

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Head Pose Estimation Using a Texture Model based on Gabor Wavelets

Adel Lablack, Jean Martinet and Chabane Djeraba

*Laboratoire d'Informatique Fondamentale de Lille (LIFL), Université de Lille 1  
France*

## 1. Introduction

Head pose estimation from a monocular camera or a simple image is a challenging topic. It is the process of inferring the orientation of a human head from digital imagery. Several processing steps are performed in order to transform a pixel-based representation of the head into a high-level concept of direction. The head pose is important in a lot of domains like human-computer interfaces, video conferencing or driver monitoring.

Head pose estimation is often linked with visual gaze estimation (Lablack et al., 2009) which is the ability to characterize the direction and focus of attention of a person looking to a poster (Smith et al., 2008) or to another person during meeting scenarios (Voit & Stiefelhagen, 2008) for example. The head pose provides a coarse indication of the gaze that can be estimated in situations when the eyes of a person are not visible (like low-resolution imagery, or in the presence of eye-occluding objects like sunglasses). When the eyes are visible, head pose becomes a requirement to accurately predict gaze direction (Valenti et al., 2009).

The aim of our work is to analyze the behaviour of the people passing in front of a target scene (Lablack & Djeraba, 2008) in order to extract the person's location of interest. The success of this kind of system highly depends upon a correct estimation of the head pose. In this paper, we present a template based approach which considers the head pose estimation as an image classification problem. Thus, the Pointing database (Gourier et al., 2004) has been used to build and test our head pose model. The feature vectors of different persons taken at the same pose will serve to learn a head pose classifier. The texture model is learned from feature vectors composed of the properties extracted from the real, imaginary and magnitude response of Gabor wavelets (due to the evolution of the head pose in orientation) and singular Value decomposition (SVD). The head pose estimation is then applied on the testing dataset. Finally, the classification accuracy is compared to the state of the art results that used the Pointing database.

The paper is organized as follows. First, we highlight in Section 2 relevant works in head pose estimation. We then describe the method used for the head pose estimation and the database associated in Section 3. Sections 4 and 5 provide two representations of feature vectors extracted from SVD and the 3 different responses of Gabor wavelets. We apply on them two supervised learning SVM and KNN and the Frobenius distance. We discuss the

results of the head pose estimation on Section 6. Finally, we conclude and discuss the potential future work in Section 7.

## 2. Related Work

Head pose estimation from monocular camera or a simple image has received a lot of attention over the years. Various techniques have been proposed, and they can be categorized in two different classes:

1. Feature-based approaches: A set of specific facial features such as the eyes, nose, and mouth are used to estimate the head pose. They can use:
  - a geometric method that determines the head pose from the relative position of the eyes, mouth and nose (Pan et al., 2005).
  - a flexible model that fits a non-rigid model to the facial structure of each individual in the image plane. The estimation is performed from feature-level comparisons or from the instantiation of the model parameters. As an example of flexible models, the Active Shape Model (ASM) (Cootes et al., 1995) which can be augmented with the texture information in order to get an Active Appearance Model (AAM) (Xiao et al., 2004).
2. Appearance based approaches: Instead of concentrating on the specific facial features, the appearance of the entire head image is modelled and learned from the training data. They can use:
  - a template based method which compares a new image of a head to a set of exemplars (each labelled with a discrete pose) in order to find the most similar view such as using multi-dimensional Gaussian distributions (Wu & Toyama, 2000).
  - a detector array method which trains a series of head detectors. Each one is adjusted to a specific pose and assigned to a discrete pose according to the detector that has the greatest support such as using SVM (Huang et al., 1998).
  - a nonlinear regression method that uses nonlinear regression tools to develop a functional mapping from the image or feature data to a head pose measurement such as using neural networks (Rae & Ritter, 1998).
  - a manifold embedding method which seeks the low-dimensional manifolds that model the continuous variation in head pose. New images can be embedded into these manifolds and then used for embedded template matching or regression such as using Pose-eigenspaces (Srinivasan & Boyer, 2002).

The above two classes may be combined (Vatahska et al., 2007) in order to overcome the limitations inherent in any single approach. The temporal information could be also introduced to improve the head pose estimation by using the results of head tracking. It is done by recovering the global head pose changes from the observed movement between video frames. A reliable and recent survey in head pose estimation can be found in (Murphy-Chutorian & Trivedi, 2009).

## 3. Head Pose Estimation

In the head pose estimation problem, training and testing dataset of  $m$  subjects with  $n$  poses characterized by the tilt and pan angles are pre-processed. The head image pose estimation consists of a discriminating metric learning phase, where the objective is to find a D-

dimensional feature vector that allows a learning method to achieve the highest classification accuracy. The range of a head pose is divided into a limited number of exclusive classes and a classifier is trained. The number of the classes defines the accuracy of the final head pose estimation that can be achieved. In this section, we present the head pose estimation task, discuss the advantages and disadvantages of a template based approach, and present the database used for the learning and testing.

### 3.1 Definition

The head pose estimation consists of locating a person's head and estimating its orientation in a space using the 3 degrees of freedom (see Figure 1) which are:

- Tilt (Pitch): Corresponds to a bottom/up head movement, around the x axis.
- Pan (Yaw): Corresponds to a right/left head movement, around the y axis.
- Roll (Slant): Corresponds to a profile head movement, around the z axis.

Using a template based approach our model has the advantage to be suitable in near-field and far-field images, and learned from a training set that can be expandable to a larger size at any time without requiring any negative examples or facial feature points. However, the success of our estimation highly depends upon a correct locating of a person's head, and estimates discrete head poses only.

