MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11-13, 2002, Nara-ken New Public Hall, Nara, Japan

6—2

Invariant Gabor Features for Face Evidence Extraction

J.-K. Kamarainen*, V. Kyrki*, H. Kälviäinen*
Laboratory of Information Processing
Lappeenranta University of Technology

M. Hamouz†, J. Kittler†
Centre for Vision, Speech, and Signal Processing
University of Surrey

Abstract

Invariant feature extraction is one of the most difficult problems in machine vision. Human face detection and recognition have recently become an important application area of computer vision. In this paper, a novel illumination, translation, rotation, and scale invariant feature extraction method based on Gabor filtering is introduced. The proposed method is successfully applied to invariant detection of facial features. In addition to face detection, the proposed theories and methods can be applied to a wide variety of object detection problems.

1 Introduction

Face recognition is an important issue in developing of security applications for personal identification. Many approaches have been studied for face recognition (e.g. [4, 9, 14, 15]). In practical applications, face detection/localization is often a prior step before face recognition and various methods have been proposed (e.g. [3, 8, 11, 15]).

Recently, authors (Hamouz and Kittler) have proposed a complete framework for more general problem of object detection [2, 8] based on detection and combination of discriminative regions. If a scale, rotation, and translation invariant detection of discriminative regions can be performed and relationships between different classes of discriminative regions of object is known, a detection of an object can be carried out. The approach can be applied also to the detection of faces in images [2, 8]. The success and computational performance of the framework depend on a successful selection and efficient extraction of discriminative regions, e.g., face evidences.

In this study, responses of Gabor filters are used to extract the face evidences from gray-scale images. The Gabor filter responses can be used in a translation, scale, and orientation invariant manner and they are very robust to small changes of illumination and in the presence of noise [5, 6, 7]. The accuracy of the method is evaluated with experiments on real face images. The proposed method performs well in the task and can be integrated into framework described in [2, 8].

*Address: P.O. Box 20, FIN-53851 Lappeenranta, Finland. {Joni.Kamarainen,Ville.Kyrki,Heikki.Kalviainen}@lut.fi.

†Guildford GU2 7XH, United Kingdom. {m.hamouz,j.kittler}@eim.surrey.ac.uk.

2 Face detection by evidence combination

A complete framework for invariant object detection based on local features and affine correspondences has been proposed by Hamouz et al. [2] and Matas et al. [8]. In this paper, the detected objects are frontal human faces and discriminative regions are salient subparts, such as nostrils and eyes. Other similar detection approaches, classified as bottom-up feature-based face detection, have been surveyed in [15]. The proposed framework can be used as a front-end at face authentication systems as shown in Fig. 1.

The framework architecture allows a separate development and evaluation of different sub-system solutions. In this study the focus is on the first phase of the face detection system, the evidence extraction. The evidence extraction is a major concern for the success of the sub-system in Fig. 1, as it should reliably provide the relevant information for the face validation stage. In particular, 10 face evidences are used to represent discriminative regions: outer eye corners (LEO,REO), eye centers (LEC,REC), inner eye corners (LEI,REI), nostrils (NL,NR), and mouth corners (ML,MR). The problem of invariant detection of the selected features is illustrated in Fig. 2.

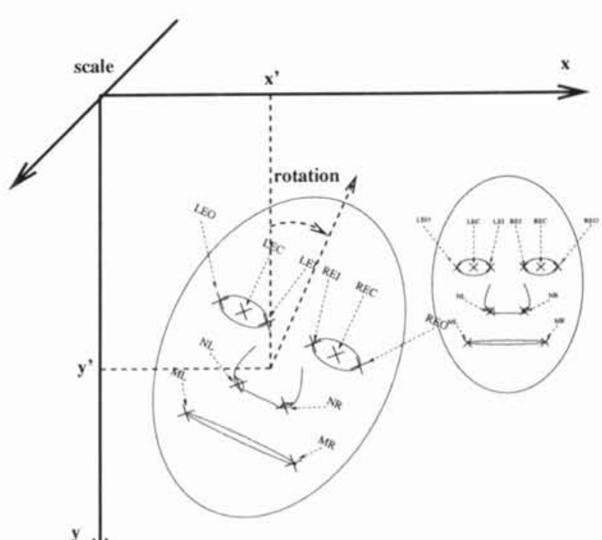


Figure 2: Invariant extraction of face evidences.

