

Generic and Robust Method for Head Pose Estimation

Abdellatif Hajraoui*, Mohamed Sabri

Faculty of Science and Technology, University Sultan Moulay Slimane, Beni Mellal, Morocco

*Corresponding author, e-mail: abdohajraoui@gmail.com

Abstract

Head pose estimation has fascinated the research community due to its application in facial motion capture, human-computer interaction and video conferencing. It is a pre-requisite to gaze tracking, face recognition, and facial expression analysis. In this paper, we present a generic and robust method for model-based global 2D head pose estimation from single RGB image. In our approach we use of the one part the Gabor filters to conceive a robust pose descriptor to illumination and facial expression variations, and that target the pose information. Moreover, we ensure the classification of these descriptors using a SVM classifier. The approach has proved effective view the rate for the correct pose estimations that we got.

Keywords: pose estimation, face detection, gabor filters, SVM, PCA-LDA.

Copyright © 2016 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

Knowing the pose of the head is an essential link in processing chain of many applications because it offers machine new capacity for analysis and interaction. Head pose estimation is widely used in many applications such as: biometrics (face recognition), video surveillance, road safety (monitoring head for the driver's attention control), behavioral analysis for marketing, augmented reality, video games field, assistive technology. Is given in Figure 1 some examples of applications for estimating the pose of the head. Knowing the pose of the head is an essential link in processing chain of many applications because it offers machine new capacity for analysis and interaction. Head pose estimation is widely used in many applications such as: biometrics (face recognition), video surveillance, road safety (monitoring head for the driver's attention control), behavioral analysis for marketing, augmented reality, video games field, assistive technology. Is given in Figure 1 some examples of applications for estimating the pose of the head.



Figure 1. Some Examples of Applications For Estimating the Pose of the Head: Analysis of Driver Vigilance, Virtual Test of Glasses with Augmented Reality, Behavioral Analysis Using Video Surveillance

Head pose estimation is intrinsically linked with visual gaze estimation, i.e., the ability to characterize the direction and focus of a person's eyes. Geometrically talking, automatic pose estimation of the head in pictures or videos is to define an associated landmark in the face and estimate its rotation relative to a reference landmark. This could for example be associated with the body and defined by the planes: sagittal, frontal and transverse (Figure 2). One may also

choose an independent landmark body: it may be the position of the capture camera as a reference marker (Figure 2) or a face of the reference position (the face image, for example) (Figure 3). Commonly found a final performance that involves only set a direction in space with the rotation angles of the head: pan, tilt and roll (Figure 2).

The pose estimation is an area of research has been studied since 1994 [1]. It is an open research axis attracting researchers within different disciplines: psychology, image processing, artificial intelligence, robotics, computer vision, computer graphics, etc. This diversity of research disciplines hatched many methods for estimating pose. These methods are global in nature if they use the entire image 2D of a face in order to deduce its pose by means of a descriptor pose. In the literature, the descriptors exploited to estimate the pose, are usually inspired tools used in the analysis of faces or forms, such as: Sobel or Canny filters which are used to extract the facial contours, Gabor filters, nonlinear operator Local Binary Patterns (LBP), the form descriptors (Fourier descriptors, geometrical moments) The color histograms, etc. Among these methods may be mentioned those elaborated in the works: [2-6] and [7]. Otherwise, there are approaches that estimate the pose by aligning a 3D model. Notable techniques for 3D head pose estimation employ features, pose-specific classifiers or registration to reference 3D head models [8-14].

To have a thorough idea about other methods, the reader can refer to the survey [1] and state of the art in the PhD thesis [15] that investigates and compares a large number of methods for estimating the pose developed in recent years. Similarly, the work [16] presents an analysis and an evaluation of some approaches using different types of regression.

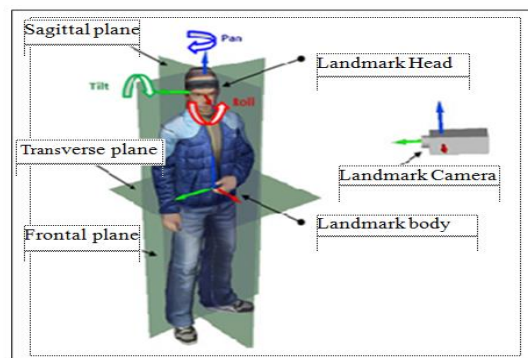


Figure 2. Illustration of Some Reference Landmarks Used in the Pose Estimation: Landmark Body, Rotation Angles and Landmark Camera