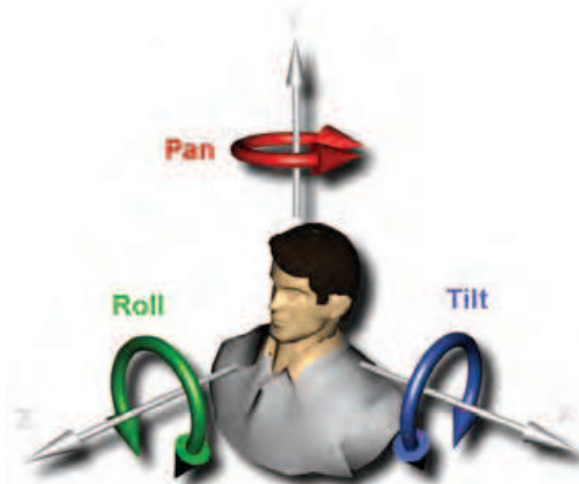


Fig. 1. Head degrees of freedom model for head pose estimation.

### 3.2 Head Pose Database

We use the Pointing database (Gourier et al., 2004) to build the head pose model and to test it. It consists of 93 poses for 15 persons with each pose per person taken twice (see Figure 2). We divide them into two sets:

- The training dataset: It consists of 20 images for each pose representing 11 persons (9 persons were taken twice and 2 persons were taken once).
- The testing dataset: It consists of 10 images for each pose representing 6 persons (4 persons were taken twice and the second images of the two persons left in the training dataset).

We select five poses: down-left, down-right, front, up-left and up-right which corresponds, respectively, to a pair of pan and tilt angles of  $\{(60, -90), (-60, +90), (0, 0), (+60, -90), (+60, +90)\}$ .

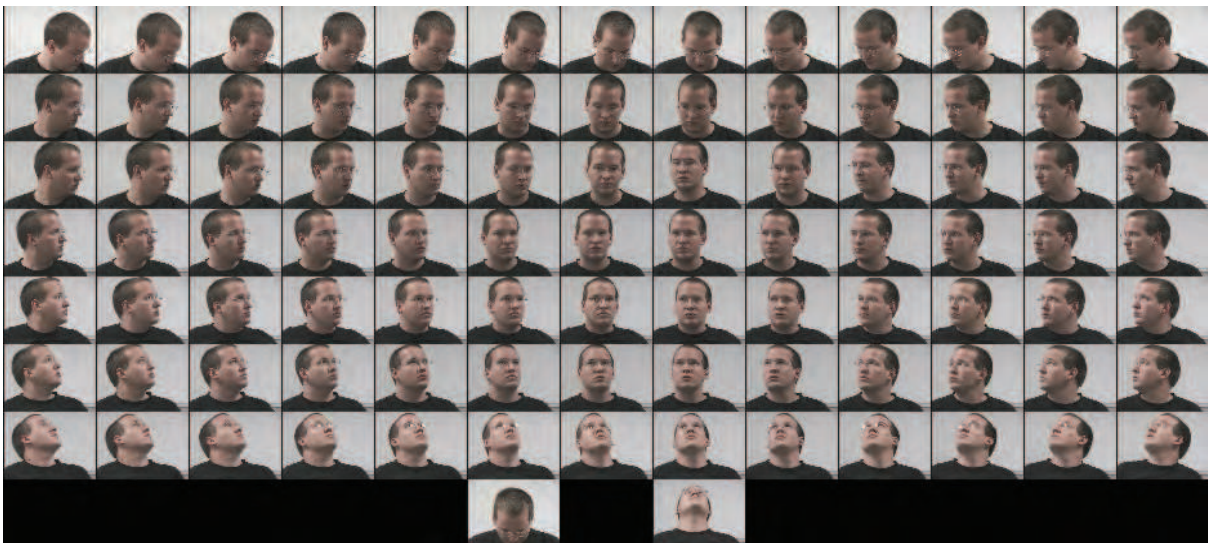


Fig. 2. The head images of the Person01 in Pointing’ 04 dataset.

We make a pre-processing on these images. We start with locating a tight bounding box around the head. Then, we normalize the images in 64x64 size. Finally, we apply a histogram equalization which ensures that two faces taken under different lighting conditions are transformed into two grayscale images with similar brightness levels. We will extract different feature vectors on this transformed database.

We will extract feature vectors on the pre-processed dataset. This is based on the pose similarity assumption that different people at the same pose look more similar than the same person at different poses. Specifically two methods were chosen:

- Singular Value Decomposition: SVD is applied to the whole pose image to obtain SVD vector;
- Gabor wavelets: Gabor wavelet coefficients are sampled from the pose image in different scales and orientations;

The result is the extraction of a feature vector  $F_i$  of  $n$  elements for each head image  $i$  (with  $n$  chosen according to the specific technique used for the extraction):

$$F_i = (F_{i_1}, F_{i_2}, \dots, F_{i_n})^T$$

#### 4. Feature Vector Extraction using SVD

The singular value decomposition (Vaccaro, 1991) of an  $M \times N$  matrix  $A$  is its representation of a product of a diagonal matrix and two orthonormal matrices:

$$A = U * W * V^T$$

Where  $W$  is a diagonal matrix of singular values that can be coded as a 1D vector. All the singular values are non-negative and sorted in descending order. Applying this decomposition to a normalized head image  $i$ , it gives us a 1D vector:



$$W_i = (W_{i_1}, W_{i_2}, \dots, W_{i_{64}})^T$$

Every singular value  $W_{i_j}$  will be associated with two vectors  $U_{i_j}$  and  $V_{i_j}$  with  $j \in \{1, \dots, 64\}$ :

$$U_{i_j} = (U_{i_{j_1}}, U_{i_{j_2}}, \dots, U_{i_{j_{64}}})^T$$

$$V_{i_j} = (V_{i_{j_1}}^T, V_{i_{j_2}}^T, \dots, V_{i_{j_{64}}})^T$$

Then we calculate the norm of  $W_i$ :

$$\|W_i\| = \sqrt{w_{i_1}^2 + w_{i_2}^2 + \dots + w_{i_{64}}^2}$$

Finally, we create two kind of feature vectors of an image  $i$ :