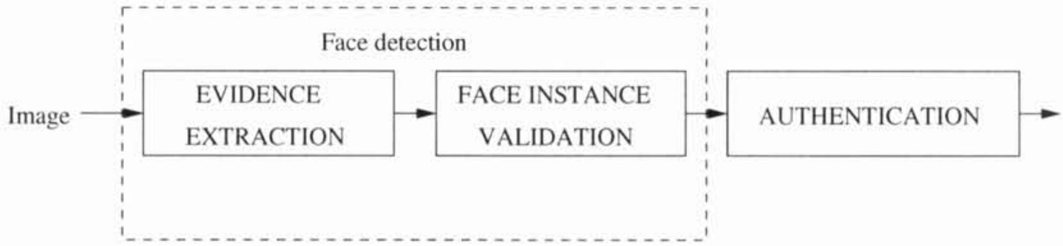


Figure 1: Authentication system.

3 Gabor features

3.1 Extraction

Gabor features have been successfully used in face detection and recognition (e.g., [1, 12, 14]). However, often only the tolerance induced by the smooth behavior of Gabor filter responses is used as the basis for invariance. It is true that a small tolerance is natural phenomenon of certain Gabor features [13], but also an arbitrary level of invariance for translation, scaling, and rotation (similarity transforms) can be achieved utilizing Gabor filter responses. Realizing invariance with Gabor filters is considered next.

A 2-dimensional Gabor filter can be defined as

$$\begin{aligned} \psi(x,y;f,\theta) =& e^{-\left(\frac{f^2}{\gamma^2}x'^2+\frac{f^2}{\eta^2}y'^2\right)}e^{j2\pi f x'}, \\ x' =& x\cos\theta + y\sin\theta, \\ y' =& -x\sin\theta + y\cos\theta, \end{aligned} \tag{1}$$

where f is the frequency, θ orientation, and γ and η control the frequency and orientation bandwidths. The normalized response of the Gabor filter, $\psi(x,y)$, for image $\xi(x,y)$ [5]

$$r_\xi(x,y;f,\theta) = \frac{f^2}{\pi\gamma\eta}\psi(x,y;f,\theta) * \xi(x,y) \tag{2}$$

can be used to construct a Gabor feature at any location $(x,y) = (x_0,y_0)$. If several frequencies and orientations of Gabor filter are used, a feature matrix of filter responses at a single point can be constructed as

$$\begin{bmatrix} r_\xi(x_0,y_0;f_0,\theta_0) & \ldots & r_\xi(x_0,y_0;f_0,\theta_{N-1}) \\ \vdots & \ddots & \vdots \\ r_\xi(x_0,y_0;f_{M-1},\theta_0) & \ldots & r_\xi(x_0,y_0;f_{M-1},\theta_{N-1}) \end{bmatrix}. \tag{3}$$

The orientation invariant similarity measure can be introduced utilizing horizontal circular shifts of the feature matrix (3) [7, 12].

The scale invariance can be realized by noticing that when the frequency parameter is logarithmically spaced,

$$f_k = a^{-k}f_{highest},\ k = \{0,1,\ldots\},\ a > 1, \tag{4}$$

scaling corresponds to vertical shifts in the feature matrix. The filter matrix can be normalized to unity to provide illumination invariance. Often only the magnitude parts of the complex filter responses are used neglecting the phase information, but it should be noted that all theorems hold also for responses containing both the magnitude and the phase information.

3.2 Classification

In this paper, two classifiers, 1-NN and SCC, are used for classifying the feature matrices (3) represented as a single vector. Since the rotation and scale invariant comparison of the features can be established, standard distance functions can be used as similarity measures, making it possible to use elementary classification methods such as the k-nearest neighbor (k-NN) decision rule. Since every pixel in the image is an evidence conveying candidate, a confidence measure for the classifier output is necessary to eliminate false hypotheses. For the 1-NN classifier, the distance to the closest stored example is used as the confidence value. However, an assumption is made that the probabilities of the evidence classes are represented by mixtures of Gaussians in the feature space, and thus, more sophisticated methods can be applied.

A new clustering based classifier, referred to as subcluster classifier (SCC), was developed especially for this application. The training of the classifier consists of the following steps: 1) cluster sample feature vectors using the c-means algorithm, 2) for each cluster, estimate the number of vectors belonging to each class, 3) identify the clusters that consist mostly of vectors of a same class, 4) using these clusters, assign each vector to the closest cluster of the same class, and 5) calculate the means and covariances for the clusters derived in the previous step.

The resulting model can be used to classify unknown filter outputs based on their Mahalanobis distance from the respective clusters, which also provides the associated confidence measure. The parameters of the method are the number of clusters and the proportion of samples from a single class that a cluster must contain in order to be accepted at step 3.

4 Experiments

In this study, only five separate classes of face features were distinguished (see Fig. 2): C1 (LEO, REO); C2 (LEC, REC); C3 (LEI, REI); C4 (NL, NR); C5 (ML, MR). Experiments were conducted using the XM2VTS face database consisting of 600 training images and 560 test images of size 576×720 [10]. Now, the benefits of invariant measures and the effect of a classifier are examined first by evidence classification experiments. Then, the evidence extraction capability and the effect of the phase information are evaluated.

