

REVIEW

Open Access



Entropy-weighted feature-fusion method for head-pose estimation

Xiao-Meng Wang*, Kang Liu and Xu Qian

Abstract

This paper proposes a novel entropy-weighted Gabor-phase congruency (EWGP) feature descriptor for head-pose estimation on the basis of feature fusion. Gabor features are robust and invariant to differences in orientation and illuminance but are not sufficient to express the amplitude character in images. By contrast, phase congruency (PC) functions work well in amplitude expression. Both illuminance and amplitude vary over distinctive regions. Here, we employ entropy information to evaluate orientation and amplitude to execute feature fusion. More specifically, entropy is used to represent the randomness and content of information. For the first time, we seek to utilize entropy as weight information to fuse the Gabor and phase matrices in every region. The proposed EWGP feature matrix was verified on Pointing'04 and FacePix. The experimental results demonstrate that our method is superior to the state of the art in terms of MSE, MAE, and time cost.

Keywords: EWGP, Head-pose estimation, Entropy weighted, Gabor, Phase congruency, Feature fusion

1 Review

1.1 Introduction

Visual focus of attention (VFoA) is emphasized to estimate at what or whom a person is looking and is highly correlated with head-pose estimation [1]. To study head-pose estimation, three-dimensional orientation parameters from human head images are explored. Head poses convey an abundance of information in natural interpersonal communication (NIC) and human-computer interaction (HCI) [2]; therefore, an increasing number of researchers is seeking more effective and robust methodologies to estimate head pose. Head poses also play a critical role in artificial intelligence (AI) applications and reveal considerable latent significance of personal intent. For example, people nod their heads to represent understanding in conversations and shake their heads to show dissent, confusion, or consideration. Head orientation with a specific finger-pointing direction generally indicates the place that a person wants to go. The combination of head pose and hand gestures is used to assess the target of an individual's interest [3]. Mutual orientation indicates that people are involved in discussion. If a person shifts the head toward a specific direction, it is highly likely that there is an object of interest in this direction. Therefore, the study of VFoA

as an indicator of conversation target in human-computer interaction and facial-expression recognition is increasingly of interest.

Analyzing head poses is a natural capability of humans but is difficult for AI. However, head-pose estimation has been researched for years, and the state of the art in head-pose estimation can contribute greatly to bridging the gap between humans and AI [4, 5]. Head-pose estimation is generally interpreted as the capability to infer orientation relative to the observation camera. For example, head pose is exploited to determine the focus point on the screen based on the gaze direction [6]. The factors influencing the estimation of head pose and their relationships have been introduced in detail, and the crucial significance of head pose was emphasized in [7]. These factors are mostly related to the surroundings, including camera calibration, head features, glasses, hair, beard, illuminance variations, and image transformations.

To address the shortcomings of existing methods, we concentrate on regional feature extraction based on entropy information. We aim to utilize an information entropy model to assess randomness and content as feature metrics for a specific region for the first time. We then employ the more adaptive feature to represent the virtual region. In addition, the normalized entropy information is regarded as a weight metric to fuse the ultimate feature

* Correspondence: xiao_meng_wang@163.com
China University of Mining and Technology, Beijing 100083, China

matrix. The experimental results demonstrate that our feature matrix is superior to the state-of-the-art.

This paper is structured as follows: Section 1.2 provides an exhaustive overview of previous related work in head-pose estimation. Section 1.3 presents the proposed methodology step by step, including a skin model for face detection using Gabor features, PC features, and entropy-weighted Gabor phase congruency (EWGP). Section 1.4 describes the experiments using the Pointing'04 dataset. Finally, in Section 2, we present our conclusions and discuss future work.

1.2 Related work

Despite significant improvements in head-pose estimation techniques [5, 6, 8, 9], robust and effective recognition remains challenging for various poses. Head motion can be characterized by 6 degrees of freedom (DOF), as shown in the schematic representation in Fig. 1. Under the common assumption of regarding the head as a rigid object in a fixed position, we also address the head as a rigid ball with 3 DOF. We ignore translation in the horizontal and vertical planes. Then, the head pose can be categorized into pitch, roll, and yaw, which are indicated with red arrows in Fig. 1.

In general, head-pose estimation approaches can be classified into two types: coarse level and fine level [5]. The former commonly employ algorithms to calculate a few discrete head orientations, such as left, right, and looking up. The latter generally utilize methodologies to compute the continuous pose in accurate angles. Here, we redefine the coarse level and fine level: coarse-level approaches recognize the head-orientation variations using discrete estimation and accurate computation, and

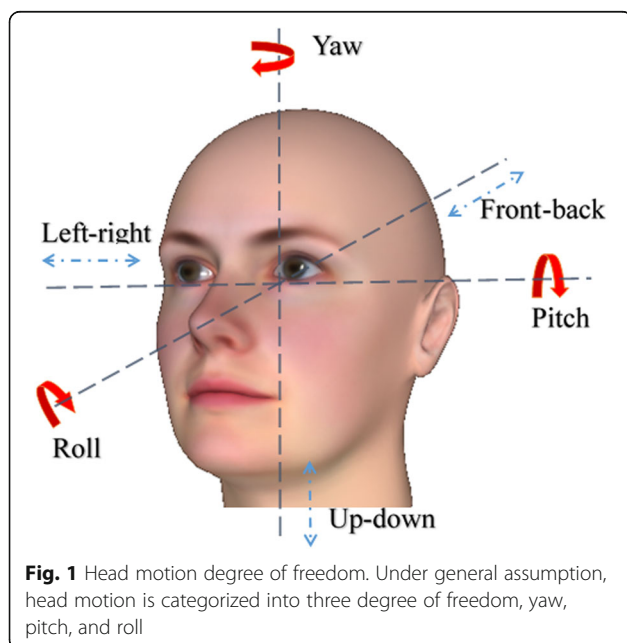
fine-level approaches indicate the intentions or interests of the experimental subjects. The computational approaches of both layers can be divided into statistical and non-statistical types based on their dependence on statistical methods or not.

1.2.1 Statistical approaches

The most classical statistical method is to exploit classifiers or regression methods to recognize specific discrete head poses. Multi-classification tools such as a support vector machine (SVM) are utilized to estimate discrete head poses. SVM has been employed to locate the iris centers in approximately detected eye regions [10] and to distinguish frontal and look-up head-pose variations in a Carnegie Mellon University (CMU) face dataset [11]. Support vector regression (SVR) is an alternate version that is used for the continuous problem. The differences in head-pose estimation between SVM and SVR have been described in detail [12]. SVR performs well for either horizontal or vertical head-pose variations, whereas SVM performs better for vertical variations than for horizontal. If the search range is not extensive, the combination of SVM and SVR is a good option. In addition, whenever the number of classes changes, the SVMs must be re-trained from scratch.