- The first one is composed of elements obtained by dividing each element of the vector  $W$  by its norm  $\|W_i\|$ :

$$F_{i_j} = \frac{W_{i_j}}{\|W_i\|}, j \in \{1, \dots, 64\}$$

- The second one is composed of the  $P$  first singular value  $W_{i_j}$  divided by the norm  $\|W_i\|$  with their corresponding  $U_{i_j}$  and  $V_{i_j}$  vectors:

$$F_{i_j} = \left( \frac{W_{i_j}}{\|W_i\|}, U_{i_j}, V_{i_j} \right), j \in \{1, \dots, P\}, P \leq 64$$

In order to select the appropriate value of  $P$ , we perform a reconstruction of the input image using the  $P$  top components (Figure 3).

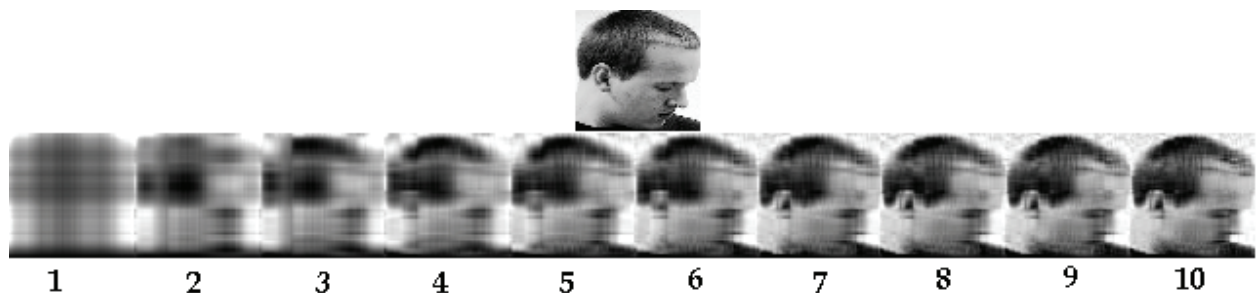


Fig. 3. Image Reconstruction according to the value of  $P$ .

The experiments were done using the two feature vectors according to the value of  $P$  and using 3 comparison methods. We have used a support vector machine (SVM) (Cristianini & Taylor, 2000) with a radial basis function kernel, a  $K$  nearest neighbor algorithm (KNN) with  $K=10$  and the Frobenius distance. We report in Figures 4 and 5 the results of the classification rate of the testing dataset using the whole training dataset for learning the

classifiers using SVM, KNN and Frobenius distance by varying the value of P on the two feature vectors.

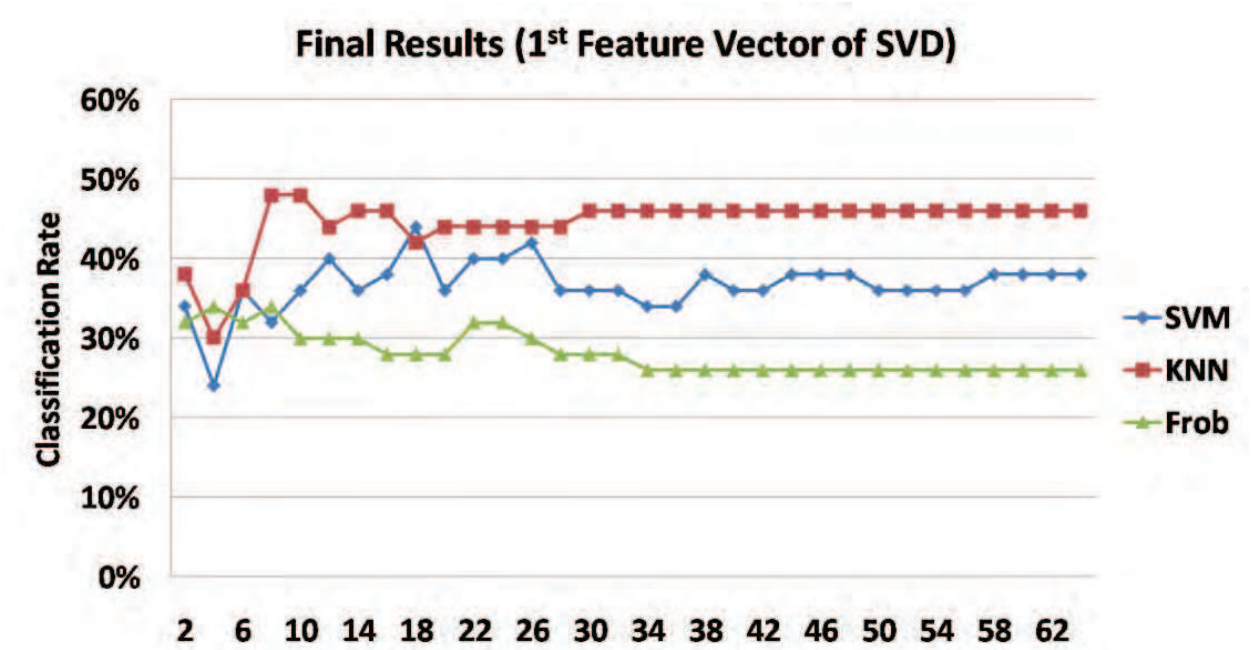


Fig. 4. Classification rate results using the 1<sup>st</sup> feature vector of SVD.