Table 1: Classification accuracies for 1-NN and SCC classifiers.

Image manipulation	None	Rotating	Rotating	Scaling	Scaling
Applied invariance	None	None	Rotation	None	Scale
1-NN	76%	54%	64%	51%	62%
SCC	74%	52%	58%	49%	61%

4.1 Evidence classification

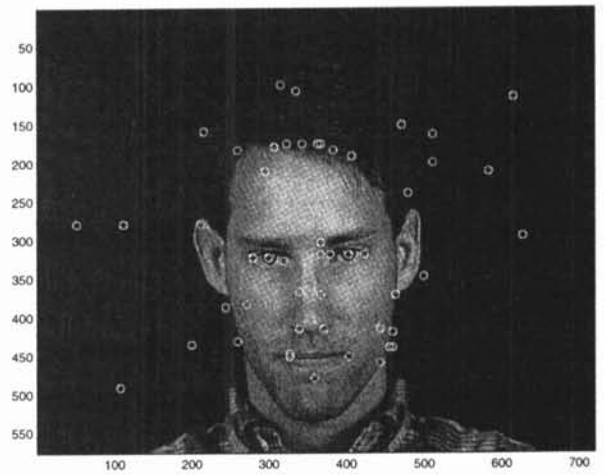
The discriminatory power and invariance capability of the proposed features were assessed in the first experiment. The feature matrix was generated at manually labeled feature locations of normalized training images. The features in the test set images were classified in three different poses, normalized pose, random rotation ($0^\circ - 45^\circ$), and random scaling ($1.0\times - 2.0\times$). A 4×4 feature matrix of filter magnitudes was used. 4 different frequencies f_0 corresponding to wavelengths of 4, 8, 16, and 32 pixels, and 4 equally spaced orientations were selected with filter bandwidth corresponding to $\gamma = 1, \eta = 1$. 1 scale shift (recognition in normal and double scale) and 3 (maximum for 4 orientations) orientation shifts were used. While 1-NN decision rule performs little better (Table 1), its computational complexity is notably higher than the proposed SCC scheme, so SCC is not an optimal classifier for the data. More importantly, the results show that compared to natural tolerance (second and fourth column of Table 1), applied invariance (third and fifth columns) improves classification accuracy.

4.2 Evidence extraction

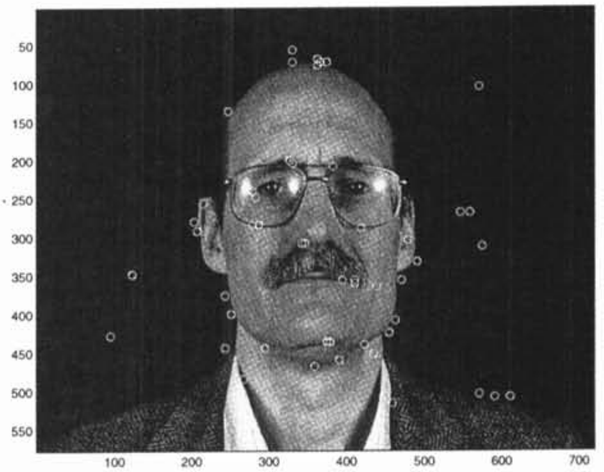
Without a proper confidence measure all points can be classified, but the points cannot be ordered to treat the most probable candidates first. It was found out that the confidence measure for 1-NN works slowly and somewhat unreliably. For that reason, only the SCC classifier was used to extract face feature evidences.

For successful face instance validation in the next processing step (see Fig. 1), at least three correct evidences must be extracted within a distance of 5 pixels. As a performance measure for the proposed method, the number of point sets (one point from each class) required for validation was used. Using only the magnitude of the filter responses, the SCC classifier succeeded in 521 images and failed in 39 (after extraction of 200 point sets). Examples of successful and unsuccessful extraction are shown in Fig. 3. It should be noted that faulty points in the low contrast background are due to illumination invariance and can be avoided by selecting some minimum contrast.

The experiment was conducted using also the phase information. With phase information the method succeeded with 542 images and failed only with 18 images. Still, the most important improvement can be seen in Fig. 4 where histograms of the number of extractions needed for face detection are shown. Clearly, the phase information improves the method significantly. For most cases, only one set of points was required to obtain the 3 necessary evidences (see Fig. 4(b)).

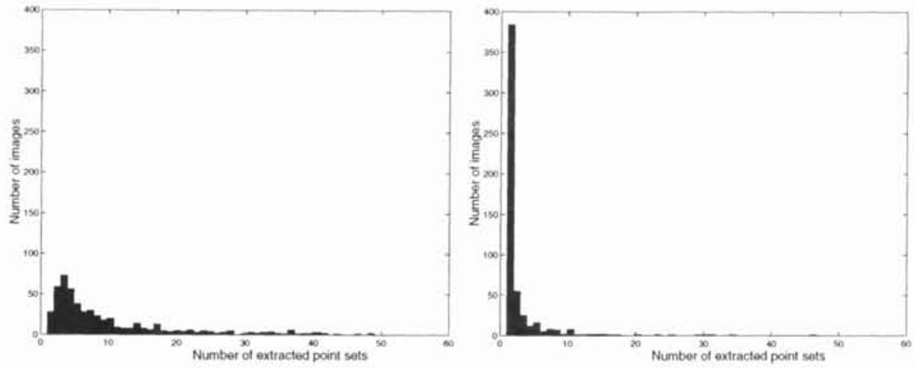


(a) Successful.



(b) Unsuccessful.

Figure 3: 10 Best evidence candidates for each evidence classes (C1-C5) extracted from images.



(a) Magnitude information.

(b) Magnitude and phase information.

Figure 4: Number of point sets extracted for successful detection.

5 Conclusions

A novel method for extracting local discriminative features was proposed. The method can be used in object detection, e.g., in face detection as demonstrated in this study. The method is based on Gabor filter features, invariant treatment of the feature matrix, and a clustering based classifier. In order to detect faces according to [2, 8] at least 3 different local features have to be correctly detected. Using the complex information, the proposed method succeeded in 97% of cases. In addition, experiments pointed out the significance of phase information for accurate evidence extraction.

Further experiments with more complex facial images are needed to stress the relevance of phase information. Two interesting problems to asses in further studies are automatic selection of evidences and synthesis of a better classifier based on studies on Gaussian mixture models. In the future the proposed method will be integrated into the system proposed in [2, 8] to create a complete face authentication application.

References

[1] J. Buhmann, J. Lange, and C. Von Der Malsburg. Distortion invariant object recognition by matching hierarchically labeled graphs. In *IJCNN: International Joint Conference on Neural Networks*, volume 1, pages 155–159, 1989.

[2] M. Hamouz, J. Kittler, J. Matas, and P. Bílek. Face detection by learned affine correspondences. In *Joint international Workshop on Syntactical & Structural & Statistical Pattern Recognition*, pages 566–575, Windsor (Ontario), Canada, August 2002.

[3] Ing-Sheen Hsieh, Kuo-Chin Fan, and Chiunhsiun Lin. A statistic approach to the detection of human faces in color nature scene. *Pattern Recognition*, 35:1583–1596, 2002.

[4] K. Jonsson, J. Kittler, Y.P. Li, and J. Matas. Support vector machines for face authentication. *Image and Vision Computing*, 20:369–375, 2002.

[5] J.-K. Kamarainen, V. Kyrki, and H. Kälviäinen. Fundamental frequency Gabor filters for object recognition. In *16th International Conference on Pattern Recognition ICPR2002*, volume 1, pages 628–631, Quebec, Canada, 2002.

[6] J.-K. Kamarainen, V. Kyrki, and H. Kälviäinen. Noise tolerant object recognition using Gabor filtering. In *Proceedings of DSP2002 14th International Conference on Digital Signal Processing*, pages 1349–1352, Santorini, Greece, 2002.

[7] V. Kyrki, J.-K. Kamarainen, and H. Kälviäinen. Invariant shape recognition using global Gabor features. In *12th Scandinavian Conference on Image Analysis*, pages 671–678, Bergen, Norway, 2001.

[8] J. Matas, P. Bílek, M. Hamouz, and J. Kittler. Discriminative regions for human face detection. In *Proceedings of ACCV2002: The 5th Asian Conference on Computer Vision*, 2002.

[9] J. Matas, K. Jonsson, and J. Kittler. Fast face localisation and verification. *Image and Vision Computing*, 17(8):575–581, June 1999.

[10] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The extended M2VTS Database. In R. Chellapa, editor, *Proceedings of Second International Conference on Audio and Video-based Biometric Person Authentication*, pages 72–77, 1999.

[11] Jeffrey Ng and Shaogang Gong. Composite support vector machines for detection of faces accross views and pose estimation. *Image and Vision Computing*, 20:359–368, 2002.

[12] Hyun Jin Park and Seung Yang. Invariant object detection based on evidence accumulation and Gabor features. *Pattern Recognition Letters*, 22:869–882, 2001.

[13] Eero P. Simoncelli, William T. Freeman, Edward H. Adelson, and David J. Heeger. Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, 38(2):587–607, March 1992.

[14] Laurentz Wiskott, Jean-Marc Fellous, Norbert Krüger, and Christoph von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.

[15] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.