EUSIPCO 2013 1569741911

HEAD POSE ESTIMATION BASED ON STEERABLE FILTERS AND LIKELIHOOD PARAMETRIZED FUNCTION

Nawal Alioua[1,3]*, Aouatif Amine*[2]*, Mohammed Rziza*[1]*, Abdelaziz Bensrhair*[3]*, Driss Aboutajdine*[1].

[1] LRIT, unité associée au CNRST, Faculty of Sciences, Mohammed V-Agdal University, Rabat, Morocco;
[2] LGS, ENSA-Kenitra, Ibn Tofail University, Morocco;
[3] LITIS, INSA-Rouen, Saint-Etienne-du-Rouvray, France.

ABSTRACT

In this paper, we propose a new discrete head pose estimator that combines appearance template and discriminative learning. Our approach consists on constructing a reference model for each considered head orientation. To do this, we first locate the head patch using a skin color based filter. The reference models are then elaborated from steerable filters which are applied in order to extract feature vectors. We chose to apply such filters since they are robust to global geometric deformations and view point changes. Next, we learn parameters of likelihood function from training data with a discriminative approach. When a new image is considered, a feature vector based on steerable filters is extracted from the localized head patch. Subsequently, head pose is estimated using likelihood parametrized function. The performance of our estimator is evaluated on PRIMA-POINTING database showing that the proposed approach is very competitive compared to other existing methods.

Index Terms— Discrete head pose estimation, steerable filters, likelihood parametrized function.

1. INTRODUCTION

Head pose estimation from images is an interesting research domain required by a large number of applications such as human-machine interfaces, game industry, driver monitoring systems and analysis of visual focus of attention. In addition, it represents a crucial step in visual gaze estimation techniques since it provides coarse indication of gaze direction. In [1], two types of head pose estimators are distinguished. The first type allows to identify few discrete orientations such as frontal view compared with left and right profiles. The second type provides a more accurate estimation presented by continuous angles values according to a fixed number of degrees of freedom. In general, head pose modeling is limited to 3 degrees of freedom which are represented by pitch (top to down movement), yaw (left to right movement) and roll (rotation). Like any facial processing approach, techniques allowing head pose estimation must be robust to some factors such as identity variation, facial expressions, lighting conditions and image resolution. In [1], head pose estimation approaches are classified into eight conceptual categories. Approaches based on appearance template perform the best match between a new input and a set of exemplars. Templates can be created using the entire image information or some specific features especially orientation-selective features which can be extracted by Gabor filters [2] or steerable filters [3]. The second category composed from detector array methods trains one detector for each pose and assigns a discrete pose to the detector providing the greatest support [4]. Approaches based on nonlinear regressions such as neuronal networks or Support Vector Regressors (SVR) [5] develop a functional mapping from image features to a head pose. Techniques based on manifold embedding methods use low-dimensional reduction to model continuous variation in head pose and ignore the other sources of image variation. Principal component Analysis (PCA) is the most popular dimensionality reduction technique used to estimate head pose [1]. Methods using flexible models [6] adjust a non-rigid model to facial structure and estimate head pose from comparison of features or model parameters. Geometric methods explore facial features such as eyes, mouth, or nose to determine head pose from their location [7]. Tracking based methods [8] retrieve pose variation from the movement between video frames. The last category contains approaches that merge at least two techniques from the previous categories in order to overcome the limitations of each single approach.

In this paper, we propose to estimate head pose using an approach that combines appearance template and likelihood parametrized function in which parameters are learned from training data. Our template representation is based on steerable filters that are robust to global geometric deformations and view point changes [3]. Another advantage of this orientation selective filtering is the ability to obtain a filtered image by linearly combining images filtered by a small set of basis filters permitting to reduce considerably the processing time. Likelihood parametrized function employed to estimate the congruence between current input and reference models is inspired from the works presented in [9, 8]. The remainder of this paper is organized as follows. Section 2 describes head

pose modeling using steerable filters. Section 3 presents the proposed head pose estimation algorithm based on Steerable Filters and Likelihood Parametrized Function (SFLPF). Section 4 exposes experimental results. Finally, section 5 concludes the paper and provides directions for future work.

2. HEAD POSE MODELING USING STEERABLE FILTERS

Modeling head pose is an essential step of pose estimation allowing to construct a representation of head appearance taking into consideration image variations produced by orientation changes. We chose to consider steerable filters to model head pose since they are able to analyze oriented structures in images. In addition, they can produce a filtered image at any orientation by linearly combining its filtered versions obtained by a small set of basis filters. This concept reduces considerably the processing time.