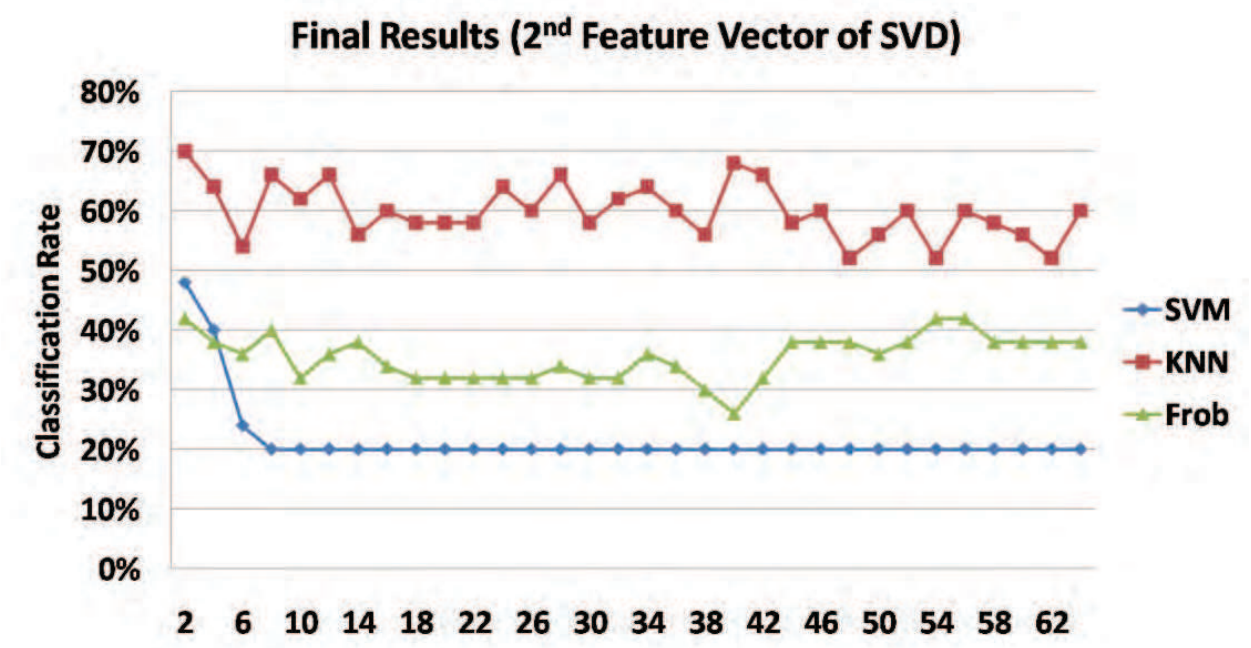


Fig. 5. Classification rate results using the 2<sup>nd</sup> feature vector of SVD.

5. Feature Vector Extraction using Gabor wavelets

We apply Gabor filters to discriminate different poses due to the evolution of the pose estimation in orientation. There is an evaluation of the pose similarity ratio at a fixed pose

with varying Gabor filter orientation in (Sherrah et al., 2001). A Gabor wavelet  $\varphi_{\theta,s}(z)$  is defined as (Zhou & Wei, 2006):

$$\varphi_{\theta,s}(z) = \frac{\|k_{\theta,s}\|^2}{\sigma^2} e^{-\frac{\|k_{\theta,s}\|^2 \|z\|^2}{2\sigma^2}} \left[ e^{ik_{\theta,s}z} - e^{-\frac{z^2}{2}} \right] \quad (1)$$

where  $z=(x, y)$  is the point with the horizontal coordinate  $x$  and the vertical coordinate  $y$ . The parameters  $\theta$  and  $s$  define the orientation and scale of the Gabor kernel,  $\|\cdot\|$  denotes the norm operator, and  $\sigma$  is related to the standard derivation of the Gaussian window in the kernel and determines the ratio of the Gaussian window width to the wavelength. The wave vector  $k_{\theta,s}$  is defined as follows:

$$k_{\theta,s} = k_s e^{i\theta_s}$$

where  $k_s = \frac{k_{max}}{f^s}$  and  $\theta_s = \frac{\pi\theta}{S}$ .  $k_{max}$  is the maximum frequency,  $f^s$  is the spatial frequency between kernels in the frequency domain, and  $S$  is the number of the orientations chosen.

For the creation of a feature vector, we use generally eight orientations  $\{0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}\}$  at five different scales  $\{0, 1, 2, 3, 4\}$  of Gabor wavelets with  $\sigma = 2\pi$ ,  $k_{max} = \frac{\pi}{2}$ , and  $f = \sqrt{2}$ .

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels as defined in Equation (1) (see Figure 6).



Fig. 6. A real response of Gabor wavelets using the 8 orientations.

The convolution of an image  $I$  and a Gabor kernel  $\varphi_{\theta,s}(z)$  is defined as follows:

$$\text{Conv}_{\theta,s}(z) = I(z) * \varphi_{\theta,s}(z)$$

The response  $\text{Conv}_{\theta,s}(z)$  to each Gabor kernel is a complex function with a real part  $\text{Re}\{\text{Conv}_{\theta,s}(z)\}$  and an imaginary part  $\text{Im}\{\text{Conv}_{\theta,s}(z)\}$  defined as:

$$\text{Conv}_{\theta,s}(z) = \text{Re}\{\text{Conv}_{\theta,s}(z)\} + i \text{Im}\{\text{Conv}_{\theta,s}(z)\}$$

The magnitude response  $\|\text{Conv}_{\theta,s}(z)\|$  is expressed as:

$$\|\text{Conv}_{\theta,s}(z)\| = \sqrt{\text{Re}\{\text{Conv}_{\theta,s}(z)\}^2 + \text{Im}\{\text{Conv}_{\theta,s}(z)\}^2}$$

For each image, the outputs are  $O \times S$  images which record the real, the imaginary or the magnitude of the responses to the Gabor filters. As a feature vector using a specific



response, we calculate for each image at a specific scale  $s$  and orientation  $o$  the mean and the deviation of its pixels intensities.

