid \mathbf{x}_{k-1}^{(n)}, u_{k,p|1,\ldots,P})$, hence simplifying the definition of the particle weights to,

$$w_k^{(n)}(\alpha) = w_{k-1}^{(n)}(\alpha) \frac{p(u_{k,p|1,\ldots,P} \mid \mathbf{x}_k^{(n)}) p(\mathbf{x}_k^{(n)} \mid \mathbf{x}_{k-1}^{(n)})}{q(\mathbf{x}_k^{(n)} \mid \mathbf{x}_{k-1}^{(n)}, u_{k,p|1,\ldots,P})} \tag{10}$$

for the head yaw and similarly for the head pitch angles. In order to avoid degeneration of the particle set, where all but one of the particle weights are equal to zero, a bootstrap re-sampling algorithm was implemented to re-sample

the particle set with replacement and hence preserve the particles having the highest weights at every time instance [1]. Furthermore, we approximate the state evolution of the implemented particle filter by a Gaussian random walk model that serves to propagate the particles to the next time step. Hence, the state evolution model may be defined by,

$$p(x_k \mid x_{k-1}) = \mathcal{N}(\mu_k, \sigma) \tag{11}$$

where $\mathcal{N}(.)$ denotes a Gaussian distribution having mean, $\mu_k = x_{k-1}$, and constant standard deviation, σ.

3 Experimental Results and Discussion

To evaluate the proposed head pose estimation method, we selected several video clips from the Head Pose and Eye Gaze (HPEG) Dataset owing to the availability of various head yaw and pitch rotations, and corresponding ground truth information [2]. The HPEG dataset aggregates webcam recordings of 10 different participants into two separate sets, the first of which was recorded while the participants performed various head rotations in different directions, while the second set of recordings was more focused on changes in gaze direction. Hence, we opted to evaluate our method on webcam videos selected from the first set of recordings given their relevance to our work. Each video in the set has been captured at 30 frames per second and spatial resolution of 640×480 pixels, and lasts for 10 seconds. The ground truth information has been extracted from the relative positioning of three green light emitting diodes mounted on the head and tracked across all image frames.

We compare our results to those obtained through the implementation of two additional methods. The first method estimates the yaw and pitch angles by factorising the feature trajectories generated via a standard KLT feature tracker alone, in order to evaluate the error in head pose attributed to the occurrence of outliers and missing entries in the measurement matrix from drifting or lost feature trackers respectively. The KLT algorithm is used extensively in the factorisation literature due to its ease of implementation and low computational cost, nonetheless the feature trackers tend to drift slowly off the feature of interest especially across long image sequences, or tracking is lost entirely if the feature of interest is occluded [20]. The second method is a model-based approach which adapts the geometric face model proposed by Gee et al. in [6], originally proposed to infer the gaze direction by estimating the orientation of near-frontal head poses in static paintings, to a real-time gaming application which operates by estimating the head pose in a stream of webcam image frames [15]. In their approach, Sapienza and Camilleri [15] fit a generic face model to previously detected facial features, specifically the eyes, nose and mouth regions, and subsequently estimate the head pose from the relative tracked positions of these facial features. The resulting mean absolute error (MAE) and standard deviation (SD) of the head yaw and pitch angles estimated by the proposed method

Table 1. Mean absolute error (MAE) and standard deviation (SD) of the head yaw and pitch angles estimated by the proposed method and a KLT-based method alone to generate the feature trajectories, for different subjects in the HPEG dataset.

	Proposed Method		KLT-based Method	
Subject Number	Yaw (MAE(°), SD(°))	Pitch (MAE(°), SD(°))	Yaw (MAE(°), SD(°))	Pitch (MAE(°), SD(°))
1	(3.29, 3.13)	(2.63, 1.75)	(8.32, 8.73)	(4.45, 4.03)
4	(3.04, 2.39)	(4.52, 4.04)	(12.53, 9.45)	(3.90, 2.19)
5	(7.33, 3.87)	(4.85, 4.06)	(8.76, 5.69)	(6.00, 6.14)
6	(6.05, 3.70)	(3.61, 1.95)	(4.03, 3.30)	(8.29, 4.58)
7	(4.64, 3.80)	(3.87, 1.76)	(31.20, 14.75)	(18.85, 21.29)
8	(2.86, 3.23)	(6.33, 4.83)	(20.51, 11.89)	(40.00, 45.00)
9	(2.51, 1.13)	(0.99, 0.65)	(8.61, 6.23)	(0.01, 0.01)
Mean	(4.25, 3.04)	(3.83, 2.72)	(13.42, 8.58)	(11.64, 11.89)

Table 2. Mean absolute error (MAE) and standard deviation (SD) of the head yaw and pitch angles estimated by a model-based method, for different subjects in the HPEG Dataset.