2.1. Steerable filters

A function $f(x,y)$ is steerable (see Eq (1)) if its rotated versions $f^{\theta}(x,y)$ around the angle θ can be expressed by a linear combination of M basis functions $f^{\theta_j}(x,y)$. $k_j(\theta)$ are the corresponding interpolation functions ($j = 1 \cdots M$).

$$f^{\theta}(x,y) = \sum_{j=1}^{M} k_j(\theta) f^{\theta_j}(x,y) \tag{1}$$

If polar coordinates are considered ($r = \sqrt{x^2+y^2}$, $\phi = arg(x,y)$) and if f can be expanded in a Fourier series in polar angle ϕ, f will be expressed by Eq (2).

$$f^{\theta}(r,\phi) = \sum_{n=-N}^{N} a_n(r) e^{in\phi} \tag{2}$$

The steering condition (see Eq (1)) holds for functions expandable in the form of Eq (2) if and only if the interpolation functions $k_j(\theta)$ are solutions of Eq (3).

$$\begin{bmatrix} 1 \\ exp(i\theta) \\ \vdots \\ exp(iN\theta) \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ exp(i\theta_1) & exp(i\theta_2) & \cdots & exp(i\theta_M) \\ \vdots & \vdots & \vdots & \vdots \\ exp(iN\theta_1) & exp(iN\theta_2) & \cdots & exp(iN\theta_M) \end{bmatrix} \begin{bmatrix} k_1(\theta) \\ k_2(\theta) \\ \vdots \\ k_M(\theta) \end{bmatrix} \tag{3}$$

From Eq (1) and Eq (2), we can get Eq (4).

$$f^{\theta}(r,\phi) = \sum_{j=1}^{M} k_j(\theta) g_j(r,\phi) \tag{4}$$

Where $g_j(r,\phi)$ can be any set of functions. The minimum number T of basis functions is the number of $a_n(r) \neq 0$ in Eq (2), which implies that $M \geq T$. If rotated versions of f are chosen as basis functions, the T orientations θ_j of basis functions spaced equally in angle between 0 and π are computed by $\theta_j = \frac{j\pi}{T}, (j = 0 \cdots T-1)$.

We chose a simple steerable function which is the circularly symmetric Gaussian function (see Eq (5)) to model head pose.

$$f(x,y) = e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{5}$$

According to Freeman et al. [10], the directional derivative operator is steerable. If we note f_n the n^{th} derivative of f, we obtain Eq (6). $f_1^{0^\circ}$ et $f_1^{90^\circ}$ are respectively represented by Fig.1-a and Fig.1-b.

$$\begin{aligned} \frac{\partial}{\partial x} f(x,y) = f_1^{0^\circ} &= -\frac{1}{\sigma^2} x e^{-\frac{(x^2+y^2)}{2\sigma^2}} \\ \frac{\partial}{\partial y} f(x,y) = f_1^{90^\circ} &= -\frac{1}{\sigma^2} y e^{-\frac{(x^2+y^2)}{2\sigma^2}} \end{aligned} \tag{6}$$

The filter f_1 at an arbitrary orientation θ can be synthesized by taking a linear combination of $f_1^{0^\circ}$ and $f_1^{90^\circ}$ which are the basis filters in this case (see Eq (7)). Fig1-c represents the filter at $\theta = 30^\circ$

$$f_1^{\theta} = cos(\theta) f_1^{0^\circ} + sin(\theta) f_1^{90^\circ} \tag{7}$$

Since convolution (denoted by $*$) is a linear operation, an image I (for example Fig.1-d) filtered at an arbitrary orientation can be synthesized by taking linear combinations of the images filtered with $f_1^{0^\circ}$ and $f_1^{90^\circ}$ (see Eq (8)) corresponding to Fig.1-e and Fig.1-f. If we consider $\theta = 30^\circ$, Fig.1-g depicts the resulting image.

$$\begin{aligned} R_1^{0^\circ} &= f_1^{0^\circ} * I \\ R_1^{90^\circ} &= f_1^{90^\circ} * I \\ R_1^{\theta} &= cos(\theta) R_1^{0^\circ} + sin(\theta) R_1^{90^\circ} \end{aligned} \tag{8}$$

2.2. Construction of head pose models

If we need to estimate K head poses, we must construct K different pose models. To obtain good models, each pose must be represented by enough training images. For each considered image, we locate head patch by a skin color filter based on connected components. Next, to reduce the noise effect, we filter the resized patch by Gaussian filter. Then, we apply steerable filters as described by Eq (8). After several experiments, we find that the best representation of head pose information is given by three filter orientations $\Theta = \{0^\circ; 50^\circ; 100^\circ\}$. The result obtained by these three filters are then concatenated in a single feature vector v_i. After processing all training images, we compute the mean E_k of training vectors associated with each head pose k given by Eq (9).