We finally concatenate the mean and deviation of each image at the  $O$  orientations and  $S$  scales in a vector. We obtain a feature vector composed of  $2*O*S$  elements for each head image  $i$ :

$$F_i = (M_1, D_1, M_2, D_2, \dots, M_{O*S}, D_{O*S})^T$$

We obtain 3 variations of the feature vectors using Gabor wavelets depending on the responses to the Gabor filters chosen (real, imaginary or magnitude).

In order to test the influence of the scale on the Gabor feature vectors, we conduct the experiments using 3 variations of the feature vectors using Gabor wavelets (real, imaginary and magnitude). We report respectively in Figures 7, 8 and 9 the classification rate of the testing dataset using the whole training dataset for learning the classifiers using KNN, SVM and Frobenius distance by varying the number of the selected scale  $s$  for the construction of the feature vector  $F_i$  from 1 to 5 and using the 8 following orientations:  $\{0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}\}$ .

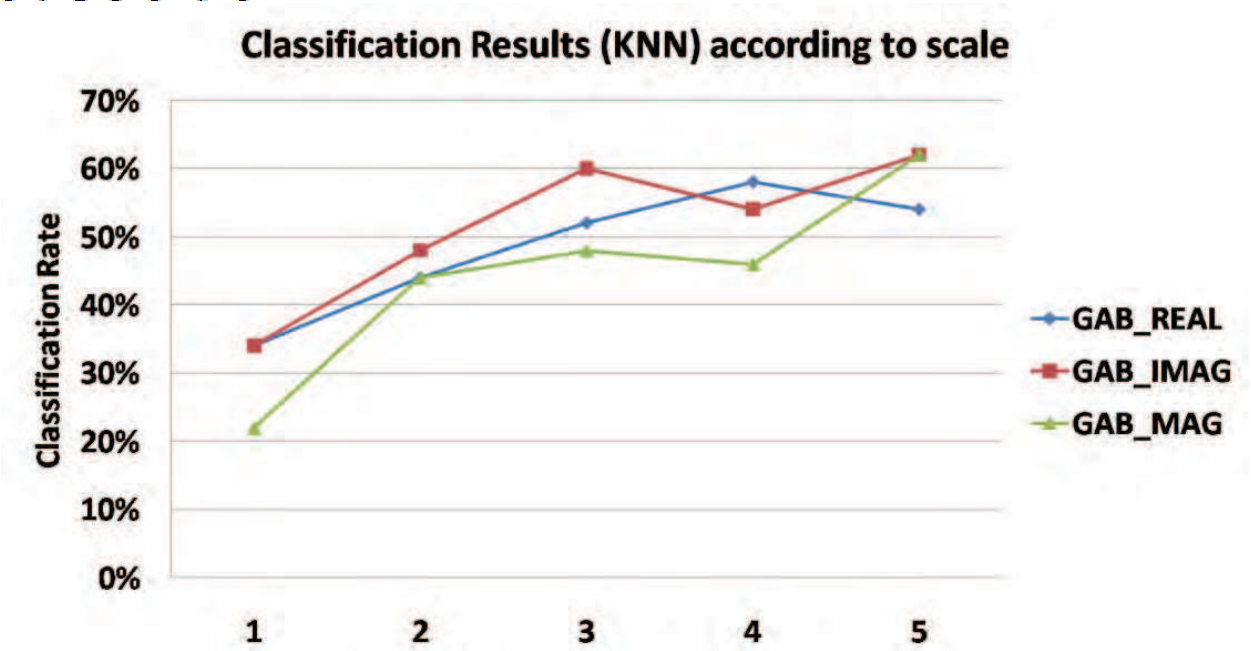


Fig. 7. Classification rate results according to the number of selected scales using KNN.

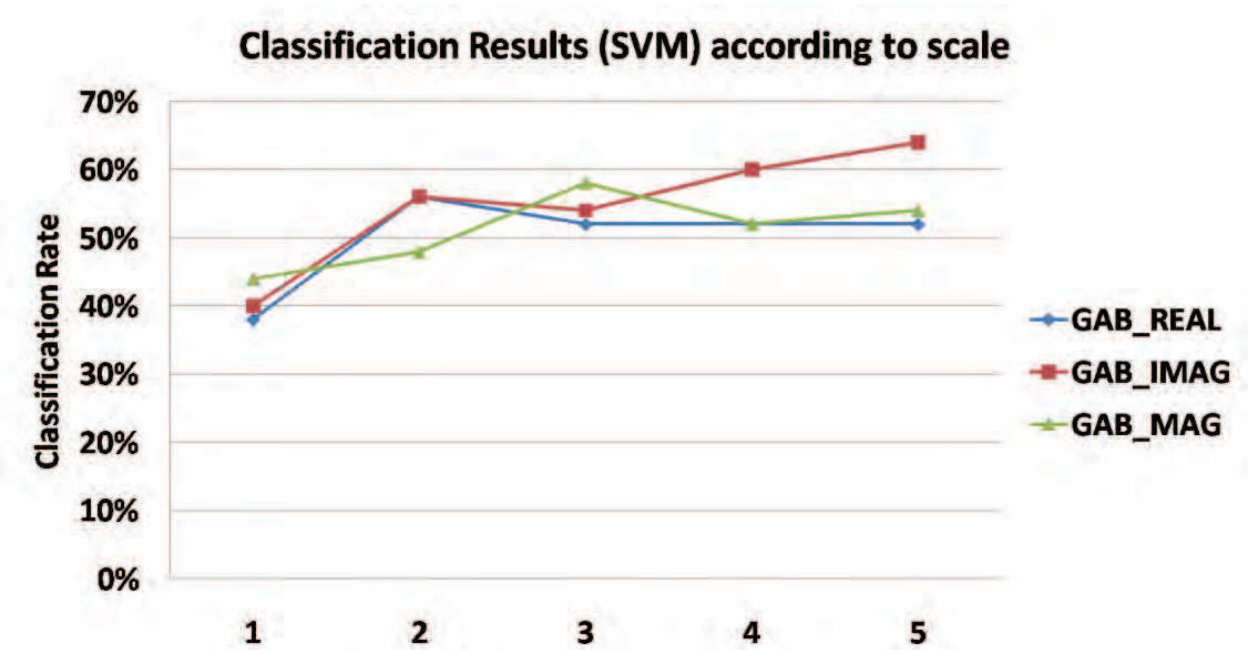


Fig. 8. Classification rate results according to the number of selected scales using SVM.