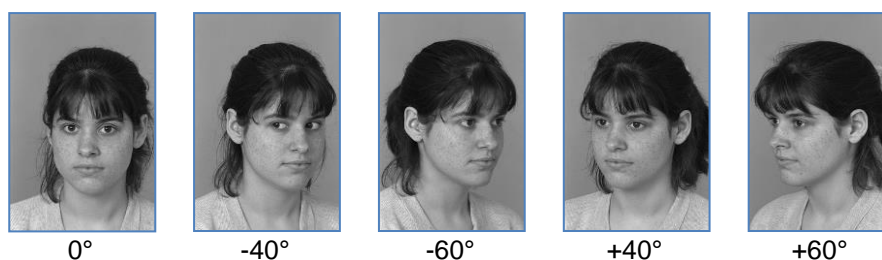


Figure 3. Head Pose with Respect to the Frontal Plane of the Face as Reference

The implementation of a pose estimation system of the head fulfilling the criteria of reliability and robustness remains a very open problem. Indeed, there are still many challenges to overcome. In particular, in an uncontrolled environment where the appearance of faces varies widely, and this can also be related to conditions in which images are acquired and the quality of the acquisition itself. This variability is not tolerated by the majority of proposed solutions that only work in controlled environments. That is to say that all conditions can degrade the expected performance by its solutions are mastered.

Our work in such problematic. Its objective is design an algorithm for an automated system for head pose estimation. This algorithm must be robust to many sources of variations that can affect the image of a face: variations related to facial expressions, illumination change, occlusion, variations related to the device acquisition.

This article presents a new approach to estimate the pose of the head. This proposed approach is part of the global learning methods. First, we use the Gabor filters to extract a robust descriptor of pose from the 2D face image. And secondly, we used the pose templates to ensure learning an SVM classifier. The results obtained with this approach proved its efficiency in terms of reliability and robustness of character when tested on images of faces that have large variations of expression and illumination.

2. Proposed Approach for Pose Estimation

Before starting the presentation of the proposed approach for pose estimation, we note that the objective is to estimate a discrete way the pose and give the value closest to that of the face from following values: **-90°, -60°, -30°, 0°, +30°, +60°, +90°** and not the exact value of the actual pose of the face. Moreover, in our case, we take the front plane of the face as a reference.

2.1. Basic Principle of the Proposed Approach

The main idea of our design pose estimator is based on the construction of a classifier with supervised learning using support vectors machine (SVM). To ensure learning of this classifier, first of all we collect some face images (pose templates) so as to form seven groups. Each group contains face images that have the closest orientations to finally define 7 classes $\{0^\circ, 30^\circ, \dots, 90^\circ\}$. Then these images undergo treatment to extract for each a pose descriptor (figure 4). And it is these descriptors, after reduction that will be stimulated the learning of our SVM classifier. At the presentation of a face image which the pose estimates (the face is not necessarily available when learning the SVM classifier) at the entrance of the classifier, it determines its orientation class of belonging from its pose descriptor. In other words, the class that provides a score of the highest similarity.

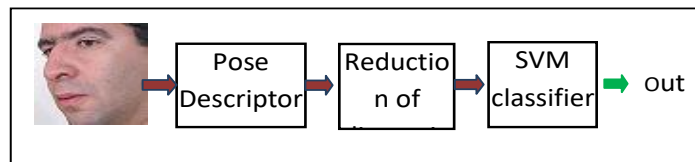


Figure 4. Block Diagram of Basic Principle of the Proposed Approach

2.2. The Developed Descriptor of Pose

The role of a descriptor pose is to extract relevant information in the face image for best modeling pose. This new model must be sensitive to changes in poses and invariant to other sources of variation of the face image (lighting, facial expression, illocutions ...). Concerning our pose estimator, we opted for the Gabor filters. These have been introduced in the field of analysis of facial images for the first time by Lades and al [17] for detecting face characteristics. Then they became a powerful tool in many algorithms of face recognition [18] [19] and face detection [20] [21]. The work that exploited Gabor filters to extract relevant information from facial images argues that this representation is robust to changes caused by variations in illumination or changes facial expressions.