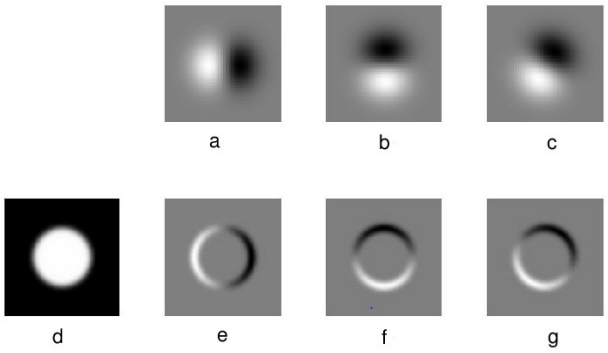

Fig. 1. Steerable filter representation [10]. (a) $f_1^{0^\circ}$; (b) $f_1^{90^\circ}$; (c) $f_1^{30^\circ}$; (d) Image of circular disk; (e) $R_1^{0^\circ}$; (f) $R_1^{90^\circ}$; (g) R_1^{θ} with $\theta = 30$.

$$E_k = mean(v_k^1, \cdots, v_k^n) \tag{9}$$

n refers to the number of training images representing pose k. Each mean vector E_k is considered as the reference model corresponding to pose k. We note ξ the total mean vectors $\xi = (E_1, \cdots, E_K)$ with K the number of head poses that must be estimated. We also compute the total diagonal covariance matrix Σ which will be used to process likelihood parametrized function (see Eq (10)).

$$\begin{aligned} \Sigma &= (\sigma_1, \cdots, \sigma_K) \\ \sigma_k &= diag(cov(v_k^1, \cdots, v_k^n)) \end{aligned} \tag{10}$$

3. HEAD POSE ESTIMATION USING LIKELIHOOD PARAMETRIZED FUNCTION

Ricci and Odobez [8] define the likelihood function as a measure of compatibility between current observation and the reference model of a specific pose. This function is expressed as a set of parameters learned offline in order to obtain a high similarity between an input and a reference model if their poses are close. In the work proposed by Toyama and Blake [11], a new method for visual tracking using exemplar-based approach and a probabilistic mechanism is introduced. Exemplars are used to represent probabilistic mixture distributions of object configurations. Instead of using standard learning algorithms, the authors employ a Metric Mixture approach based on likelihood function. Our approach is inspired from likelihood function presented by these two works but unlike them we do not address a tracking problem. In the following, we present the definition and the learning of likelihood parametrized function from training data.

3.1. Learning likelihood parameters

The likelihood function of an image characterized by its extracted feature vector v given a head pose k is expressed by Eq (11)

$$p(v|k) = \frac{1}{Z_k} e^{(-\lambda_k \rho_k(v, E_k))} \tag{11}$$

With ρ_k normalized Mahalanobis distance (see Eq (12)):

$$\rho_k(v, E_k) = \frac{1}{n} \sum_{i=1}^{n} max \left\{ \frac{(v(i) - E_k(i))}{\sigma_k}, T^2 \right\} \tag{12}$$

T is a threshold allowing to exclude inappropriate values. Z_k represents the normalization constant or partition function and λ_k corresponds to the exponential parameter. Computing these parameters is difficult in general, but straightforward when ρ_k is a quadratic function since it can be approximated by a Gaussian distribution. In this case, likelihood function parameters are given by Eq (13)

$$\begin{aligned} \lambda_k &= \frac{1}{2\delta_k^2} \\ Z_k &= \delta_k^{d_k} \end{aligned} \tag{13}$$

δ_k is an image-plane distance constant. The distance ρ_k can be considered as a random variable $\delta_k^2 \chi_{d_k}^2$ following a χ^2 distribution with δ_k its standard variation and d_k its dimension. This constraint allows the parameters δ_k and d_k to be learned from training data. For this, we construct a set F_v from training data and for each $f_v \epsilon F_v$ we determine the pose p allowing to minimize the distance ρ between f_v and all reference models E_k (see Eq (14)). We note $\rho_p(f_v) = \rho_p(f_v, E_p)$ to simplify.