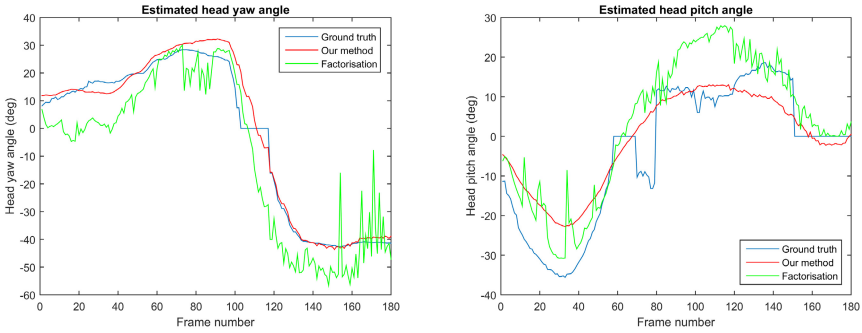
Subject Number	Yaw (MAE(°), SD(°))	Pitch (MAE(°), SD(°))
1	(5.68, 4.00)	(3.41, 3.14)
4	(5.51, 4.03)	(6.30, 2.15)
5	(10.02, 8.89)	(11.97, 9.19)
6	(4.47, 2.65)	(12.74, 5.64)
7	(5.79, 4.14)	(15.77, 7.81)
8	(11.24, 10.61)	(7.98, 5.19)
9	(5.63, 3.51)	(6.01, 0.82)
Mean	(6.90, 5.40)	(9.17, 4.85)

in Section 2 in comparison to the results obtained by the KLT-based method and the model-based method are presented in Tables 1 and 2 respectively.

The results presented in Table 1 indicate a significant reduction in the calculated MAE and SD values when the head yaw and pitch angles were estimated by the method proposed in Section 2, in comparison to the results obtained by generating the feature trajectories prior to factorisation via the KLT-based method alone. It may be noted that several of the highest MAE and SD values for the KLT-based method, such as the results for subjects 7 and 8, correspond to the widest ranges of head yaw or pitch rotations as tabulated in Table 3. The increased error corresponding to larger head rotation angles is caused by an increase in the occurrence of outliers inside the measurement matrix $\mathbf{W}$, and in the absence of a suitable mechanism that detects and corrects the outlying information, the factorisation method produces incorrect head yaw and pitch estimates. Indeed, as shown in Figure 1(d), during larger head rotations several feature points become self-occluded causing the corresponding feature trackers to gradually drift off and collect outlying information, and eventually lose the fea-

Fig. 1. Head pose estimation results obtained through our method (a-c), factorisation of the feature trajectories generated via a standard KLT feature tracker alone (d-f) and the geometric model-based method in [15] (g-i), for subject 8 in the HPEG Dataset.

ture of interest as indicated by the lost feature trackers marked in red, in Figures 1(e) and 1(f). In comparison, our method addresses this problem by exploiting the 3-dimensional shape of the surface of interest in order to correct drifting feature trackers, while permitting the trajectories of correctly tracked features to contribute towards the improvement of the 3-dimensional shape. Hence, the occurrence of outliers in the measurement matrix is reduced in real-time, which allows for increased robustness in estimating larger head rotation angles as shown in Figures 1(a-c), where the relative configuration of the feature trackers is preserved by preventing the trackers from drifting off the object of interest during head rotations. Furthermore, Figure 2 compares the head yaw and pitch angles estimated through our method to the motion information recovered by factorisation in real-time during tracking, for subject 8 in the HPEG dataset. This figure indicates a reduction in jitter for the results obtained by our method,



(a) (b)

Fig. 2. Head pose estimation results obtained through our method (red) and via factorisation during tracking (green), in comparison to ground truth data (blue) for subject 8 in the HPEG Dataset.

while higher MAE and SD values were obtained for the head yaw, (9.74, 6.53), and pitch, (7.55, 6.33), angles estimated by the factorisation algorithm in comparison to our method, hence indicating the validity of combining factorisation with particle filtering.

Furthermore, the results in Table 2 also indicate a reduction in the calculated MAE and SD values for the head yaw and pitch estimates obtained by our method, in comparison to those obtained through an implementation of the model-based method in [15]. In evaluating the results for the model-based method, it has been noted that distortion and partial occlusion of the tracked facial features of interest contribute significantly to the error in the estimated head yaw and pitch angles. Indeed, as shown in Figures 1(g-i), a leftward and rightward rotation of the head produces a displacement of the feature bounding boxes to the opposite direction from their true image positions as the appearance of these features distorts, resulting in reduced head pose estimation accuracy. As expected and similar to the KLT-based results discussed earlier, several of the highest MAE and SD values for the model-based method also correspond to large head yaw or pitch angles as tabulated in Table 3, due to increased distortion of the facial features during extensive out-of-plane head rotation. The effectiveness of our method, on the other hand, is not contingent on a specific head-model and hence a larger set of salient features to track may be better distributed over the surface of interest without being constrained to specific model landmarks. As discussed earlier, this permits the feature trackers latched onto visible feature points to collectively compensate for partially or fully occluded trackers without compromising the estimation accuracy.