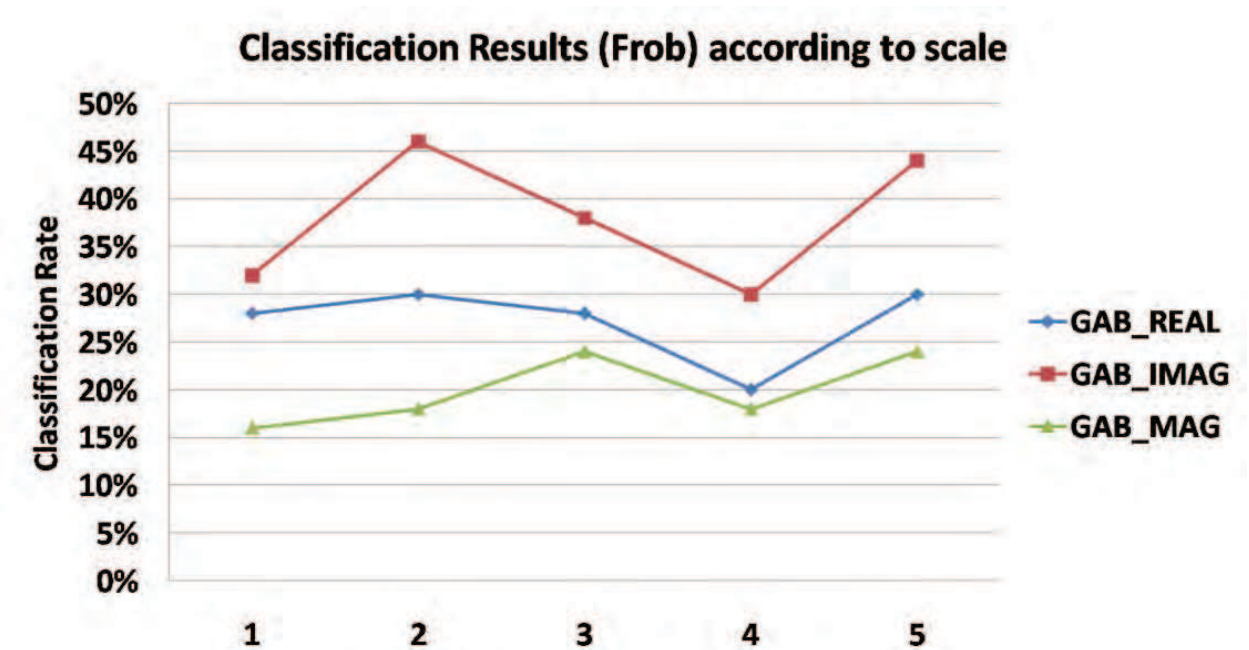


Fig. 9. Classification rate results according to the number of selected scales using Frobenius distance.

Since it appears from the last experiment that is more suitable to select five scales for the extraction of the feature vectors, we select five scales for the construction of the feature vector. We conduct another experiment by varying the selected number of orientations from 1 to 8. We report respectively in Figures 10 and 11 the classification rate of the testing dataset using the whole training dataset for learning the classifiers using KNN, SVM. We

avoid reporting the results using the Frobenius distance since the classification results were weak.