$$p = arg \min_k \rho_k(f_v, E_k) \tag{14}$$

$\rho_p(f_v)$ can be estimated by $\delta_k^2 \chi_{d_p}^2$. An approximate but simple approach to estimate parameters can be done via simple moments (see Eq (15)).

$$\begin{aligned} \overline{\rho}_k &= \frac{1}{N_k} \sum_{f_v} \rho_k(f_v) \\ \overline{\rho^2}_k &= \frac{1}{N_k} \sum_{f_v} \rho_k^2(f_v) \end{aligned} \tag{15}$$

From the forms of mean and standard deviation of χ^2 statistic, δ_k and d_k can be estimated by Eq (16).

$$\begin{aligned} d_k &= 2 \frac{\overline{\rho}_k^2}{\overline{\rho^2}_k - \overline{\rho}_k^2} \\ \delta_k &= \sqrt{\frac{\overline{\rho}_k}{d_k}} \end{aligned} \tag{16}$$

3.2. Estimating head pose of an input image

We consider that the reference models for each head pose are determined and the parameters of likelihood function are learned. When a new image is presented, we apply the following procedure:

- Locate head patch by the same skin color filter used to construct the reference models .
- Reduce noise effect by applying a Gaussian filter.
- Apply steerable filters to construct the feature vector v using the three orientations described in section 2.2.
- Compute likelihood between v and all reference models as described by Eq (11). The estimated head pose k^* is chosen according to Eq (17).

$$k^* = arg \max_k p(v|k) \tag{17}$$

4. EXPERIMENTAL RESULTS

To evaluate our head pose estimation approach, we consider PRIMA-POINTING database [7] which represents 15 different subjects in 93 discrete head poses. For each subject, two series of images in all specified poses was acquired. Head orientations are described by pitch included in the set $\{0; \pm 90; \pm 60; \pm 30; \pm 15\}$ and yaw belonging to the interval $[-90°; +90°]$ with a displacement of $15°$.

We conduct two types of experiments. The first experimental setup is proposed in [8] as CLEAR evaluation workshop protocol where the first serie of each subject is used as training set and the second one as test set. This experiment is the most used in literature to evaluate head pose estimators when PRIMA-POINTING database is considered for evaluation process. In the second experiment, we avoid to consider the same subject in training and test sets. This experiment is conducted to evaluate if our estimator can recognize poses even if the subject is not represented in the training set. To enhance the two experiments, we propose to locate head patches manually and using skin color filter. Manually head localisation is proposed to evaluate more precisely our pose estimator without including errors that can be generated by automatic head localisation. We present experimental results of head pose estimator as the average of absolute difference between the ground truth and the estimated pose. Hence, results are exposed as mean errors of pitch and yaw angles.

4.1. Experimental results using CLEAR evaluation protocol

In this experiment, the first serie of each subject is used as training set. In other words, 15 images are considered to construct one reference model corresponding to one pose (see section 2.2). For each test image, we apply the process defined in section 3.2 to determine the estimated head pose k^*. Fig. 2 represents a frontal image example ($pitch = yaw = 0°$) and its corresponding features extracted using steerable filters with three orientations $\Theta = \{0°; 50°; 100°\}$. Fig. 3 shows the reference model extracted from all frontal heads in the training set.

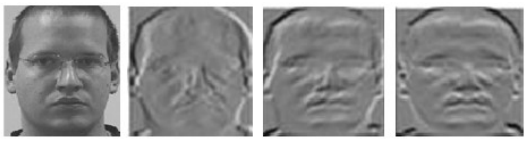

Fig. 2. Frontal image and its corresponding features using steerable filters with three orientations $\Theta = \{0°; 50°; 100°\}$

Fig. 3. Reference model extracted from frontal heads in the training set

Table 1 shows results of our proposed approach compared to other methods that can perform head pose estimation. The term (A) corresponds to automatic head localisation while (M) refers to manually head localisation. According to the results presented in Table 1, the proposed head estimation method based on steerable filters and likelihood parametrized function is competitive with respect to several methods in literature.