Table 3. Ranges of head rotation yaw and pitch angles for different subjects in the HPEG Dataset.

Subject Number	Yaw [Min (°), Max (°)]	Pitch [Min (°), Max (°)]
1	[-27.44, 14.72]	[-21.53, 0.00]
4	[-27.57, 29.85]	[-4.63, 2.98]
5	[-33.41, 26.00]	[-36.70, 0.00]
6	[-17.21, 16.81]	[0.00, 22.30]
7	[-30.87, 39.13]	[-4.20, 30.56]
8	[-42.53, 28.42]	[18.59, -35.61]
9	[-18.17, 11.40]	[0.00, 0.00]

4 Conclusion

In this paper, we proposed a method to estimate the head pose based on the trajectories of salient feature points distributed randomly over the face region rather than specific facial features that fit the landmarks of typical face models, hence allowing larger head rotations without requiring prior training or accurate initialisation of specific feature points. In the absence of specific facial landmarks, we proposed the application of factorisation theory to the problem of head pose estimation in combination with particle filtering. This allowed us to exploit the recovered sparse 3-dimensional shape information in order to prevent the feature trackers from drifting off the features of interest, while at the same time permitting correctly tracked feature points to improve upon the initial estimation of the sparse 3-dimensional shape during tracking. The experimental results revealed a reduction in the head yaw and pitch estimation error when compared to the results obtained by a KLT-based method and a model-based method, hence indicating increased robustness especially in the presence of feature distortion and self-occlusion typically associated with larger head rotation angles.

Future work aims to focus upon increasing the estimated degrees-of-freedom of the head movement, such as translational movement which has not been considered in this work.

Acknowledgement. This work forms part of the project *Eye-Communicate* funded by the Malta Council for Science and Technology through the National Research & Innovation Programme (2012) under Research Grant No. R&I-2012-057.

References

1. Arulampalam, M., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. IEEE Transactions on Signal Processing **50**(2), 174–188 (2002)
2. Asteriadis, S., Soufleros, D., Karpouzis, K., Kollias, S.: A natural head pose and eye gaze dataset. In: Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots (AFFINE 2009) (2009)

3. Ba, S., Odobez, J.: A probabilistic framework for joint head tracking and pose estimation. In: Proceedings of the 7th International Conference on Pattern Recognition, vol. 4, pp. 264–267 (2004)
4. Chen, C., Schonfeld, D.: A particle filtering framework for joint video tracking and pose estimation. IEEE Transactions on Image Processing **19**(6), 1625–1634 (2010)
5. Fanelli, G., Gall, J., Van Gool, L.: Real time head pose estimation with random regression forests. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 617–624 (2011)
6. Gee, A., Cipolla, R.: Determining the gaze of faces in images. Image and Vision Computing **12**(10), 639–647 (1994)
7. Gurbuz, S., Oztop, E., Inoue, N.: Model free head pose estimation using stereovision. Pattern Recognition, 33–42 (2012)
8. Hansen, D.W., Ji, Q.: In the eye of the beholder: A survey of models for eyes and gaze. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(3), 478–500 (2010)
9. Ho, H., Chellappa, R.: Automatic head pose estimation using randomly projected dense sift descriptors. In: Proceedings of the 19th IEEE International Conference on Image Processing, pp. 153–156 (2012)
10. Kim, J., Kim, H., Park, R.: Head pose estimation using a coplanar face model for human computer interaction. In: Proceedings of the IEEE Conference on Consumer Electronics, pp. 560–561 (2014)
11. Kwolek, B.: Model based facial pose tracking using a particle filtering. In: Proceedings of the Geometric Modeling and Imaging - New Trends, pp. 203–208 (2006)
12. La Cascia, M., Sclaroff, S., Athitsos, V.: Fast, reliable head tracking under varying illumination: an approach based on registration of texture-mapped 3d models. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(4), 322–336 (2000)
13. Murphy-Chutorian, E., Trivedi, M.: Head pose estimation in computer vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(4), 607–626 (2009)
14. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: 3d head tracking for fall detection using a single-calibrated camera. Image and Vision Computing **31**, 246–254 (2013)
15. Sapienza, M., Camilleri, K.: Fasthpe: A recipe for quick head pose estimation. Tech. Rep. TR-SCE-2011-01, University of Malta. https://www.um.edu.mt/library/oar/handle/123456789/859 (2011)
16. Shi, J., Tomasi, C.: Good features to track. In: Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 593–600 (1994)
17. Tomasi, C., Kanade, T.: Detection and tracking of point features. Tech. Rep. CMU-CS-91-132, Carnegie Mellon University (1991)
18. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: a factorization method. International Journal of Computer Vision **9**(2), 137–154 (1992)
19. Viola, P., Jones, M.: Robust real-time object detection. International Journal of Computer Vision (2001)
20. Wang, G., Wu, Q.M.J.: Introduction to structure and motion factorization. Advances in Pattern Recognition, pp. 63–86 (2011)
21. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Biometrics Compendium, pp. 2879–2886 (2012)