The majority of work use a Gabor filter bank constructed from the expression following set:

$$G_{u,v}(x, y) = \frac{f_u^2}{\pi k \rho} e^{-\left[\left(\frac{f_u^2}{k^2}\right)x^2 + \left(\frac{f_u^2}{\rho^2}\right)y^2\right]} e^{j2\pi f_u x} \quad (1)$$

$$\begin{cases} x' = x \cdot \cos \theta + y \cdot \sin \theta \\ y' = -x \cdot \sin \theta + y \cdot \cos \theta \end{cases} \tag{2}$$

$$f_u = f_{max} / 2^{\left(\frac{u}{2}\right)} \tag{3}$$

$$\theta_v = v\pi/8 \tag{4}$$

f_u and θ_v , respectively designate the center frequency and the orientation of the plane wave parameterized by u and v . The parameters k and ρ determine the ratio between the center frequency and the size of the Gaussian envelope along the x and y axes respectively (the standard deviations of the Gaussian: σ_x and σ_y). Choosing different values for these parameters allows building the Gabor filter bank.

To select filters that will be optimal for carrying out the descriptor pose (better discriminate of the pose) and at the same time minimize the number of filters to use to gain speed and size of the descriptor, we made the following treatment :

We first built a bank of 40 filters from the expression 1, corresponding to:

1.8 orientations: $\theta_v = v\pi/8$, $v=\{0,1,2,\dots,7\}$ (5)

2.5 center frequencies: $f_u = 0.25 / 2^{(u/2)}$, $u=\{0,1,2,3,4\}$ (6)

Then we proceeded to a variety of filtering operations of several face images in different poses with the filters of the bank (Figure 5). The result of such experimental work shows that the Gabor filters with respectively the center frequency and the orientation angle (u, v): (1,4), (2,4) and (3,4) are capable of targeting the pose information in the face image. Indeed these filters to extract descriptors vectors of pose which ensure discrimination between different classes of pose.

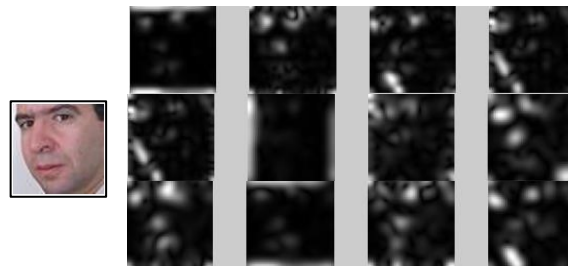


Figure 5. Example of Result of A Filtering Operation for Face RGB Image with Different Gabor Filters of the Bank

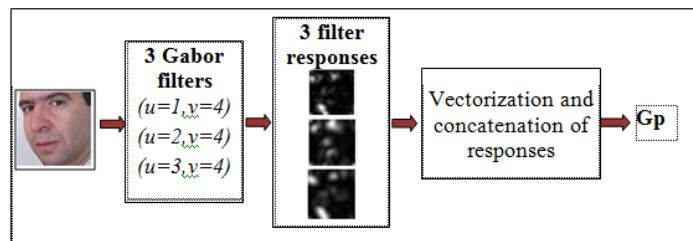


Figure 6. Block Diagram of the Pose Descriptor Developed

Modeling the pose of a face by the Gabor filters is carried out with the process illustrated in Figure 6. The face image which pose to estimate is submitted to 3 Gabor filters. The 3 filter responses are then converted into 3 vectors and thereafter concatenated into a single vector named G_p . It is this resulting vector that represents the pose descriptor.

2.3. Dimension Reduction of the Pose Descriptor Vector

The vector of pose descriptor derived from 3 Gabor filters after concatenation has a large size. ($3 \times l \times h$, with (l, h) is the size of the window of a Gabor filter). It would therefore be necessary to reduce its size. That is why we opted for an algorithm of dimension reduction that combines the two techniques known in this area: Principal Component Analysis (PCA) and Linear Discriminate Analysis (LDA). These two techniques seek a low-dimensional projection subspace that best represents the variations of the original data. The purpose of that combination is to preserve the advantages of both methods. For one, these two algorithms are based on well-known statistical properties using linear algebra. They are relatively quick to implement, especially in the online phase, which consists of a simple projection (matrix product). Furthermore, The PCA reduces the dimensionality of the space of representation of the input image while optimizing preservation of relevant information that image and the LDA to retrieve feature vectors that best discriminant poverant the data from several classes. This algorithm proposed and described in detail in our work [22], has proved her effectiveness in dimensionality reduction while ensuring discrimination of features vectors.

2.4. SVM Classifier

The support vectors machines (SVM) are a set of learning techniques to solve classification problems. In other words, decide which class a sample of a database containing several classes. SVM were introduced by Vladimir Vapnik in 1995 [23]. SVM currently attracting much attention in the community of machine learning, this proves their gain in popularity and use in many applications.

For details on obtaining certain mathematical formulas necessary for the implementation of the SVM or an understanding of SVM, the reader may refer to the memory of the thesis [24] the pages: 31-37 and 53-56.