Regression is another typical statistical method that is available for both discrete and continuous head-orientation angle estimation. Examples of regression approaches include the aforementioned SVR and multi-layer perceptrons (MLP). Regression approaches are classified as linear and nonlinear based on the causal relationships between independent variables and dependent variables. An MLP can also be trained for fine head-pose estimation over a continuous pose range. In this configuration, the network has one output for each DOF. The activation of the output is proportional to its corresponding orientation [13–15]. The high dimension of an image presents a challenge for some regression tools. More specially, regression methods cannot resolve the need for long, sophisticated training and are highly sharply sensitive to head localization. In summary, dimension reduction via principle component analysis (PCA) [16] or its nonlinear kernel version (KPCA) [17] or localized gradient-orientation histograms [18] is necessary during the above procedure.

Instead of comparing images to a large set of discrete class labels or a series regression values, the probe image can be measured by a detector array that is also trained on many images with supervised learning methods. Detector array methods are well suited for both high- and low-resolution images. In addition, they are superior in sub-regional operations. Most importantly, these methods do not require separate head detection and localization. The drawbacks of these schemes are the



necessary scale of training, binary output of detectors, and low accuracy; in practice, a maximum of 12 different detectors can be formed, which limits the pose-estimation definition to less than 12 states [5].

High-dimensional image samples can lie on a low-dimensional manifold that is constrained to meet the pose variations. Manifold-embedding methodologies, including isometric feature mapping (Isomap) [19, 20], locally linear embedding (LLE) [21], and Laplacian eigenmaps (LE) [22], have shown promise for head-pose estimation by mapping high-dimensional data into low-dimensional space. Such low-dimensional spaces can be formed by classification or regression. However, the limitation of typical PCA is not averted for nonlinear head-pose variations. Since unsupervised methods are utilized during the classification or regression, these methods are not available to incorporate the class labels during head-pose training. Most importantly, the aforementioned techniques cannot ensure that each class is expressed as a single label.

1.2.2 Non-statistical approaches

Experimental results have revealed considerable differences between statistical methods and non-statistical measurements [23–26]. The former mainly focus on appearance-based measurements, whereas the latter usually consider geometric relationship cues, such as the deviation of the nose from the mid-line and the deviation between the new head pose and the original state. In non-statistical methods, flexible models, geometric information, and motion trajectory are employed to estimate head pose.

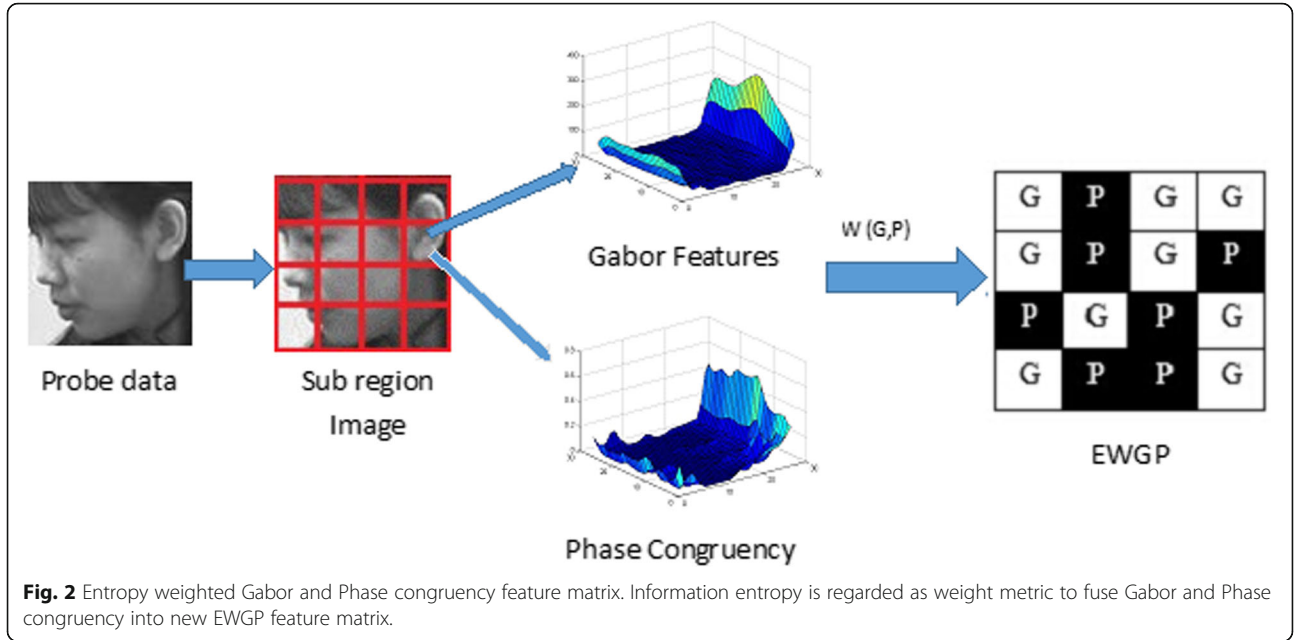
Flexible models seek to fit non-rigid models with facial features and contribute to the exploration of the facial structure in both discrete and continuous head orientations. Among flexible models, active shape models (ASM) [27, 28] and active appearance models (AAM) exhibit higher accuracy and robustness [29]. These approaches permit the direct prediction of head pose when an inherent 3D model constrains the fitting of 2D points. Combination of the 3D model and 2D points enables direct head-pose computation using structure-from-motion algorithms. In summary, flexible models have great potential for both high accuracy and good robustness in head-pose estimation, but these qualities are strictly correlated with the relative extracted feature positions and image definition. Additionally, geometric methods exploit relative feature positions to estimate head pose; however, the accuracy is highly related to the feature-point extraction [30]. Importantly, the highest accuracies of the presented approaches are at least 1–2 pixels. Unfortunately, each pixel error generally relates to an angle error of approximately 5°. Consequently, geometric measurements cannot serve as precise head-

pose estimates in cases of limited feature-point detection. The use of motion-trajectory tracking methods between subsequent video frames outperforms the other aforementioned methods [27, 31, 32]. In previous work, we employed a SIFT feature-point and bio compound-eye mechanism to explore object-tracking measurements with superior robustness and accuracy [33]. Tracking methods operate in a bottom-up manner, following low-level facial landmarks from frame to frame. Typically, the subject must maintain a frontal pose before the system started. The track system must be reinitialized whenever the object of interest is lost. As a result, geometrical approaches often rely on manual initialization or a camera view in which the subject's neutral head pose is forward-looking and easily reinitialized with a frontal face detector [5]. Recently, a number of hybrid approaches have been proposed [33–38] that integrate the remarkable advances in the above statistical and non-statistical methods to provide the best accuracy and robustness in head-pose estimation.

Our proposed method is a hybrid approach, and we seek to estimate head poses on the coarse level to compute the orientation angles using some machine classifiers and geometrical information. Information entropy is a good indicator of information representation with respect to randomness and content. Histogram of gradient (HoG), Gabor, and phase congruency (PC) are effectively and commonly used in direction estimation. However, the dimensionality of these feature matrices is usually too high for image representation. With the development of technology, image-definition has increased abruptly. More specifically, dimension-disaster frequency has clearly risen. This paper presents an entropy-weighted method to fuse Gabor and PC features and exploits entropy as a weight metric to reinforce randomness for the first time. Additionally, entropy plays an important role in dimension reduction and image annotation. The experimental results prove that our solution is effective in reducing the dimension and shows good accuracy and robustness to variations of head pose.

1.3 Proposed methodology

Face detection is an important preceding step for head-pose estimation [3, 32, 39, 40]. A schematic representation of our method is shown in Fig. 2. A skin model is exploited to perform face detection and abstraction, and on the basis of the image blocks, probe images are divided into sub-regions. In our case, an elliptical skin-color model is employed to detect facial regions in YCbCr color space. Initially, the probe input image is split into various sub-regions, and the features of each sub-region are then extracted using Gabor and PC methods. Gabor features and PC features show remarkable advantages in representing orientation and gradient



information of images. Information entropy is capable of assessing feature randomness and content. Therefore, information entropy is utilized to take advantage of Gabor and PC features. Next, a new image-feature matrix, called the entropy-weighted Gabor and PC feature matrix (EWGP), is constructed. Ultimately, the matrix is used as a classifier to estimate head pose on the coarse level. Most importantly, the matrix shows promising results in dimension reduction and pose estimation, with improved effectiveness and robustness.

1.3.1 Skin-color model

The use of skin-color characteristics in face detection is receiving an increasing amount of attention. An HSV-YCbCr color-space skin model was suggested in [41]. An adaptive skin-color model in the normalized RGB space was examined in [42]. In this paper, we employ an elliptical skin model in a non-linearly transformed YCbCr color space, which was proposed in [43]. This algorithm detects face regions in probe images with good performance, and its core operation is described in Eqs. (1) and (2).

$$\frac{(x-ec_x)^2}{a^2} + \frac{(y-ec_y)^2}{b^2} = 1 \quad (1)$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} C_b - c_x \\ C_r - c_y \end{bmatrix} \quad (2)$$

where ec_x and ec_y represent the offset from the original position. Equation (2) describes the transformation with the angle θ from YCbCr to two-dimensional coordinates. C_b and C_r denote the skin cluster space. These parameters were calculated as $c_x = 109.38$, $c_y = 152.02$,

$\theta = 2.53$ (rad), $ec_x = 1.60$, $ec_y = 2.41$, $a = 25.39$, and $b = 14.03$. In addition, a Gaussian distribution is used to model the skin tones in the transformed space by the vector Z in Eq. (3), which denotes the skin-cluster space. Therefore, the skin-pixel classification rule can be reformulated using Eq. (4). We regard the current methodology as a binary classification, since we only need to determine whether the pixel is in the skin cluster or not. μ is a two-dimensional vector that describes the mean value of Z , and Σ is the covariance of Z . The left side of Eq. (4) in the first line is the Mahalanobis distance. λ indicates whether the pixel is skin or not following the elliptical model, which is an empirical threshold. This information can also be obtained by some other threshold method, such as OTSU. Here, we combine the elliptical model and Gaussian model to construct the face region more accurately. From the elliptical model, we can obtain the closed interval of the skin region. This region is simply the linear function in terms of the expectation and standard deviation in the Gaussian model. Additionally, the Gaussian distribution is exploited to generate neighborhood face orientations using manifold-embedding methods and to assign multi-labels to the images [44].

$$Z = \begin{bmatrix} C_b \\ C_r \end{bmatrix} \quad (3)$$

$$\varphi(Z) = \begin{cases} 1 & \text{if } (Z-\mu)^t \Sigma^{-1} (Z-\mu) \leq \lambda, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

$$\lambda = \frac{(x-ec_x)^2}{a^2} + \frac{(y-ec_y)^2}{b^2} \leq 1$$

1.3.2 Gabor features

Gabor features are widely employed in head-pose estimation for their advantages in the calculation of orientation and frequency. Gabor features are also used to alleviate variations in illumination in highly tunable properties. In addition, diverse sets of parameters are obtained to create multi-kernels in terms of the constructed series of filters. The Gabor transform is applied with the assistance of the window Fourier transform, which extracts features in different directions and scales from an image in the frequency domain. The covariance of the Gabor descriptor and regional symmetry information were applied to estimate head orientation, and principal component analysis (PCA) was employed to reduce the dimension of the data, obtaining promising results. In our case, the original Gabor filter is utilized [45]. The Gabor filter is defined in Eq. (5), where σ is the standard deviation of the Gaussian kernel, which is typically linearly dependent on the wavelength; λ is the wavelength of the sinusoid; θ denotes the orientation of the filter; and x and y are the spatial position coordinates of the filter.

$$g(x, y, \lambda, \theta) = e^{-\frac{x^2+y^2}{\sigma^2}} e^{\frac{2\pi i}{\lambda}(x \cos \theta + y \sin \theta)}. \quad (5)$$

1.3.3 Phase congruency

Gabor-based features cannot be identified for a sizable portion of the features within an image on which current pose-sensitive systems depend. PC-based edge descriptors, which are sensitive to pose variations, can be used

to process a higher density of features. The discrete Fourier transform based PC-based edge descriptor (6) produces a higher density of features and is sensitive to head orientation.

$$PC(x) = \frac{\sum_{i=1}^n A_i \cos(\varphi_i(x) - \varphi(x))}{\sum_{i=1}^n A_i} \quad (6)$$

$$\varphi_i(x) = Tw_i x$$

A_i indicates the amplitude of the i th Fourier component, $\varphi_i(x)$ represents the i th local phase of the components, and $\varphi(x)$ is the weighted mean of all local phase angles at the objective location. Additionally, for each frequency w_i , A_i is the amplitude of the cosine wave, and $\varphi_i(x) - \varphi(x)$ is the phase offset of that wave. The term T is related to the size of the image window, and we will assume it a value of 1. It is important to assume that phase-congruency features differ from one another when dealing with different head-orientation probe images. Consequently, it is necessary to distinguish which filter orientation is more effective in pose estimation. In our case, the Pointing'04 head-pose dataset was utilized to evaluate the phase-congruency features after face detection by the eclipse skin model. To this end, binary-edge images were collected.

1.3.4 EWGP feature fusion

Face extraction, Gabor features, and PC features play important roles in head-pose estimation. After the processing described in Sections 1.3.1 to 1.3.3, Gabor and PC maps are collected for a specific image, which are

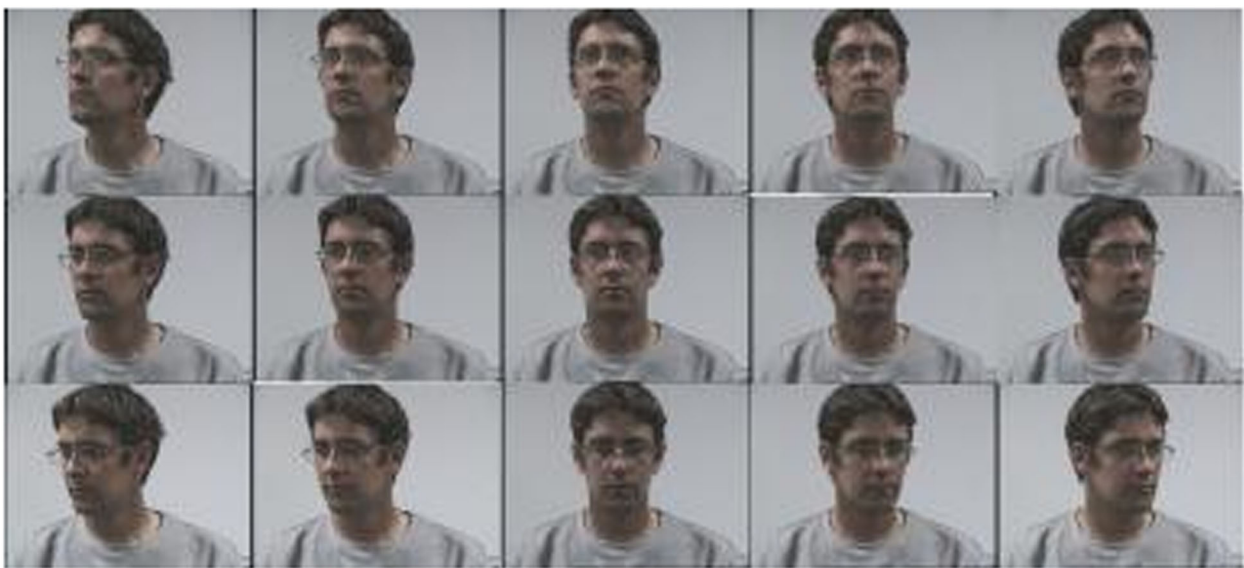
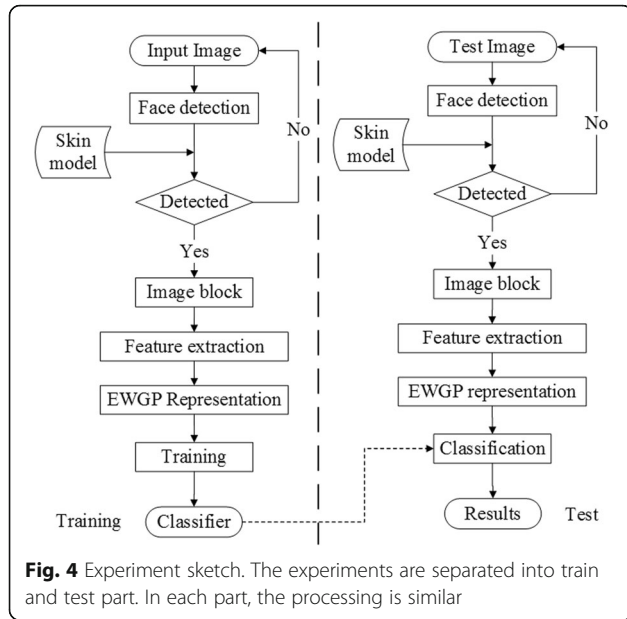


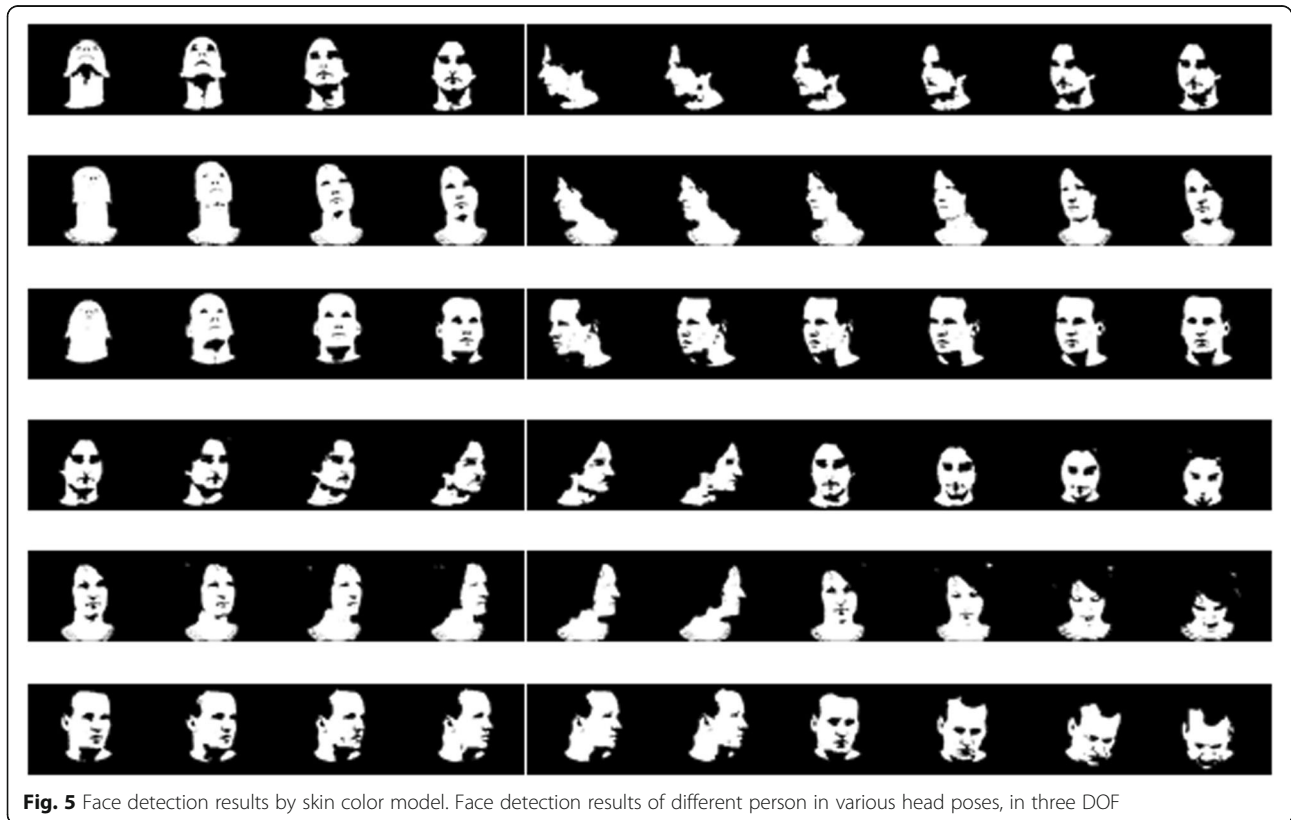
Fig. 3 Sample of Pointing'04 Database. An example of the dataset. Each set contains of 2 series of 93 images of the same person at different poses. There are 15 people in the database, wearing glasses or not and having various skin color. The pose, or head orientation is determined by 2 directions yaw and pitch, which varies from -90° to $+90^\circ$



relevant on the fine level. Region-based methods are highly effective and robust in head-pose estimation. A local directional quaternary patterns (LDQP) descriptor has been proposed to estimate head pose and obtained surprising results. In addition, the whole image must be reconstructed with complex and uncertain patterns [46].

Consequently, we employ regional thinking and implement feature extraction in each sub-region. Both Gabor and PC features are computed and are described by the distributions R_G and R_P in Eqs. (5) and (6), respectively, in the specific sub-region R_I . The features exhibit different similarities with the original data. Gabor features are robust and invariant to different levels of illuminance but are not sufficient to express the amplitude character. By contrast, phase features function well for amplitude expression. Both illuminance and amplitude vary over different regions. Here, we regard entropy information as the metric to evaluate the two aforementioned features. More specifically, entropy information represents the randomness and information content. Different principles are applied to measure the distance or similarity between two distributions using the entropy. For instance, the distance between R_G and R_I can be represented by Jeffrey's divergence, given in Eq. (7). Similarly, Eq. (8) is given for R_P and R_I . The two distributions must be normalized to the same scale before the similarity calculation is performed. Finally, the results of (7) and (8) are normalized.

$$D_{Jg}(R_G||R_I) = \sum (R_G - R_I) \ln \frac{R_G}{R_I}, \quad (7)$$



$$D_{Ip}(R_p||R_l) = \sum (R_p - R_l) \ln \frac{R_p}{R_l}, \quad (8)$$

$$F = W_{ij}(D_{Ig}, D_{Ip}) [R_{G,ij} R_{P,ij}]. \quad (9)$$

A larger value of D represents a closer relationship with the probe representation. We employ Jeffrey's entropy as the weight to construct the new feature matrix. Meanwhile, dimension-reduction operations are utilized to optimize the Gabor feature matrix and PC feature matrix, such as PCA and SVD. In summary, in our case, the advantages of Gabor features and PC features are combined for the first time to estimate head pose in Eq. (9), where W is the normalized entropy weight for Gabor and PC in the $i \times j$ th sub-region.

1.4 Experiments and analysis of results

In the experiments, we first utilized the Pointing'04 head-pose database, which comprises 15 sets of images. Each set contains 2 series of 93 images of the same

Table 1 Estimation results on Pointing'04

Features	MSE (SVR)	CVA (SVM) %
HOG	3.13	85.88
Gabor	3.32	25.92
PC	3.24	86.88
EWGP	0.93	96.79

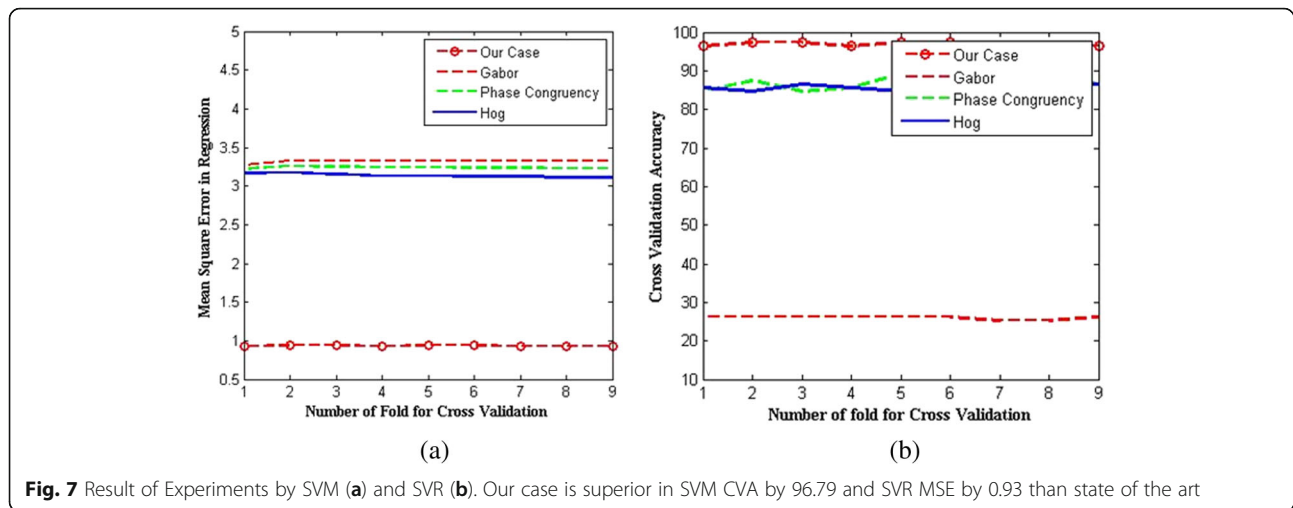
List best result of head-pose estimation of MSE and CVA. Our case outperform others by 0.93 and 97.79% in MSE and CVA for SVR and SVM, respectively

person in different poses. The database includes images of 15 people who are wearing or not wearing glasses and who have various skin colors. The pose or head orientation is determined in 2 directions, yaw and pitch, which vary from -90° to $+90^\circ$. An example is given in Fig. 3.