Table 1. Head pose estimation error using CLEAR evaluation protocol

	Pitch (°)	**Yaw** (°)
SFLPF (A)	12.4	9.6
SFLPF (M)	10.1	8.7
Ricci et al. [8]	14.2	13.7
Gourier et al. [7]	15.9	10.3
Tu et al. [12]	17.9	12.9
Ba et al. (A) [9]	14.1	13.2
Ba et al. (M) [9]	11.1	11.1

4.2. Experimental results using unseen subject for test

CLEAR evaluation protocol does not provide strong information about the ability of our approach to estimate poses of unseen subjects. To obtain more reliable results, we defined

the second experiment protocol in which the two series of one subject are conserved for test while the remaining subjects are considered for training. To improve the experiment, we use the same protocol for each subject and we present the result as the average of angular errors obtained for all tests. Table 2 reports the performances of our approach using unseen subjects for test compared to the unique work proposing the same protocol [9]. We also report a recent result obtained in [5] even if the experiment is not done using the different protocol (80% of the database is used for training and 20% for test). We can conclude that the performance of the second evaluation protocol is close to results obtained by CLEAR evaluation protocol.

Table 2. Head pose estimation error using unseen subjects for test

	Pitch (°)	**Yaw** (°)
SFLPF (A)	13.8	11
SFLPF (M)	11.4	9.97
Ba et al. (A) [9]	14.4	11.7
Ba et al. (M) [9]	14.5	12.1
Al Haj et al. (M) [5]	10.52	11.29

For a new test image, the average processing time needed to estimate head pose when head patch is detected manually is approximatively 0.04 seconds. When head patch is located automatically, the average processing time does not exceed 0.065 seconds. Notice that these processing times are obtained with a non-optimized Matlab code running on an 2GHz PC.

5. CONCLUSION

We describe in this paper a new approach to estimate discrete head poses. Our proposed method begins by determining a reference model based on steerable filters for each considered head pose. Then, we learn parameters of likelihood function from a subset of training data involved to construct reference models. When a new image is presented, a head patch is extracted and a feature vector is computed based on steerable filters. Subsequently, head pose is rapidly estimated using likelihood parametrized function. Good performances are achieved by this approach evaluated using PRIMA-POINTING database and compared to other methods proposed in literature. As future work, we plan to adapt our head pose estimator in order to integrate it in our previous system that monitor driver vigilance level [13].

6. REFERENCES

[1] E. Murphy-Chutorian and M.M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Transactions Pattern Analysis and Machine Intelligence*, pp. 607–626, 2009.

[2] J. Foytik and V.K. Asari, "A two-layer framework for piecewise linear manifold-based head pose estimation," *International Journal of Computer Vision*, pp. 270–287, 2013.

[3] M. Dahmane and J. Meunier, "Oriented-filters based head pose estimation," in *Canadian Conference on Computer and Robot Vision (CRV)*, 2007, pp. 418–425.

[4] G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L.J. Van Gool, "Random forests for real time 3d face analysis," *International Journal of Computer Vision*, pp. 437–458, 2013.

[5] M. Al Haj, J. Gonzalez, and L.S Davis, "On partial least squares in head pose estimation: How to simultaneously deal with misalignment," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[6] J. Wu and M.M. Trivedi, "A two-stage head pose estimation framework and evaluation," *Pattern Recognition*, vol. 41, pp. 1138–1158, 2008.

[7] N. Gourier, D. Hall, and J.L. Crowley, "Estimating face orientation from robust detection of salient facial features," in *International Workshop on Visual Observation of Deictic Gestures (Pointing)*, 2004.

[8] E. Ricci and J.M. Odobez, "Learning large margin likelihoods for realtime head pose tracking," in *International Conference on Image Processing (ICIP)*, 2009, pp. 2593–2596.

[9] S.O. Ba, "Joint head tracking and pose estimation for visual focus of attention recognition," *These Ecole Polytechnique Federale de Lausanne EPFL, n.3764*, 2007.

[10] W.T. Freeman and E.H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 891–906, 1991.

[11] K. Toyama and A. Blake, "Probabilistic tracking with exemplars in a metric space," *International Journal of Computer Vision*, pp. 9–19, 2002.

[12] J. Tu, Y. Fu, Y. Hu, and T. Huang, "Evaluation of head pose estimation for studio data," in *International evaluation conference on CLassification of Events, Activities and Relationships (CLEAR)*, 2006.

[13] N. Alioua, A. Amine, M. Rziza, and D. Aboutajdine, "Driver's fatigue and drowsiness detection to reduce traffic accidents on road," in *International Conference on Computer Analysis of Images and Patterns (CAIP)*, 2011, pp. 397–404.