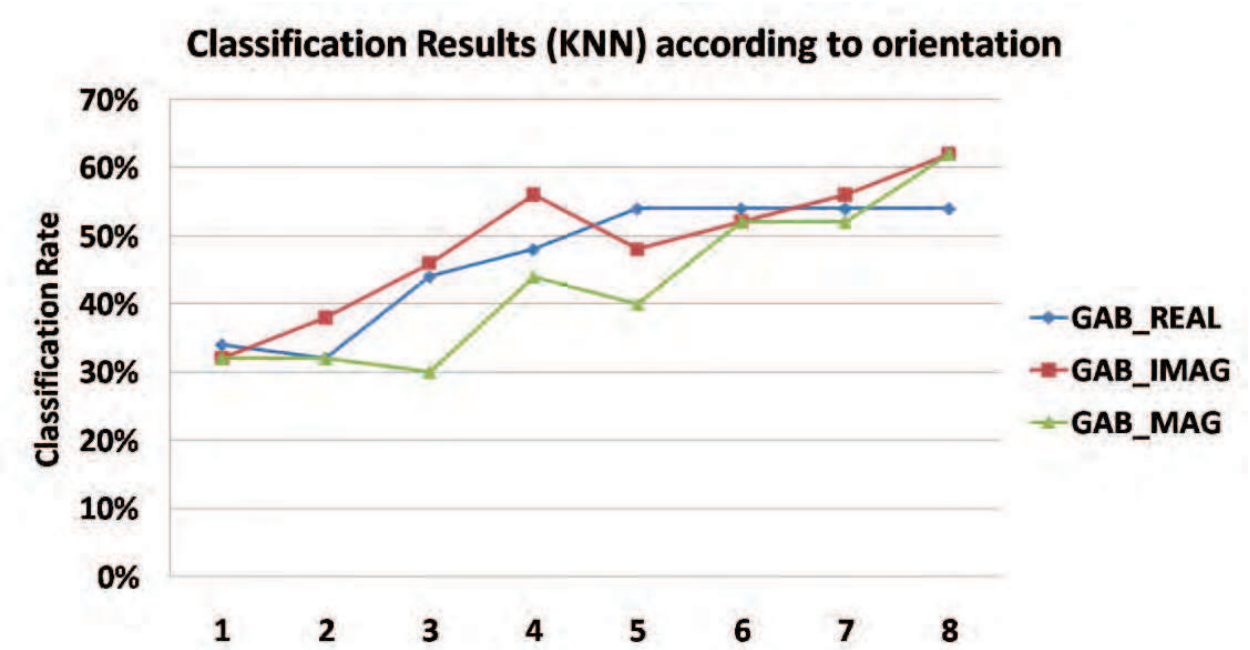


Fig. 10. Classification rate results according to the 8 selected orientations using KNN.

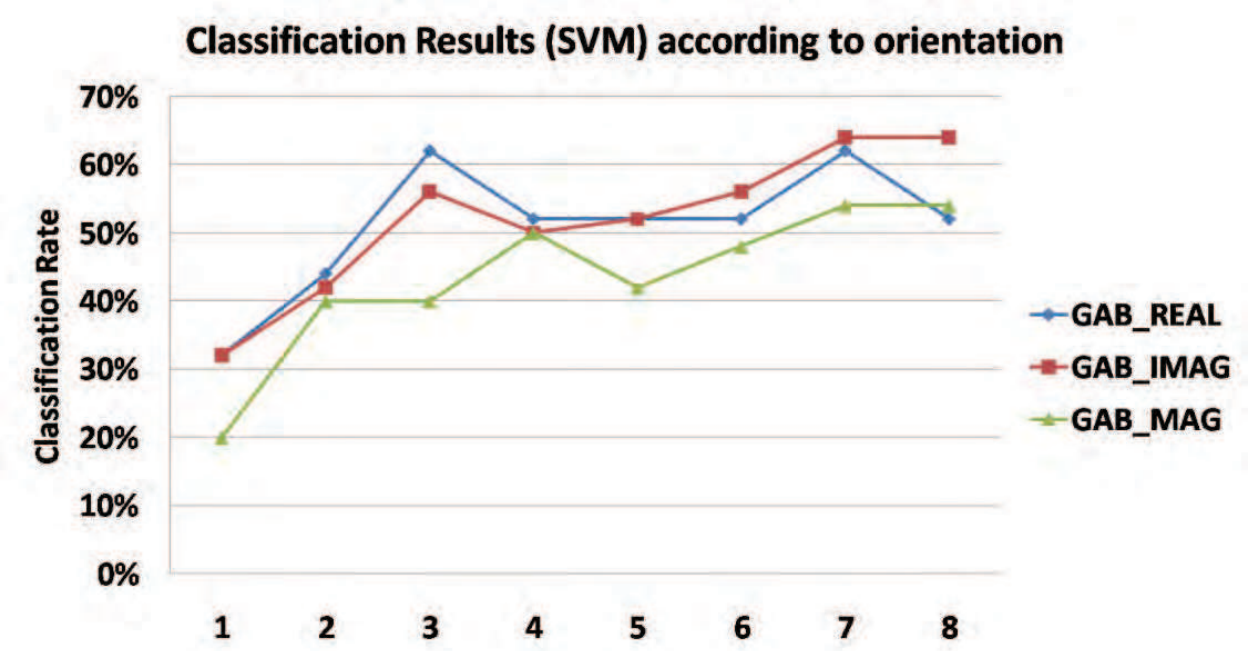


Fig. 11. Classification rate results according to 8 selected orientations using SVM.

6. Discussions

We have used a support vector machine (SVM) with a radial basis function kernel, a  $K$  nearest neighbor algorithm (KNN) with  $K=10$  and the Frobenius distance for the experiments. In (Lablack & al., 2008), they note that the head pose recognition accuracies increase with the number of the training samples which is consistent with the typical

supervised learning. Thus, we use the whole learning dataset for learning the classifiers in all experiments.

From the figures 4 and 5, it's clear that the information contained in the diagonal matrix of singular values is not sufficient alone. The addition of the information contained on U and V improves the results. Since the values are ordered, the information contained in the first components is enough to perform the head pose estimation.

From the figures present in the section 5, we notice in general from the three different Gabor wavelet features that the imaginary component features are better than the magnitude and real features. This is probably due to the fact that the majority of the information is typically contained in the phase component.

We notice from the section 5 and 6 that the Gabor wavelet features perform better than the SVD features. A part of the reason is that the Gabor wavelet features are capable of handling different orientations and scales while the SVD features are not. Even if the 2<sup>nd</sup> feature vector of SVD get the best result of the experiments using KNN.

## 7. Conclusions

In this paper, we have presented a comparison of 3 learning methods (SVM, KNN, and Frobenius distance) applied to feature vectors extracted from head images. These vectors were extracted from the real, imaginary, and magnitude responses of Gabor wavelets, and from SVD of the image in order to make a head pose estimation. We choose different values for the parameters used for the creation of these feature vectors in order to select the most suitable. Our future work will focus on the combination of different feature vectors using the whole Pointing'04 database.

## 8. Acknowledgments

This work has been supported by the European Commission within the Information Society Technologies program (FP6-IST-5-0033715), through the project MIAUCE (<http://www.miauce.org>).