In this particular application of SVM, we wish to discriminate 7 classes each of which is labeled with a fitting value among the values: -90° , -60° , -30° , 0° , $+30^\circ$, $+60^\circ$, $+90^\circ$. Samples will be to classify the descriptors vectors pose after reduction.

During the learning phase, we determine the hyperplanes separators 7 classes, while for the pose estimation stage (Test), a target face is assigned to an installation according to the position of its vector descriptor arises in the space partitioned into separate regions by the SVM hyperplanes.

In this work, the SVM implemented have the following parameters:

1. $K = (K-1) / 2$ binary and nonlinear classifiers SVM (where $K=7$ is the number of classes).
2. The adopted Kernel function is: Polynomial.
3. The multi-class classification method is: One versus One.

3. Results and Analysis

We end the presentation of this work by an experimental protocol that will allow evaluating the performance of the approach proposed for pose estimation of the head. Therefore, validate the scientific gait used to develop it. To achieve this objective, we use the Color FERET database [25] for the implementation of this evaluation. The Color FERET base contains a total of 14 126 images of 1199 people. This is an extension of the FERET base. The images were acquired under different conditions of expression, pose, lighting and time. Front face images are grouped into 5 partitions and profile face images are grouped in a named FERET Pose. The different partitions of the base Color FERET we used are:

1. The FA partition: set of 994 face images in front view, taken at the first session FERET.
2. The FB partition: set containing 992 images. The subjects are the same as those of the FA partition except that the facial expressions are different.
3. The FC partition: set containing 194 images. It contains images of some people with the FA partition but with variations in illumination.

4. The set FERET Pose : contains images of 200 people in different poses per person (-90°, -60°, -40°, -25°, -15°, 0°, +15°, +25°, +40°, +60, +90°) . It is this set (figure 7) that we used to ensure the learning of our estimator pose.

In the face images of Color FERET Database there is the presence of a background. To locate and extract only the thumbnails of faces, we used an algorithm for detection and localization of the face (Figure 8). This algorithm which gave very good results, we have developed in the work [20].



Figure 7. Examples of Face Images
Extracted from the Color FERET Database

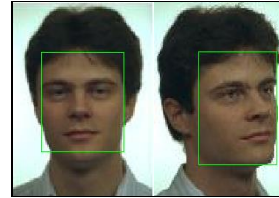


Figure. 8 Examples of Face Detection and
Location by the Algorithm [21]

The structure of the proposed pose estimator is designed to estimate a discrete way the pose of a head and give the nearest value among the 7 following values: -90°, -60°, -30°, 0°, +30°, +60°, +90°. But in tests, we adopted the 7 following values: {-90°, -60°, -25°, 0°, +25°, +60°, +90°}. This choice is imposed by the pose magnitudes available in Color FERET database.

To ensure learning SVM classifier of the estimator, we collected a set of 70 face images from the set FERET Pose (templates poses) to form seven groups. Each group contains 10 face images that have the closest possible orientations to finally define 7 classes {-90°, -60°, -25°, 0°, +25°, +60°, +90°}.

During testing, we submitted to the input of the pose estimator a series of face images randomly extracted from the FERET Color database. The faces in these images have distinct poses: -90°, -60°, -40°, -25°, -15°, 0°, +15°, +25°, +40°, +60, +90° and with different variations: facial expression, illumination.

As evaluation of the proposed approach, we measure the two following rates:

1. The right estimation rate (RER):

$$RER = \frac{\text{Number of correct estimates}}{\text{Total number of test images}} \quad (7)$$

2. The false estimations rate (FER):

$$RRE = \frac{\text{Number of false estimates}}{\text{Total number of test images}} \quad (8)$$

The Statistics of the experimental results obtained by our pose estimator are summarized in the Table 1.

Table 1. The Statistics of the Experimental Results

Partition and value of the test pose	Number of test images	RER (%)	RRE (%)
FA Patition (0°)	60	100	0
FB Patition (0°)	60	98.33	1.66
FC Patition (0°)	60	98.33	1.66
Set FERET Pose (11 different poses per person)	110	98.18	1.81

In analyzing the results presented in Table 1 summarize the estimation rates found in the four partitions of the Color FERET database, we note that estimation rate of 98.33% obtained in the FB partition (variation of facial expression) and FC partition (illumination variations) are very satisfying. This proves the robustness of the proposed approach even in the presence of large variations in the appearance of the face. Moreover, the rate of 98.18% achieved at the Set FERET Pose with 11 different poses affirms precision and reliability approach.

4. Conclusion

In this paper, we have presented a generic and robust method for head pose estimation from single RGB Image. The results obtained with this approach proved its robustness and reliability. This affirms the effectiveness of the tools built into this approach. On the one hand, the introduction of Gabor filters to design a descriptor pose while optimizing the number of filters to use, has can target the pose information in the face image even in the presence of variations of facial expression or illumination variations. On the other hand, a better classification is provided by the SVM after size reduction of the pose descriptors vectors using PCA and LDA.

References

- [1] EM Chutorian, MM Trivedi. Head Pose Estimation in Computer Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002; 31: 607–626.
- [2] Y Wei, et al. *Head pose estimation using gabor eigenspace modeling*. International Conference on Image Processing (ICIP'02). 2002; 1: 281-284.
- [3] B Ma, et al. *Robust head pose estimation using lgbp*. International Conference on Pattern Recognition (ICPR '06). 2006; 512-515.
- [4] EM Chutorian, et al. *Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation*. Intelligent Transportation Systems Conference (ITSC'07). 2007; 709-714.
- [5] N Malagavi, et al. Head Pose Estimation Using Convolutional Neural Networks. *International Journal of Innovative Science, Engineering & Technology*. 2014 ; 1(6): 470-475.
- [6] A Dahmane al. Head pose estimation based on face symmetry analysis. *Signal, Image and Video Processing*. 2015; 9(8): 1871-1880.
- [7] H Hatem, et al. Head Pose Estimation Based On Detecting Facial Features. *International Journal of Multimedia and Ubiquitous Engineering*. 2015; 10(3): 311-322.
- [8] JG Wang, E Sung. Em enhancement of 3d head pose estimated by point at infinity. *Image and Vision Computing*. 2007; 25(12): 1864-1874.
- [9] T Bar, et al. *Driver head pose and gaze estimation based on multi-template icp 3-d point cloud alignment*. IEEE International Conference on Intelligent Transportation Systems (ITSC 2012). 2012: 1797-1802.
- [10] M Martin, et al. *Real time head model creation and head pose estimation on consumer depth cameras*. International Conference on 3D Vision. 2014: 41-648.
- [11] M Nabati, A Behrad. 3D Head pose estimation and camera mouse implementation using a monocular video camera. *Signal, Image and Video Processing*. 2015; 9(1): 39-44.
- [12] GP Meyer, et al. *Robust Model-based 3D Head Pose Estimation*. IEEE International Conference on Computer Vision, Santiago. 2015: 3649–3657.
- [13] Y Cai, et al. Robust head pose estimation using a 3d morphable model. *Mathematical Problems in Engineering*. 2015: 1-10.
- [14] X Liu, et al. *3D head pose estimation with convolutional neural network trained on synthetic images*. IEEE International Conference on Image Processing (ICIP'2016). 2016: 1289-1293.
- [15] K Bailly. Learning methods for estimating the pose of the head in the monocular image. PhD thesis, University of Pierre and Marie Curie, Paris 6. 2010.
- [16] A Jaiswal, et al. Analysis and evaluation of regression-based methods for facial pose classification. *International Journal of Applied Pattern Recognition*. 2015; 2(1): 24-45.
- [17] M Lades, et al. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*. 1993; 42(3): 300-310.
- [18] H Kusuma, N Wirawan, A. Soeprijanto. Gabor based Face Recognition with Illumination Variation Using Subspace Linear Discriminant Analysis. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(1): 119-128.
- [19] SN Randive, AB Gonde. A Novel Approach for Face Recognition Using Fusion of Local Gabor Patterns. *International Journal of Electrical and Computer Engineering (IJECE)*. 2012; 2(3): 345-352.
- [20] A Hajraoui, M Sabri. Face Detection Algorithm based on Skin Detection, Watershed Method and Gabor Filters. *International Journal of Computer Applications*. 2014; 94(6): 33-39.

-
- [21] C Lin, et al. Face Detection Algorithm Based on Multiorientation Gabor Filters and Feature Fusion. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(10): 5986-5994.
 - [22] A Hajraoui, et al. A new approach for Face Recognition Based on PCA & Double LDA Treatment combined with SVM. *IOSR Journal of Engineering*. 2012; 2(4): 685-691.
 - [23] VN Vapnik. The Nature of Statistical Learning Theory. *Springer-Verlag New York*. 1995.
 - [24] M Feuilloy. Study of machine learning algorithms for the prediction of syncope in humans. PhD thesis, University of Angers, 2009.
 - [25] www.nist.gov/itl/iad/ig/colorferet.cfm