A sketch of the experiment is given in Fig. 4. The probe images contain the neck and shoulders, resulting in lower classification accuracy, and thus the elliptical skin model is employed to extract the primary facial region and capture the eye, mouth, and nose information.



Fig. 6 Experimental results of EWGP. Experiment results on public data Pointing'04, and the results are binary images



Then, the face region is divided into $m \times m$ sub-regions. Next, Gabor and PC features are extracted. Information entropy is utilized as a weight metric to fuse the Gabor and PC features into the EWGP representation. Finally, the EWGP matrix data are used to train multi-classifiers to distinguish the head-pose variations.

In terms of the skin model, we conducted experiments using the Pointing'04 database and extracted head information at the fine level. We tested this model with three different people in different directions. We tested three pitch and ten yaw directions for each person under conditions of varying levels of illuminance. In each direction, we executed experiments on three people. We obtained some promising results. First, we completely captured the eyes, nose, mouth and ears of the various persons in different directions. Second, after the binary operation, we clearly captured, the skin color information for different head-pose variations in both the pitch and yaw directions, as shown in Fig. 5.

On the basis of face extraction, the EWGP was compared with state-of-the-art that are utilized in head-pose estimation, such as HOG, Gabor, and PC by SVM and SVR. Experiments were conducted in both the yaw and pitch directions via k-fold cross validation, which means that the whole extracted face images used in the experiments were partitioned by a factor, related to the variation, into k parts, of which k-1 parts were used for training, and the remaining part was used for testing. The experimental results for the EWGP filter are illustrated in Fig. 6. Two types of evaluation measures, cross-validation accuracy (CVA) and mean square error (MSE), were used to compare the performance of the different descriptors. The head-pose estimation results for the Pointing'04 database were compared. The best results for MSE and CVA are shown in Table 1. Detailed information on the MSE and CVA states is depicted in Fig. 7. The HOG, Gabor, and PC descriptors yielded

similar results for MSE. However, Gabor exhibited an apparent decrease in CVA of 25.92%, which we tentatively attribute to the use of only one direction filter for all head-orientation images; this will be confirmed in future research. Our proposed hybrid representation outperformed other descriptors with respect to MSE and CVA by 0.93 and 96.79, respectively, as shown in Table 1. These experimental results indicate that the proposed EWGP representation is suitable for head-pose estimation in the yaw and pitch directions. The proposed method was also compared with other methods. Regardless of whether SVM or SVR was employed, the MAE of our method on Pointing'04 was enhanced three to fivefold, as shown in Table 2.

A pose-estimation system was implemented utilizing simple linear-subspace methods and oriented Gabor and PC features in [47]. Global Gabor and PC features were employed as global and local measurements, respectively, to estimate head orientations with Canonical Correlation Analysis (CCA) on FacePix datasets. The proposed EWGP method achieved better MAE. The comparison is illustrated in Fig. 8. The EWGP method

Table 2 Comparison of head-pose estimation results on the Point'04 database

Method	MAE	
	Yaw	Pitch
EWGP-SVM	1.03°	1.00°
EWGP-SVR	1.12°	1.31°
MLD-wJ	4.24°	2.69°
Kernel PLS	5.02°	3.54°
Kernel SVM	6.83°	5.91°
Kernel SVR	6.89°	6.59°

The proposed EWGP feature was tested by SVM with the MAE 1.03° and 1.00° in Yaw and Pitch direction. When verified by SVR MAE are 1.12° and 1.31°, respectively

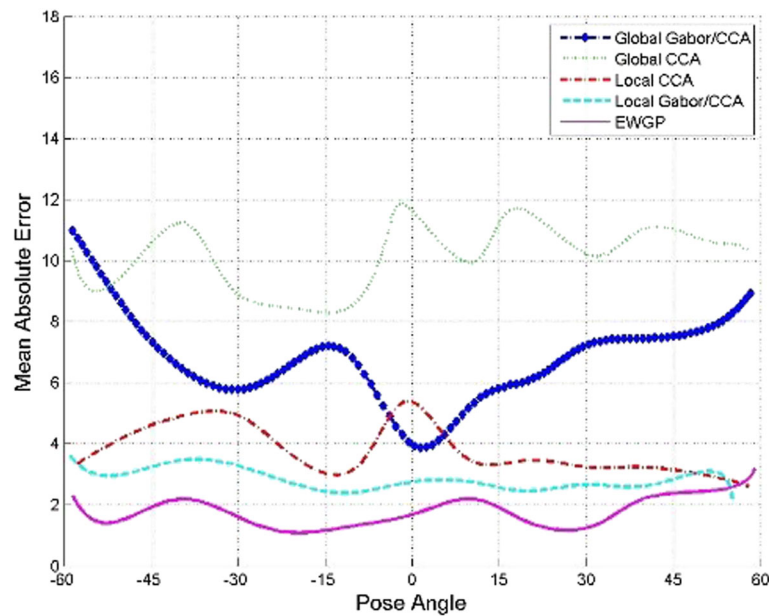


Fig. 8 Experimental results MAE of EWGP, Global CCA, Local CCA, Global Gabor/CCA, and Local Gabor/CCA. The overall MAEs are 2, 10, 8, 4, 2.8 in degrees, respectively. EWGP outperform the state of the art that also utilized Gabor and PC as feature matrix

exhibited better performance from -60 to 60 , and the MAE deviated by 2° . The overall MAEs of Global CCA, Global Gabor of CCA, Local CCA, and Local Gabor of CCA were 10° , 8° , 4° , and 2.8° , respectively. MAE increased in both directions as the pose angle increased. When the head turns with large deviation, facial features are obscured, and only a partial face image can be obtained with a single camera. Thus, key face points are missed.

Furthermore, the experiments were performed on both homogeneous and heterogeneous data. We utilized classifiers that were trained on the Pointing'04 datasets on a real-time video stream that was obtained using a consumer-level webcam (320×240) on a Windows 10 (64 bit) platform, with 8G Ram and Intel(R) Core(TM) i7-3635QM CPU@ 2.40GHz. Promising results were

obtained. Both single faces and multiple faces were detected in real time, and we captured the nose region. The results are illustrated in Fig. 9. Simultaneously, head poses were also calculated with satisfactory conclusions, as shown in Fig. 10. In the yellow label, h denotes the horizontal yaw degree, and v represents the vertical rotation degree, or pitch. For processing video information, the face-detection event-capture rate was 100%, and the maximum head-pose estimation time cost was 50 ms for each frame. The lowest time cost was approximately 5–10 ms.

2 Conclusions

In this study, a novel entropy-weighted Gabor and phase-congruency (EWGP) feature matrix was built on the condition of feature fusion. We successfully applied

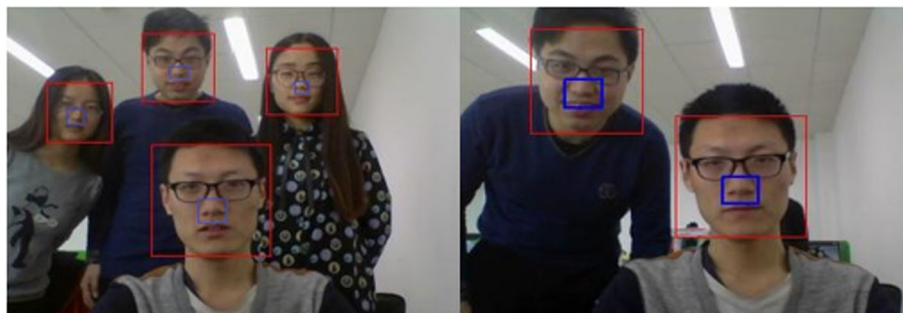


Fig. 9 Head extraction in real time. We detect both single face and multi-faces with ourselves video data, and the nose labeled by blue rectangle in real time



Fig. 10 Head-pose estimation in real time. The person yaw to left by 29.54° and pitch to up by 50.53° in left image. The person yaw to left by 8.22° and pitch to up by 18.89° in right image

EWGP in multi-classification for head-pose estimation in still imagery and a real-time video stream with homogeneous and heterogeneous data. Our experimental results demonstrated that the proposed EWGP method outperforms state-of-the-art when estimating head pose in terms of MSE, CVA, MAE, and time cost. Unfortunately, head pose only describes the direction in which a person is looking and does not provide information on the object of interest. Therefore, it is necessary to focus on additional information, such as visual saliency in head orientation, gaze direction, and hand gestures. In future works, we plan to expand head-pose estimation to include gaze estimation and obtain a better understanding of the object of an individual's interest.

Abbreviations

3D: Three dimensional; AI: Artificial intelligence; DOF: Degree of freedom; EWGP: Entropy-weighted Gabor and phase congruency; HCI: Human-computer interaction; Isomap: Isometric feature mapping; KPCA: Kernel principle component analysis; LE: Laplacian eigenmaps; MLP: Multi-layer perceptrons; NIC: Natural interpersonal communication; PC: Phase congruency; PCA: Principle component analysis; SVM: Support vector machine; SVR: Support vector regression; VfoA: Visual focus of attention

Acknowledgements

Thanks to the partners for providing their test results. Gratitude is also expressed to my mentor Xu Qian for his valuable and helpful suggestions.

Authors' contributions

XMW proposed the main idea to fuse Gabor and phase-congruency features using information entropy and participated in carrying out the experiments using Pointing'04 datasets to verify the proposed method. KL performed some experiments to verify the relationship between the skin-color model and face extraction and participated in extracting face regions from the Pointing'04 and FacePix datasets. XQ helped with the statistical analysis and drafting of the manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 2 March 2016 Accepted: 2 December 2016

Published online: 09 December 2016

References

1. SO Ba, J-M Odobez, Multiperson visual focus of attention from head pose and meeting contextual cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(1), 101–116 (2011)
2. JJ Magee, M Betke, EyeKeys: A real-time vision interface based on gaze detection from a low-grade video camera. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern. Recognit. Workshops*. **2004**, 1–8 (2004).
3. M Pateraki, H Baltzakis, P Trahanias, Visual estimation of pointed targets for robot guidance via fusion of face pose and hand orientation. *Computer Vision and Image Understanding*. **120**, 1–13 (2014)
4. H Huttunen et al., Computer vision for head pose estimation: review of a competition. *Image Analysis. 19th Scandinavian Conference, SCIA 2015*, 15–17 June 2015 **2015**, 65–75 (2015)
5. E Murphy-Chutorian, MM Trivedi, Head pose estimation in computer vision: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**(4), 607–626 (2009)
6. DW Hansen, J Qiang, 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)
7. CG Healey, JT Enns, Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics* **18**(7), 1170–88 (2012)
8. W Xiaomeng, L Kang, Q Xu, A survey on gaze estimation. *10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). Proceedings*, 2015, pp. 260–7
9. MV Sireesha, PA Vijaya, K Chellamma, A survey on gaze estimation techniques. *Lecture Notes in Electrical Engineering*, v 258 LNEE, 2013, pp. 353–361
10. MH Nguyen, J Perez, FDL Torre, Facial feature detection with optimal pixel reduction SVM. *2008 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008*, September 17, 2008 - September 19, 2008, (2008).
11. T Sim, S Baker, M Bsat, *The CMU pose, illumination, and expression (PIE) database*. *5th IEEE International Conference on Automatic Face Gesture Recognition, FGR 2002*, May 20, 2002 - May 21, 2002, 2002, pp. 53–58
12. G Guodong et al., *Head pose estimation: classification or regression?* *19th International Conference on Pattern Recognition, ICPR 2008*, 8–11 Dec. 2008, 2008, p. 4
13. Y Xinguo et al., *Head pose estimation in thermal images for human and robot interaction*. *2010 2nd International Conference on Industrial Mechatronics and Automation (ICIMA 2010)*, 30–31 May 2010, 2010, pp. 698–701
14. E Seemann, K Nickel, R Stiefelhagen, *Head pose estimation using stereo vision for human-robot interaction*. *Proceedings - Sixth IEEE International Conference on Automatic Face and Gesture Recognition FGR 2004*, May 17, 2004 - May 19, 2004, 2004, pp. 626–631
15. R Stiefelhagen, J Yang, A Waibel, Modeling focus of attention for meeting indexing based on multiple cues. *IEEE Transactions on Neural Networks*. **13**(4), 928–938 (2002)
16. Y Li, S Gong, H Liddell, *Support vector regression and classification based multi-view face detection and recognition*. *4th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2000*, March 28, 2000 - March 30, 2000, 2000, pp. 300–305
17. R Rosipal, M Girolami, L Trejo, Kernel PCA feature extraction of event-related potentials for human signal detection performance. *Computing and Information Systems*. **7**(1), 20–3 (2000)
18. E Murphy-Chutorian, A Doshi, MM Trivedi, *Head pose estimation for driver assistance systems: a robust algorithm and experimental evaluation*. *10th International IEEE Conference on Intelligent Transportation Systems, ITSC 2007*, September 30, 2007 - October 3, 2007, 2007, pp. 709–714

19. B Raytchev, I Yoda, K Sakaue, *Head pose estimation by nonlinear manifold learning*. Proceedings of the 17th International Conference on Pattern Recognition, 23-26 Aug. 2004, 2004, pp. 462–6
20. VN Balasubramanian, S Krishna, S Panchanathan, Person-independent head pose estimation using biased manifold embedding. *EURASIP J. Adv. Signal Process.* **15**, 1–15 (2008)
21. ST Roweis, LK Saul, Nonlinear dimensionality reduction by locally linear embedding. *Science*. **290**(5500), 2323–6 (2000)
22. M Belkin, P Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*. **15**(6), 1373–96 (2003)
23. K Woo Won et al., *Automatic head pose estimation from a single camera using projective geometry*. 2011 8th International Conference on Information, Communications & Signal Processing (ICICS 2011), 13–16 Dec. 2011, 2011, p. 5
24. AV Puri, H Kannan, P Kalra, *Coarse head pose estimation using image abstraction*. 2012 Canadian Conference on Computer and Robot Vision, 28–30 May 2012, 2012, pp. 125–30
25. S Sheikhi, J-M Odobez, Combining dynamic head pose-gaze mapping with the robot conversational state for attention recognition in human-robot interactions. *Pattern Recognition Letters*. **66**, 81–90 (2015)
26. S Sheikhi, J-M Odobez, *Investigating the midline effect for visual focus of attention recognition*. 14th ACM International Conference on Multimodal Interaction, ICMi 2012, October 22, 2012 - October 26, 2012, 2012, pp. 221–224
27. J Min et al., *Head pose estimation based on Active Shape Model and Relevant Vector Machine*. 2012 IEEE International Conference on Systems, Man and Cybernetics (SMC2012), 14–17 Oct. 2012, 2012, pp. 1035–8
28. Y Chen et al., *A method of head pose estimation based on Active Shape Model and stereo vision*. 2014 33rd Chinese Control Conference (CCC), 28–30 July 2014, 2014, pp. 8277–82
29. N Mahmoudian Bidgoli, AA Raie, M Naraghi, Probabilistic principal component analysis for texture modelling of adaptive active appearance models and its application for head pose estimation. *IET Computer Vision* **9**(1), 51–62 (2015)
30. HR Wilson et al., Perception of head orientation. *Vision Research* **40**(5), 459–72 (2000)
31. G Zhibo et al., *A fast algorithm face detection and head pose estimation for driver assistant system*. 2006 8th International Conference on Signal Processing, 16–20 Nov. 2006, 2006, p. 5
32. G Garau et al., *Investigating the use of visual focus of attention for audio-visual speaker diarisation*. 17th ACM International Conference on Multimedia, MM'09, with Co-located Workshops and Symposia, October 19, 2009 - October 24, 2009, 2009, pp. 681–684
33. XM Wang et al., *Moving object detection based on bionic compound eye*. 2014 International Conference on Materials Science and Computational Engineering, ICMSCE 2014, May 20, 2014 - May 21, 2014, 2014, pp. 3563–3567
34. P Yao, G Evans, A Calway, *Using affine correspondence to estimate 3-D facial pose*. IEEE International Conference on Image Processing (ICIP), October 7, 2001 - October 10, 2001, 2001, pp. 919–922
35. MAA Dewan et al., Adaptive appearance model tracking for still-to-video face recognition. *Pattern Recognition*. **49**, 129–51 (2016)
36. Sujono, AAS Gunawan, Face expression detection on Kinect using Active Appearance Model and fuzzy logic. *Procedia Computer Science* **59**, 268–74 (2015)
37. M Linna, J Kannala, E Rahtu, *Online face recognition system based on local binary patterns and facial landmark tracking*, 2015, pp. 403–414
38. C Luo et al., *Video based face tracking and animation*. 8th International Conference on Image and Graphics, ICIG 2015, August 13, 2015 - August 16, 2015, 2015, pp. 522–533
39. F Bazyari, Y Tzimiropoulos, *An active patch model for real world appearance reconstruction*. Computer Vision - ECCV 2014 Workshops, 6–12 Sept. 2014, 2014, pp. 443–56
40. K Otsuka, *Multimodal conversation scene analysis for understanding people's communicative behaviors in face-to-face meetings*. Human Interface and the Management of Information: Interacting with Information - Symposium on Human Interface 2011, Held as Part of HCI International 2011, July 9, 2011 - July 14, 2011, 2011, pp. 171–179
41. C Garcia, G Tziritis, Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Transactions on Multimedia*. **1**(3), 264–77 (1999)
42. M Soriano et al., Adaptive skin color modeling using the skin locus for selecting training pixels. *Pattern Recognition*. **36**(3), 681–690 (2002)
43. R-L Hsu, M Abdel-Mottaleb, AK Jain, Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(5), 696–706 (2002)
44. G Xin, X Yu, *Head pose estimation based on multivariate label distribution*. 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 23–28 June 2014, 2014, pp. 1837–42
45. M Bingpeng et al., CovGa: a novel descriptor based on symmetry of regions for head pose estimation. *Neurocomputing*. **143**, 97–108 (2014)
46. D ByungOk Han, HS Yang, Head pose estimation using image abstraction and local directional quaternary patterns for multiclass classification. *Pattern Recognition Letters* **45**, 145–53 (2014)
47. F Jacob, VK Asari, A two-layer framework for piecewise linear manifold-based head pose estimation. *Int J Comput Vis* **101**, 270–287 (2013)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com