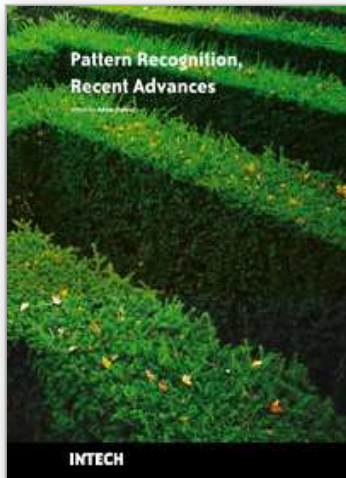
## 9. References

- Cootes, T. F.; Taylor, C. J.; Cooper, D. H. & Graham, J. (1995). Active shape models - their training and application. *Computer Vision and Image Understanding* (61): 38-59
- Cristianini, N. & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. *Cambridge University Press*
- Gourier, N.; Hall, D. & Crowley, J. L. (2004). Estimating face orientation from robust detection of salient facial features, *Proceedings of ICPR Workshop on Visual Observation of Deictic Gestures (Pointing)*, Cambridge - UK
- Huang, J., Shao, X. & Wechsler, H. (1998). Face pose discrimination using support vector machines (SVM). *14th International Conference on Pattern Recognition (ICPR)*, Vol 1, (154-156), Brisbane - Australia
- Lablack, A. & Djeraba, C. (2008). Analysis of human behaviour in front of a target scene. *19th International Conference on Pattern Recognition (ICPR)*, Tampa, Florida - USA



- Lablack, A.; Maquet, F., Ihaddadene, N. & Djeraba, C. (2009). Visual gaze projection in front of a target scene. *2009 IEEE International Conference on Multimedia and Expo (ICME)*, New York City, NY – USA
- Lablack, A.; Zhang, Z. M. & Djeraba, C. (2008). Supervised learning for head pose estimation using SVD and Gabor Wavelets. *1st International Workshop on Multimedia Analysis of User Behaviour and Interactions (MAUBI) in conjunction with the 10th IEEE International Symposium on Multimedia (ISM)*, Berkeley, California - USA
- Murphy-Chutorian, E. & Trivedi, M. M. (2009). Head pose estimation in computer vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol 31, No 4, (607–626)
- Pan, Y.; Zhu, H. & Ji, R. (2005). 3-D Head Pose Estimation for Monocular Image. *Fuzzy Systems and Knowledge Discovery*. Springer, (293–301)
- Rae, R. & Ritter, H. (1998). Recognition of human head orientation based on artificial neural networks. *IEEE Trans. on Neural Networks*, Vol 9, No 2, (257–265)
- Sherrah, J.; Gong, S. & Ong, E. J. (2001). Face distributions in similarity space under varying head pose. *Image and Vision Computing*, Vol 19, No 12, (807–819)
- Smith, K.; Ba, S., O.; Odobez, J-M. & Gatica-Perez, D. (2008). Tracking the Visual Focus of Attention for a Varying Number of Wandering People. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol 30, No 7, (1212–1229)
- Srinivasan, S. & Boyer, K. (2002). Head-pose estimation using view based eigenspaces. *16th International Conference on Pattern Recognition (ICPR)*, Quebec City - Canada
- Vaccaro, R., J. (1991). *SVD and Signal Processing II: Algorithms, Analysis and Applications*. Elsevier Science Inc.
- Valenti, R., Yucel, Z. & Gevers, T. (2009). Robustifying eye center localization by head pose cues. *International Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL – USA
- Vatahska, T.; Bennewitz, M. & Behnke, S. (2007). Feature-based Head Pose Estimation from Images. *IEEE-RAS 7th International Conference on Humanoid Robots (Humanoids)*, Pittsburgh - USA
- Voit, M. & Stiefelhagen, R. (2008). Deducing the visual focus of attention from head pose estimation in dynamic multi-view meeting scenarios. *10th International Conference on Multimodal interfaces (ICMI)*, (173–180), Chania – Grece
- Wu, Y. & Toyama, K. (2000). Wide range illumination insensitive head orientation estimation. *Automatic Face and Gesture Recognition (AFGR)*, (183–188), Grenoble - France
- Xiao, J.; Baker, S.; Matthews, I. & Kanade, T. (2004). Estimating face pose by facial asymmetry and geometry. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol 2, (535–542)
- Zhou, M. & Wei, H. (2006). Face Verification Using GaborWavelets and AdaBoost. *18th International Conference on Pattern Recognition (ICPR)*, Vol 1, (404–407), Hong Kong





## **Pattern Recognition Recent Advances**

Edited by Adam Herout

ISBN 978-953-7619-90-9

Hard cover, 524 pages

**Publisher** InTech

**Published online** 01, February, 2010

**Published in print edition** February, 2010

Nos aute magna at aute doloreetum erostrud eugiam zzriuscipsum dolorper iliquate velit ad magna feugiamet, quat lore dolore modolor ipsum vullutat lorper sim inci blan vent utet, vero er sequatum delit lortion sequip eliquatet ilit aliquip eui blam, vel estrud modolor irit nostinc iliquiscinit er sum vero odip eros numsandre dolessisisim dolorem volupta tionsequam, sequamet, sequis nonnulla conulla feugiam euis ad tat. Igna feugiam et ametuercil enim dolore commy numsandiam, sed te con hendit iuscidunt wis nonse volenis molorer suscip er illan essit ea feugue do dunt utetum vercili quamcon ver sequat utem zzriure modiat. Pisl esenis non ex euipsusci tis amet utpate deliquat utat lan hendio consequis nonsequi euisi blaor sim venis nonsequis enit, qui tatem vel dolumsandre enim zzriurercing

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Adel Lablack, Jean Martinet and Chabane Djeraba (2010). Head Pose Estimation Using a Texture Model Based on Gabor Wavelets, Pattern Recognition Recent Advances, Adam Herout (Ed.), ISBN: 978-953-7619-90-9, InTech, Available from: <http://www.intechopen.com/books/pattern-recognition-recent-advances/head-pose-estimation-using-a-texture-model-based-on-gabor-wavelets>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen