1

Feature Extraction and Representation for Face Recognition

[1]M. Saquib Sarfraz, [2]Olaf Hellwich and [3]Zahid Riaz
[1]*Computer Vision Research Group, Department of Electrical Engineering*
COMSATS Institute of Information Technology, Lahore
Pakistan
[2]*Computer Vision and Remote Sensing, Berlin University of Technology*
Sekr. FR 3-1, Franklin str. 28/29, 10587, Berlin
Germany
[3]*Institute of Informatik, Technical University Munich*
Germany

1. Introduction

Over the past two decades several attempts have been made to address the problem of face recognition and a voluminous literature has been produced. Current face recognition systems are able to perform very well in controlled environments e.g. frontal face recognition, where face images are acquired under frontal pose with strict constraints as defined in related face recognition standards. However, in unconstrained situations where a face may be captured in outdoor environments, under arbitrary illumination and large pose variations these systems fail to work. With the current focus of research to deal with these problems, much attention has been devoted in the facial feature extraction stage. Facial feature extraction is the most important step in face recognition. Several studies have been made to answer the questions like what features to use, how to describe them and several feature extraction techniques have been proposed. While many comprehensive literature reviews exist for face recognition a complete reference for different feature extraction techniques and their advantages/disadvantages with regards to a typical face recognition task in unconstrained scenarios is much needed.

In this chapter we present a comprehensive review of the most relevant feature extraction techniques used in 2D face recognition and introduce a new feature extraction technique termed as Face-GLOH-signature to be used in face recognition for the first time (Sarfraz and Hellwich, 2008), which has a number of advantages over the commonly used feature descriptions in the context of unconstrained face recognition.

The goal of feature extraction is to find a specific representation of the data that can highlight relevant information. This representation can be found by maximizing a criterion or can be a pre-defined representation. Usually, a face image is represented by a high dimensional vector containing pixel values (holistic representation) or a set of vectors where each vector summarizes the underlying content of a local region by using a high level

transformation (local representation). In this chapter we made distinction in the holistic and local feature extraction and differentiate them qualitatively as opposed to quantitatively. It is argued that a global feature representation based on local feature analysis should be preferred over a bag-of-feature approach. The problems in current feature extraction techniques and their reliance on a strict alignment is discussed. Finally we introduce to use face-GLOH signatures that are invariant with respect to scale, translation and rotation and therefore do not require properly aligned images. The resulting dimensionality of the vector is also low as compared to other commonly used local features such as Gabor, Local Binary Pattern Histogram 'LBP' etc. and therefore learning based methods can also benefit from it.
A performance comparison of face-GLOH-Signature with different feature extraction techniques in a typical face recognition task is presented using FERET database. To highlight the usefulness of the proposed features in unconstrained scenarios, we study and compare the performance both under a typical template matching scheme and learning based methods (using different classifiers) with respect to the factors like, large number of subjects, large pose variations and misalignments due to detection errors. The results demonstrate the effectiveness and weakness of proposed and existing feature extraction techniques.

2. Holistic Vs Local Features-What Features to Use?

Holistic representation is the most typical to be used in face recognition. It is based on lexicographic ordering of raw pixel values to yield one vector per image. An image can now be seen as a point in a high dimensional feature space. The dimensionality corresponds directly to the size of the image in terms of pixels. Therefore, an image of size 100x100 pixels can be seen as a point in a 10,000 dimensional feature space. This large dimensionality of the problem prohibits the use of any learning to be carried out in such a high dimensional feature space. This is called the curse of dimensionality in the pattern recognition literature (Duda et al, 2001). A common way of dealing with it is to employ a dimensionality reduction technique such as Principal Component Analysis 'PCA' to pose the problem into a low-dimensional feature space such that the major modes of variation of the data are still preserved.
Local feature extraction refers to describing only a local region/part of the image by using some transformation rule or specific measurements such that the final result describes the underlying image content in a manner that should yield a unique solution whenever the same content is encountered. In doing so, however it is also required to have some degree of invariance with respect to commonly encountered variations such as translation, scale and rotations. A number of authors (Pentland et al, 1994; Cardinaux et al, 2006; Zou et al, 2007) do not differentiate the holistic and local approaches according to the very nature they are obtained, but rather use the terms in lieu of global (having one feature vector per image) and a bag-of-feature (having several feature vectors per image) respectively. Here we want to put the both terms into their right context, and hence a holistic representation can be obtained for several local regions of the image and similarly a local representation can still be obtained by concatenating several locally processed regions of the image into one global vector, see figure 1 for an illustration. An example of the first usage is local-PCA or modular- PCA (Gottumukkal and Asari, 2004; Tan and Chen, 2005), where an image is divided into several parts or regions, and each region is then described by a vector

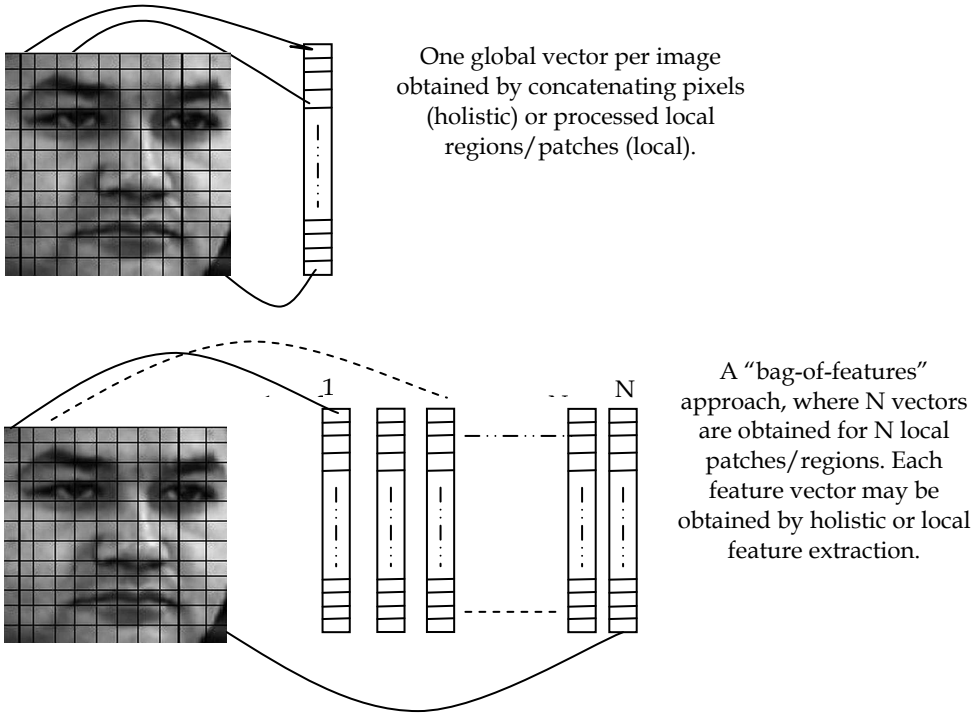

Fig. 1. Global and bag-of-feature representation for a facial image

comprising underlying raw-pixel values, PCA is then employed to reduce the dimensionality. Note that it is called local since it uses several local patches of the same image but it is still holistic in nature. An example of the second is what usually found in the literature, e.g. Gabor filtering, Discrete Cosine Transform 'DCT', Local Binary Pattern 'LBP' etc where each pixel or local region of the image is described by a vector and concatenated into a global description (Zou et al, 2007), note that they still give rise to one vector per image but they are called local in the literature because they summarize the local content of the image at a location in a way that is invariant with respect to some intrinsic image properties e.g. scale, translation and/or rotation.

Keeping in view the above discussion it is common in face recognition to either follow a global feature extraction or a bag-of-features approach. The choice, of what is optimal, depends on the final application in mind and hence is not trivial. However, there are a number of advantages and disadvantages with both the approaches. For instance, a global description is generally preferred for face recognition since it preserves the configural (i.e., the interrelations between facial parts) information of the face, which is very important for preserving the identity of the individual as have been evidenced both from psychological (Marta et al, 2006), neurobiological (Schwaninger et al, 2006; Hayward et al, 2008) and computer vision (Belhumeur et al, 1997; Chen et al, 2001) communities. On the other hand, a bag-of-features approach has been taken by a number of authors (Brunelli and Poggio, 1993; Martınez, 2002; Kanade and Yamada, 2003) and shown improved recognition results

in the presence of occlusion etc., nonetheless, in doing so, these approaches are bound to preserve the configural information of the facial parts either implicitly or explicitly by comparing only the corresponding parts in two images and hence puts a hard demand on the requirement of proper and precise alignment of facial images.
Note that while occlusion may be the one strong reason to consider a bag-of-features approach, the tendency of preserving the spatial arrangement of different facial parts (configural information) is largely compromised. As evidenced from the many studies from interdisciplinary fields that this spatial arrangement is in fact quite crucial in order to preserve the identity of an individual, we therefore, advocate the use of a global representation for a face image in this dissertation, as has also been used by many others.
One may, however, note that a global representation does not necessarily mean a holistic representation, as described before. In fact, for the automatic unconstrained face recognition, where there may be much variation in terms of scale, lighting, misalignments etc, the choice of using local feature extraction becomes imperative since holistic representation cannot generalize in these scenarios and is known to be highly affected by these in-class variations.

3. Holistic Feature Extraction

Holistic feature extraction is the most widely used feature description technique in appearance based face recognition methods. Despite its poor generalization abilities in unconstrained scenarios, it is being used for the main reason that any local extraction technique is a form of information reduction in that it typically finds a transformation that describes a large data by few numbers. Since from a strict general object recognition stand point, face is one class of objects, and thus discriminating within this class puts very high demands in finding subtle details of an image that discriminates among different faces. Therefore each pixel of an image is considered valuable information and holistic processing develops. However, a holistic-based global representation as been used classically (Turk and Pentland, 1991) cannot perform well and therefore more recently many researchers used a bag-of-features approach, where each block or image patch is described by holistic representation and the deformation of each patch is modeled for each face class (Kanade and Yamada, 2003; Lucey and Chen, 2006; Ashraf et al, 2008).

3.1 Eigenface- A global representation

Given a face image matrix F of size Y x X, a vector representation is constructed by concatenating all the columns of F to form a column vector $\vec{f}$ of dimensionality YX. Given a set of training vectors $\{\vec{f}_i\}_{i=1}^{Np}$ for all persons, a new set of mean subtracted vectors is formed using:

$$g_i = \vec{f}_i - \vec{f}_\mu, \qquad i = 1, 2, ..., N_p \tag{1}$$

The mean subtracted training set is represented as a matrix $G = [\vec{g}_1, \vec{g}_2, ..., \vec{g}_{Np}]$. The covariance matrix is then calculated using, $\sum = GG^T$. Due to the size of $\sum$, calculation of the eigenvectors of $\sum$ can be computationally infeasible. However, if the number of training vectors (N_p) is less than their dimensionality (YX), there will be only N_p-1 meaningful

eigenvectors. (Turk and Pentland, 91) exploit this fact to determine the eigenvectors using an alternative method summarized as follows. Let us denote the eigenvectors of matrix G^TG as $\vec{v}_j$ with corresponding eigenvalues Λ_j:

$$G^T G\vec{v}_j = \Lambda_j \vec{v}_j \tag{2}$$

Pre-multiplying both sides by G gives us: $GG^T G\vec{v}_j = \Lambda_j G\vec{v}_j$, Letting $\vec{e}_j = G\vec{v}$ and substituting for $\sum$ from equation 1:

$$\sum \vec{e}_j = \Lambda_j \vec{e}_j \tag{3}$$

Hence the eigenvectors of $\sum$ can be found by pre-multiplying the eigenvectors of $G^T G$ by G. To achieve dimensionality reduction, let us construct matrix $E = [\vec{e}_1, \vec{e}_1, ..., \vec{e}_D]$, containing D eigenvectors of $\sum$ with largest corresponding eigenvalues. Here, $D<N_p$, a feature vector $\vec{x}$ of dimensionality D *is* then derived from a face vector $\vec{f}$ using:

$$\vec{x} = E^T(\vec{f} - \vec{f}_\mu) \tag{4}$$

Therefore, a face vector $\vec{f}$ is decomposed into D eigenvectors, known as eigenfaces. Similarly, employing the above mentioned Eigen analysis to each local patch of the image results into a bag-of-features approach. Pentland *et al.* extended the eigenface technique to a layered representation by combining eigenfaces and other eigenmodules, such as eigeneyes, eigennoses, and eigenmouths(Pentland et al, 1994). Recognition is then performed by finding a projection of the test image patch to each of the learned local Eigen subspaces for every individual.

4. Local Feature Extraction

(Gottumukkal and Asari, 2004) argued that some of the local facial features did not vary with pose, direction of lighting and facial expression and, therefore, suggested dividing the face region into smaller sub images. The goal of local feature extraction thus becomes to represent these local regions effectively and comprehensively. Here we review the most commonly used local feature extraction techniques in face recognition namely the Gabor wavelet transform based features , discrete cosine transform DCT-based features and more recently proposed Local binary pattern LBP features.

4.1 2D Gabor wavelets

The 2D Gabor elementary function was first introduced by Granlund (Granlund, 1978). Gabor wavelets demonstrate two desirable characteristic: spatial locality and orientation selectivity. The structure and functions of Gabor kernels are similar to the two-dimensional receptive fields of the mammalian cortical simple cells (Hubel and Wiesel, 1978). (Olshausen and Field, 1996; Rao and Ballard, 1995; Schiele and Crowley, 2000) indicates that the Gabor wavelet representation of face images should be robust to variations due to illumination and

facial expression changes. Two-dimensional Gabor wavelets were first introduced into biometric research by Daugman (Daugman, 1993) for human iris recognition. Lades et al. (Lades et al, 1993) first apply Gabor wavelets for face recognition using the Dynamic Link Architecture framework.

A Gabor wavelet kernel can be thought of a product of a complex sinusoid plane wave with a Gaussian envelop. A Gabor wavelet generally used in face recognition is defined as (Liu, 2004):

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{-\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}} [e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}}] \tag{5}$$

where z = (x, y) is the point with the horizontal coordinate x and the vertical coordinate y in the image plane. The parameters u and v define the orientation and frequency of the Gabor kernel, $\|.\|$ denotes the norm operator, and σ 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 $k_{\mu,v}$ is defined as $k_{u,v} = k_v e^{i\phi_u}$.

Following the parameters suggested in (Lades et al, 1993) and used widely in prior works (Liu, 2004) (Liu and Wechsler, 2002) $k_v = \frac{k_{\max}}{f^v}$ and $\phi_u = \frac{\pi u}{8}$. k_{max} is the maximum frequency, and f^v is the spatial frequency between kernels in the frequency domain. $v \in \{0,...,4\}$ and $u \in \{0,...,7\}$ in order to have a Gabor kernel tuned to 5 scales and 8 orientations. Gabor wavelets are chosen relative to $\sigma = 2\pi$, $k_{\max} = \frac{\pi}{2}$ and $f = \sqrt{2}$. The parameters ensures that frequencies are spaced in octave steps from 0 to π, typically each Gabor wavelet has a frequency bandwidth of one octave that is sufficient to have less overlap and cover the whole spectrum.

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels as defined by equation (6). The convolution of image *I* and a Gabor kernel $\psi_{u,v}(z)$ is defined as follows:

$$G_{u,v}(z) = I(z) * \psi_{u,v}(z) \tag{6}$$

where $z = (x, y)$ denotes the image position, the symbol '$*$' denotes the convolution operator, and $G_{u,v}(z)$ is the convolution result corresponding to the Gabor kernel at scale v and orientation u . The Gabor wavelet coefficient is a complex with a real and imaginary part, which can be rewritten as $G_{u,v}(z) = A_{u,v}(z).e^{i\theta_{u,v}(z)}$, where $A_{u,v}$ is the magnitude response and $\theta_{u,v}$ is the phase of Gabor kernel at each image position. It is known that the magnitude varies slowly with the spatial position, while the phases rotate in some rate with positions, as can be seen from the example in figure 2. Due to this rotation, the phases taken from image points only a few pixels apart have very different values, although representing almost the same local feature (Wiskott et al, 1997). This can cause severe problems for face

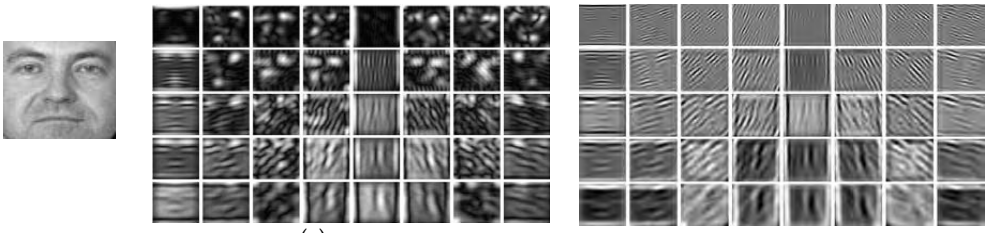

Fig. 2. Visualization of (a) Gabor magnitude (b) Gabor phase response, for a face image with 40 Gabor wavelets (5 scales and 8 orientations).

matching, and it is just the reason that all most all of the previous works make use of only the magnitude part for face recognition. Note that, convolving an image with a bank of Gabor kernel tuned to 5 scales and 8 orientations results in 40 magnitude and phase response maps of the same size as image. Therefore, considering only the magnitude response for the purpose of feature description, each pixel can be now described by a 40 dimensional feature vector (by concatenating all the response values at each scale and orientation) describing the response of Gabor filtering at that location.

Note that Gabor feature extraction results in a highly localized and over complete response at each image location. In order to describe a whole face image by Gabor feature description the earlier methods take into account the response only at certain image locations, e.g. by placing a coarse rectangular grid over the image and taking the response only at the nodes of the grid (Lades et al, 1993) or just considering the points at important facial landmarks as in (Wiskott et al, 1997). The recognition is then performed by directly comparing the corresponding points in two images. This is done for the main reason of putting an upper limit on the dimensionality of the problem. However, in doing so they implicitly assume a perfect alignment between all the facial images, and moreover the selected points that needs to be compared have to be detected with pixel accuracy.

One way of relaxing the constraint of detecting landmarks with pixel accuracy is to describe the image by a global feature vector either by concatenating all the pixel responses into one long vector or employ a feature selection mechanism to only include significant points (Wu and Yoshida, 2002) (Liu et al, 2004). One global vector per image results in a very high and prohibitive dimensional problem, since e.g. a 100x100 image would result in a 40x100x100=400000 dimensional feature vector. Some authors used Kernel PCA to reduce this dimensionality termed as Gabor-KPCA (Liu, 2004), and others (Wu and Yoshida, 2002; Liu et al, 2004; Wang et al, 2002) employ a feature selection mechanism for selecting only the important points by using some automated methods such as Adaboost etc. Nonetheless, a global description in this case still results in a very high dimensional feature vector, e.g. in (Wang et al, 2002) authors selected only 32 points in an image of size 64x64, which results in 32x40=1280 dimensional vector, due to this high dimensionality the recognition is usually performed by computing directly a distance measure or similarity metric between two images. The other way can be of taking a bag-of-feature approach where each selected point is considered an independent feature, but in this case the configural information of the face is effectively lost and as such it cannot be applied directly in situations where a large pose variations and other appearance variations are expected.

The Gabor based feature description of faces although have shown superior results in terms of recognition, however we note that this is only the case when frontal or near frontal facial images are considered. Due to the problems associated with the large dimensionality, and thus the requirement of feature selection, it cannot be applied directly in scenarios where large pose variations are present.

4.2 2D Discrete Cosine Transform

Another popular feature extraction technique has been to decompose the image on block by block basis and describe each block by 2D Discrete Cosine Transform 'DCT' coefficients. An image block $f(p,q)$, where $p,q = \{0,1..,\mathrm{N}-1\}$ (typically N=8), is decomposed terms of orthogonal 2D DCT basis functions. The result is a NxN matrix $C(v,u)$ containing 2D DCT coefficients:

$$C(v,u) = \alpha(v)\alpha(u)\sum_{y=0}^{N-1}\sum_{x=0}^{N-1} f(p,q)\beta(p,q,v,u) \tag{7}$$

where $v,u = 0,1,2,...,N-1$, $\alpha(v) = \sqrt{\frac{1}{N}}$ for v=0, and $\alpha(v) = \sqrt{\frac{2}{N}}$ for v=1,2,…,N-1 and

$$\beta(p,q,v,u) = \cos\left[\frac{(2p+1)v\pi}{2N}\right]\cos\left[\frac{(2q+1)u\pi}{2N}\right] \tag{8}$$

The coefficients are ordered according to a zig-zag pattern, reflecting the amount of information stored (Gonzales and Woods, 1993). For a block located at image position (x,y), the baseline 2D DCT feature vector is composed of:

$$x = [c_o^{(x,y)} \quad c_1^{(x,y)} ... \quad c_{M-1}^{(x,y)}]^T \tag{9}$$

Where $c_n^{(x,y)}$ denotes the *n-th* 2D DCT coefficient and M is the number of retained coefficients3. To ensure adequate representation of the image, each block overlaps its horizontally and vertically neighbouring blocks by 50% (Eickeler et al, 2000). M is typically set to 15 therefore each block yields a 15 dimensional feature vector. Thus for an image which has Y rows and X columns, there are $N_D = (2\frac{Y}{N}-1)\times(2\frac{X}{N}-1)$ blocks.

DCT based features have mainly been used in Hidden Markov Models HMM based methods in frontal scenarios. More recently (Cardinaux et al, 2006) proposed an extension of conventional DCT based features by replacing the first 3 coefficients with their corresponding horizontal and vertical deltas termed as DCTmod2, resulting into an 18-dimensional feature vector for each block. The authors claimed that this way the feature vectors are less affected by illumination change. They then use a bag-of-feature approach to derive person specific face models by using Gaussian mixture models.

4.2 Local Binary Pattern Histogram LBPH and its variants

Local binary pattern (LBP) was originally designed for texture classification (Ojala et al, 2002), and was introduced in face recognition in (Ahonen et al, 2004). As mentioned in

(Ahonen et al, 2004) the operator labels the pixels of an image by thresholding some neighbourhood of each pixel with the centre value and considering the result as a binary number. Then the histogram of the labels can be used as a texture descriptor. See figure 3 for an illustration of the basic $LBP_{P,R}^{U2}$ operator. The face area is divided into several small windows. Several LBP operators are compared and $LBP_{8,2}^{U2}$ the operator in 18x21 pixel windows is recommended because it is a good trade-off between recognition performance and feature vector length. The subscript represents using the operator in a (P, R) neighbourhood. Superscript U2 represent using only uniform patterns and labelling all remaining patterns with a single label, see (Ahonen et al, 2004) for details. The χ^2 statistic and the weighted χ^2 statistic were adopted to compare local binary pattern histograms.

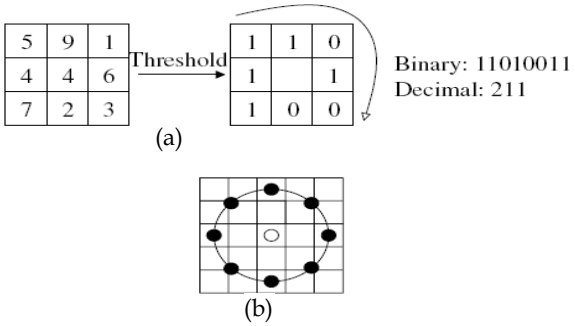

Fig. 3. (a) the basic LBP operator. (b) The circular (8,2) neighbourhood. The pixel values are bilinearly interpolated whenever the sampling point is not in the centre of a pixel (Ahonen et al, 2004)

Recently (Zhang et al, 2005) proposed local Gabor binary pattern histogram sequence (LGBPHS) by combining Gabor filters and the local binary operator. (Baochang et al, 2007) further used LBP to encode Gabor filter phase response into an image histogram termed as Histogram of Gabor Phase Patterns (HGPP).

5. Face-GLOH-Signatures –introduced feature representation

The mostly used local feature extraction and representation schemes presented in previous section have mainly been employed in a frontal face recognition task. Their ability to perform equally well when a significant pose variation is present among images of the same person cannot be guaranteed, especially when no alignment is assumed among facial images. This is because when these feature representations are used as a global description the necessity of having a precise alignment becomes unavoidable. While representations like 2D-DCT or LBP are much more susceptible to noise, e.g. due to illumination change as noted in (Zou et al, 2007) or pose variations, Gabor based features are considered to be more invariant with respect to these variations. However, as discussed earlier the global Gabor representation results in a prohibitively high dimensional problem and as such cannot be directly used in statistical based methods to model these in-class variations due to pose for instance. Moreover the effect of misalignments on Gabor features has been studied

(Shiguang et al, 2004), where strong performance degradation is observed for different face recognition systems.

As to the question, what description to use, there are some guidelines one can benefit from. For example, as discussed in section 3.1 the configural relationship of the face has to be preserved. Therefore a global representation as opposed to a bag-of-features approach should be preferred. Further in order to account for the in-class variations the local regions of the image should be processed in a scale, rotation and translation invariant manner. Another important consideration should be with respect to the size of the local region used. Some recent studies (Martınez, 2002; Ullman et al, 2002; Zhang et al, 2005) show that large areas should be preferred in order to preserve the identity in face identification scenarios.

Keeping in view the preceding discussion we use features proposed in (Mikolajczyk and Schmid, 2005), used in other object recognition tasks, and introduce to employ these for the task of face recognition for the first time (Sarfraz, 2008; Sarfraz and Hellwich, 2008) Our approach is to extract whole appearance of the face in a manner which is robust against misalignments. For this the feature description is specifically adapted for the purpose of face recognition. It models the local parts of the face and combines them into a global description We use a representation based on gradient location-orientation histogram (GLOH) (Mikolajczyk and Schmid, 2005), which is more sophisticated and is specifically designed to reduce in-class variance by providing some degree of invariance to the aforementioned transformations.

GLOH features are an extension to the descriptors used in the scale invariant feature transform (SIFT) (Lowe, 2004), and have been reported to outperform other types of descriptors in object recognition tasks (Mikolajczyk and Schmid, 2005). Like SIFT the GLOH descriptor is a 3D histogram of gradient location and orientation, where location is quantized into a log-polar location grid and the gradient angle is quantized into eight orientations. Each orientation plane represents the gradient magnitude corresponding to a given orientation. To obtain illumination invariance, the descriptor is normalized by the square root of the sum of squared components.

Originally (Mikolajczyk and Schmid, 2005) used the log-polar location grid with three bins in radial direction (the radius set to 6, 11, and 15) and 8 in angular direction, which results in 17 location bins. The gradient orientations are quantized in 16 bins. This gives a 272 bin histogram. The size of this descriptor is reduced with PCA. While here the extraction procedure has been specifically adapted to the task of face recognition and is described in the remainder of this section.

The extraction process begins with the computation of scale adaptive spatial gradients for a given image $I(x,y)$. These gradients are given by:

$$\nabla_{xy} \equiv \sum_{t} w(x,y,t)\sqrt{t}\nabla^{t}_{xy} L(x,y;t) \tag{10}$$

where $L(x,y;t)$ denotes the linear Gaussian scale space of $I(x,y)$ (Lindeberg, 1998) and $w(x,y,t)$ is a weighting, as given in equation 11.

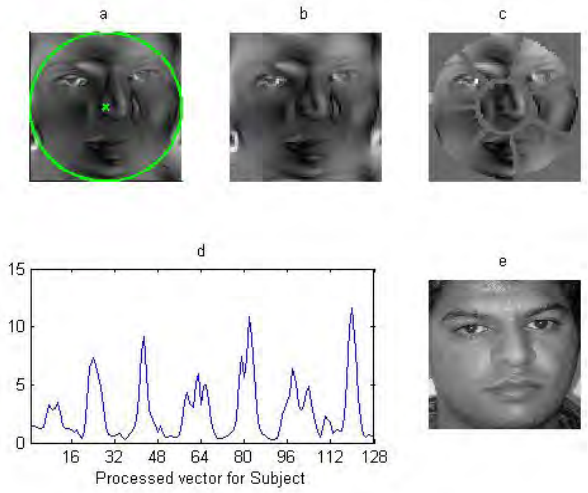

Fig. 5. Face-GLOH-Signature extraction (a-b) Gradient magnitudes (c) polar-grid partitions (d) 128-dimentional feature vector (e) Example image of a subject.

$$w(x,y,t) = \frac{\left|\sqrt{t}\nabla^{t}_{xy} L(x,y;t)\right|^{4}}{\sum_{t}\left|\sqrt{t}\nabla^{t}_{xy} L(x,y;t)\right|^{4}} \tag{11}$$

The gradient magnitudes obtained for an example face image (Figure 5 e) are shown in Figure 5 b. The gradient image is then partitioned on a grid in polar coordinates, as illustrated in Figure 5 c. As opposed to the original descriptor the partitions include a central region and seven radial sectors. The radius of the central region is chosen to make the areas of all partitions equal. Each partition is then processed to yield a histogram of gradient magnitude over gradient orientations. The histogram for each partition has 16 bins corresponding to orientations between 0 and 2π, and all histograms are concatenated to give the final 128 dimensional feature vector, that we term as face-GLOH-signature, see Figure 5 d. No PCA is performed in order to reduce the dimensionality.

The dimensionality of the feature vector depends on the number of partitions used. A higher number of partitions results in a longer vector and vice versa. The choice has to be made with respect to some experimental evidence and the effect on the recognition performance. We have assessed the recognition performance on a validation set by using ORL face database. By varying the partitions sizes from 3 (1 central region and 2 sectors), 5, 8, 12 and 17, we found that increasing number of partitions results in degrading performance especially with respect to misalignments while using coarse partitions also affects recognition performance with more pose variations. Based on the results, 8 partitions seem to be the optimal choice and a good trade off between achieving better recognition performance and minimizing the effect of misalignment. The efficacy of the descriptor is demonstrated in the presence of pose variations and misalignments, in the next section. It

should be noted that, in practice, the quality of the descriptor improves when care is taken to minimize aliasing artefacts. The recommended measures include the use of smooth partition boundaries as well as a soft assignment of gradient vectors to orientation histogram bins.

6. Performance Analysis

In order to assess the performance of introduced face-GLOH-signature with that of various feature representations, we perform experiments in two settings. In the first setting, the problem is posed as a typical multi-view recognition scenario, where we assume that few number of example images of each subject are available for training. Note that, global feature representations based on Gabor, LBP and DCT cannot be directly evaluated in this setting because of the associated very high dimensional feature space. These representations are, therefore, evaluated in a typical template matching fashion in the second experimental setting, where we assess the performance of each representation across a number of pose mismatches by using a simple similarity metric. Experiments are performed on two of the well-known face databases i.e. FERET (Philips et al, 2000) and ORL face database (http://www.cam-orl.co.uk).

6.1 Multi-view Face recognition

In order to perform multi-view face recognition (recognizing faces under different poses) it is generally assumed to have examples of each person in different poses available for training. The problem is solved form a typical machine learning point of view where each person defines one class. A classifier is then trained that seek to separate each class by a decision boundary. Multi-view face recognition can be seen as a direct extension of frontal face recognition in which the algorithms require gallery images of every subject at every pose (Beymer, 1996). In this context, to handle the problem of one training example, recent research direction has been to use specialized synthesis techniques to generate a given face at all other views and then perform conventional multi-view recognition (Lee and Kim, 2006; Gross et al, 2004). Here we focus on studying the effects on classification performance when a proper alignment is not assumed and there exist large pose differences. With these goals in mind, the generalization ability of different conventional classifiers is evaluated with respect to the small sample size problem. Small sample size problem stems from the fact that face recognition typically involves thousands of persons in the database to be recognized. Since multi-view recognition treats each person as a separate class and tends to solve the problem as a multi-class problem, it typically has thousands of classes. From a machine learning point of view any classifier trying to learn thousands of classes requires a good amount of training data available for each class in order to generalize well. Practically, we have only a small number of examples per subject available for training and therefore more and more emphasis is given on choosing a classifier that has good generalization ability in such sparse domain.

The other major issue that affects the classification is the representation of the data. The most commonly used feature representations in face recognition have been introduced in previous sections. Among these the Eigenface by using PCA is the most common to be used in multi-view face recognition. The reason for that is the associated high dimensionality of other feature descriptions such as Gabor, LBPH etc. that prohibits the use of any learning to

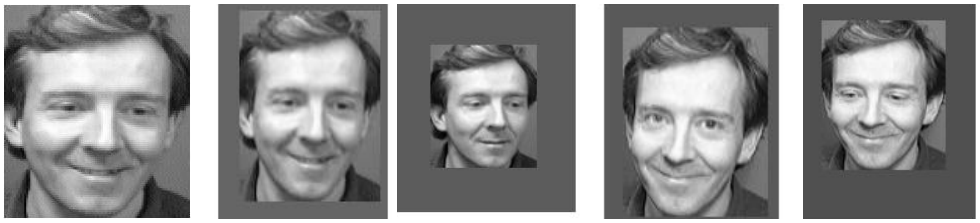

Fig. 6. An example of a subject from O-ORL and its scale and shifted examples from SS-ORL

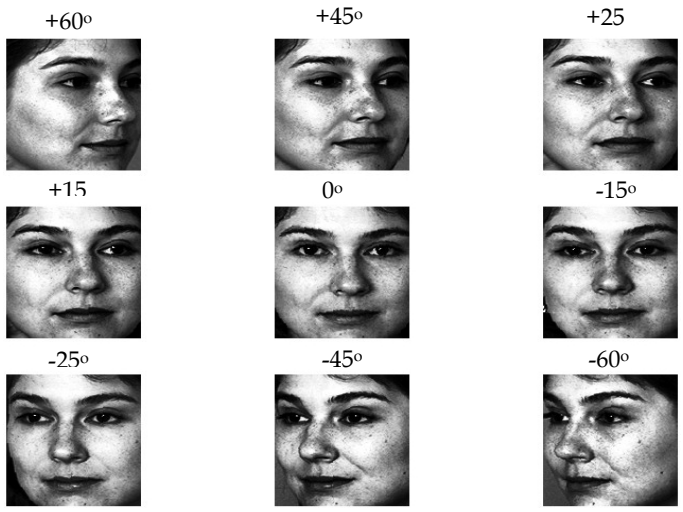

Fig. 7. Cropped faces of a FERET subject depicting all the 9 pose variations.

be done. This is the well known curse of dimensionality issue in pattern recognition (Duda et al, 2001) literature and this is just the reason that methods using such over complete representations normally resort to performing a simple similarity search by computing distances of a probe image to each of the gallery image in a typical template matching manner. While by using PCA on image pixels an upper bound on the dimensionality can be achieved.

In line with the above discussion, we therefore demonstrate the effectiveness of the proposed face-GLOH signatures with that of using conventional PCA based features in multi-view face recognition scenarios with respect to the following factors.

When facial images are not artificially aligned

When there are large pose differences

Large number of subjects

Number of examples available in each class (subject) for training.

In order to show the effectiveness of face-GLOH signature feature representation against misalignments, we use ORL face database. ORL face database has 400 images of 40 subjects (10 images per subject) depicting moderate variations among images of same person due to expression and some limited pose. Each image in ORL has the dimension of 192x112 pixels.

All the images are depicted in approximately the same scale and thus have a strong correspondence among facial regions across images of the same subject. We therefore generate a scaled and shifted ORL dataset by introducing an arbitrary scale change between 0.7 and 1.2 of the original scale as well as an arbitrary shift of 3 pixels in random direction in each example image of each subject. This has been done to ensure having no artificial alignment between corresponding facial parts. This new misaligned dataset is denoted scaled-shifted SS-ORL (see Figure 6). The experiments are performed on both the original ORL denoted O-ORL and SS-ORL using PCA based features and face-GLOH signatures. ORL face database is mainly used to study the effects on classification performance due to misalignments since variations due to pose are rather restricted (not more than 20^o). To study the effects of large pose variations and a large number of subjects, we therefore repeat our experiments on FERET database pose subset. The FERET pose subset contains 200 subjects, where each subject has nine images corresponding to different pose angles (varying from 0^o frontal to left/right profile $\pm 60^o$) with an average pose difference of 15^o. All the images are cropped from the database by using standard normalization methods i.e. by manually locating eyes position and warping the image onto a plane where these points are in a fixed location. The FERET images are therefore aligned with respect to these points. This is done in order to only study the effects on classifier performance due to large pose deviations. All the images are then resized to 92x112 pixels in order to have the same size as that of ORL faces. An example of the processed images of a FERET subject depicting all the 9 pose variations is shown in Figure 7.

We evaluate eight different conventional classifiers. These include nearest mean classifier 'NMC', linear discriminant classifier 'LDC', quadratic 'QDC', fisher discriminant, parzen classifier, k-nearest neighbour 'KNN', Decision tree and support vector machine 'SVM', see (Webb, 2002) for a review of these classifiers.

6.1.1 Experiments on ORL database

We extract one global feature vector per face image by using lexicographic ordering of all the pixel grey values. Thus, for each 92 x 112 ORL image, one obtains a 10384 dimensional feature vector per face. We then reduce this dimensionality by using unsupervised PCA. Where the covariance matrix is trained using 450 images of 50 subjects from FERET set. The number of projection Eigen-vectors are found by analysing the relative cumulative ordered eigenvalues (sum of normalized variance) of the covariance matrix. We choose first 50 largest Eigen vectors that explain around 80% of the variance as shown in figure 4-3. By projecting the images on these, we therefore obtain a 50-dimentional feature vector for each image. We call this representation the PCA-set.

The second representation of all the images is found by using face-GLOH-signature extraction, as detailed in section 5.

In all of our experiments we assume equal priors for training, SVM experiments on O-ORL use a polynomial kernel of degree 2, to reduce the computational effort, since using RBF kernel with optimized parameters C and kernel width σ did not improve performance. For SS-ORL a RBF kernel is used with parameter C=500 and σ = 10, these values were determined using 5-fold cross validation and varying sigma between 0.1 and 50 and C between 1 and 1000. All the experiments are carried out for classifiers on each of two representations for both O-ORL and SS-ORL.

We use a 10-fold cross validation procedure to produces 10 sets of the same size as original dataset each with a different 10% of objects being used for testing. All classifiers are evaluated on each set and the classification errors are averaged. The results from this experiment on both O- ORL and SS-ORL for both feature representations are reported in table 1.

Classifiers	O-ORL Representation sets		SS-ORL Representation sets	
	PCA	face-GLO H	PCA	face-GLOH
NMC	0.137	0.152	0.375	0.305
LDC	0.065	0.020	0.257	0.125
Fisher	0.267	0.045	0.587	0.115
Parzen	0.037	0.030	0.292	0.162
3-NN	0.097	0.062	0.357	0.255
Decision Tree	0.577	0.787	0.915	0.822
QDC	0.64	0.925	0.760	0.986
SVM	0.047	0.037	0.242	0.105

Table 1. Classification errors in 10-fold cross validation tests on ORL

Table 1 shows how classification performance degrades, when the faces are not aligned i.e. arbitrarily scaled and shifted, on PCA based feature representation. The robustness of the face-GLOH-signature representation against misalignments can be seen by comparing the results on O-ORL and SS-ORL, where it still gives comparable performance in terms of classification accuracy. Best results are achieved by using LDC or SVM in both cases.

6.1.2 Experiments on FERET database

As stated earlier, FERET database pose subset is used to assess the performance with regards to large pose variations and large number of subjects. 50 out of 200 FERET subjects are used for training the covariance matrix for PCA. The remaining 1350 images of 150 subjects are used to evaluate classifier performance with respect to large pose differences. In order to assess the small sample size problem (i.e. number of raining examples available per subject), experiments on FERET are performed with respect to varying training/test sizes by using 2, 4, 6, and 8 examples per subject and testing on the remaining. Similarly, tests at each size are repeated 5 times, with different training/test partitioning, and the errors are averaged. Figure 8 shows the averaged classification errors for all the classifiers on FERET set for both the feature representations with respect to varying training and test sizes. As shown in figure 8, increasing number of subjects and pose differences has an adverse affect on the performance of all the classifiers on PCA-representation set while face-GLOH-Signature representation provides relatively better performance.

6.2 Template matching Setting

As stated earlier, due to the associated high dimensionality of the extracted features of GABOR, LBP, DCT etc, we assess the performance of these feature descriptions with that of face-GLOH signature across a number of pose mismatches in a typical template matching

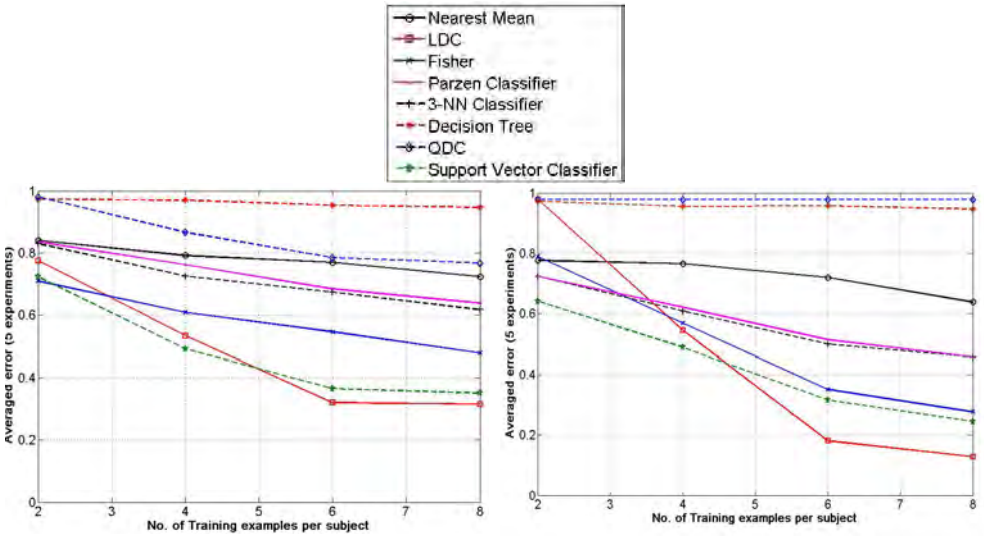

Fig. 8. Classifiers evaluation On FERET by varying training/test sizes (a) Using PCA-set (b) Using face-GLOH-signature set

setting. Frontal images of 200 FERET subjects are used as gallery while images for the remaining eight poses of each subject are used as test probes. Each probe is matched with each of the gallery images by using the cosine similarity metric. Probe is assigned the identity of the gallery subject for which it has the maximum similarity.

6.2.1 Test Results

We obtain each of the three feature descriptions as described in section 4. Gabor features are obtained by considering real part of the bank of Gabor filter kernel response tuned to 8 orientations and 5 scales, at each pixel location. This resulted in 40x92x112=412160 dimensional feature vector for each image. Due to memory constraints we used PCA to reduce the dimensionality to 16000-dimensional vector. For the LBPH feature representation, we use $LBP_{8,2}^{U2}$ operator in 18x21 window as described in (Ahonen et al, 2004) which resulted in a 2124 dimensional feature vector. The recognition scores in each pose are averaged. Table 2 depicts the performance comparison of different feature representations with that of Face-GLOH-Signature across a number of pose mismatches.

Feature Description	Average Recognition across each FERET Probe Pose			
	$\pm 15^o$	$\pm 25^o$	$\pm 45^o$	$\pm 60^o$
Eigenface	70.1%	56.2%	31.1%	13.4%
Gabor	91.4%	81.2%	68.5%	32.1%
LBPH	87.3 %	71.4%	56%	18.3%
Face-GLOH-Signature	100%	94.5%	81.1%	53.8%

Table 2. Comparison of face identification performance across pose of different feature representations

7. Conclusion

A comprehensive account of almost all the feature extraction methods used in current face recognition systems is presented. Specifically we have made distinction in the holistic and local feature extraction and differentiate them qualitatively as opposed to quantitatively. It is argued that a global feature representation should be preferred over a bag-of-feature approach. The problems in current feature extraction techniques and their reliance on a strict alignment is discussed. Finally we have introduced to use face-GLOH signatures that are invariant with respect to scale, translation and rotation and therefore do not require properly aligned images. The resulting dimensionality of the vector is also low as compared to other commonly used local features such as Gabor, LBP etc. and therefore learning based methods can also benefit from it.

In a typical multi-view face recognition task, where it is assumed to have several examples of a subject available for training, we have shown in an extensive experimental setting the advantages and weaknesses of commonly used feature descriptions. Our results show that under more realistic assumptions, most of the classifiers failed on conventional features. While using the introduced face-GLOH-signature representation is relatively less affected by large in-class variations. This has been demonstrated by providing a fair performance comparison of several classifiers under more practical conditions such as misalignments, large number of subjects and large pose variations. An important conclusion is to be drawn from the results on FERET is that conventional multi-view face recognition cannot cope well with regards to large pose variations. Even using a large number of training examples in different poses for a subject do not suffice for a satisfactory recognition. In order to solve the problem where only one training example per subject is available, many recent methods propose to use image synthesis to generate a given subject at all other views and then perform a conventional multi-view recognition (Beymer and Poggio, 1995; Gross et al, 2004). Besides the fact that such synthesis techniques cause severe artefacts and thus cannot preserve the identity of an individual, a conventional classification cannot yield good recognition results, as has been shown in an extensive experimental setting. More sophisticated methods are therefore needed in order to address pose invariant face recognition. Large pose differences cause significant appearance variations that in general are larger than the appearance variation due to identity. One possible way of addressing this is to learn these variations across each pose, more specifically by fixing the pose and establishing a correspondence on how a person's appearance changes under this pose one could reduce the in-class appearance variation significantly. In our very recent work (Sarfraz and Hellwich, 2009), we demonstrate the usefulness of face-GLOH signature in this direction.

8. References

Ahonen, T., Hadid, A. & Pietikainen, M. (2004). Face recognition with local binary patterns, *Proceedings of European Conference on Computer Vision ECCV*, pp. 469–481.

Ashraf A.B., Lucey S., and Chen T. (2008). Learning patch correspondences for improved viewpoint invariant face recognition. *Proceedings of IEEE Computer Vision and Pattern Recognition CVPR*, June.

Baochang Z., Shiguang S., Xilin C., and Wen G. (2007). Histogram of Gabor Phase Patterns (HGPP):A Novel Object Representation Approach for Face Recognition, *IEEE Trans. on Image Processing*, vol. 16, No. 1, pp. 57-68.

Beymer D. (1996). Pose-invariant face recognition using real and virtual Views. *M.I.T., A.I. Technical Report* No.1574, March.

Beymer, D. Poggio, T. (1995). Face recognition from one model view. *Proceedings of International conference on computer vision.*

Belhumeur, P. N., Hespanha, J. P. & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711-720.

Brunelli R. and Poggio T. (1993). Face recognition: Features versus templates. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1052.

Chen, L.F., Liao, H.Y., Lin, J.C. & Han, C.C. (2001). Why recognition in a statistics-based face recognition system should be based on the pure face portion: a probabilistic decision- based proof, *Pattern Recognition*, Vol. 34, No. 7, pp. 1393-1403.

Cardinaux, F., Sanderson, C. & Bengio, D S. (2006). User authentication via adapted statistical models of face images. *IEEE Trans. Signal Processing*, 54(1):361–373.

Daugman, J. (1993). High confidence visual recognition of persons by a test of statistical independence. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 15 1148–1161.

Duda, R.O., Hart, P.E. & Stork, D.G. (2001). *Pattern Classification*, 2nd edition, Wiley Interscience.

Eickeler, S., Müller, S. & Rigoll, G. (2000). Recognition of JPEG Compressed Face Images Based on Statistical Methods, *Image and Vision Computing*, Vol. 18, No. 4, pp. 279-287.

Gonzales,R. C. & Woods, R. E. (1993) *Digital Image Processing*, Addison-Wesley, Reading, Massachusetts.

Granlund, G. H. (1978). Search of a General Picture Processing Operator, *Computer Graphics and Image Processing*, 8, 155-173.

Gottumukkal, [R. & Asari, V. K. (2004). An improved face recognition technique based on modular PCA approach, *Pattern Recognition Letter*, vol. 25, no. 4, pp. 429–436.

Hubel, D., Wiesel, T. (1978). Functional architecture of macaque monkey visual cortex, *Proceedings of Royal Society on Biology,* 198 (1978) 1–59.

Gross R., Matthews I. and Baker S. (2004). Appearance-based face recognition and light-fields, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 449–465.

Hayward, W. G., Rhodes, G. & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition, *Cognition* 106(2), 1017-1027.

Kanade, T., & Yamada, A. (2003). Multi-subregion based probabilistic approach towards pose-Invariant face recognition. *Proceedings of IEEE international symposium on computational intelligence in robotics automation,* Vol. 2, pp. 954–959.

Lindeberg T. (1998) Feature detection with automatic scale selection. *Int. Journal of computer vision*, vol. 30 no. 2, pp 79-116.

Lowe D. (2004). Distinctive image features from scale-invariant keypoints. *Int. Journal of computer vision*, 2(60):91-110.

Liu, C. (2004). Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 26, pp. 572–581.

Liu, C., Wechsler, H. (2002). Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 11, pp. 467–476.

Lucey, S. & Chen, T. (2006). Learning Patch Dependencies for Improved Pose Mismatched Face Verification, *Proceedings of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 17-22.

Lades, M., Vorbruggen, J., Budmann, J., Lange, J., Malsburg, C., Wurtz, R. (1993). Distortion invariant object recognition on the dynamic link architecture. *IEEE Transactions on Computers,* 42, 300–311.

Lee H.S., Kim D. (2006). Generating frontal view face image for pose invariant face recognition. *Pattern Recognition letters,* vol. 27, No. 7, pp. 747-754.

Liu, D. H., Lam, K. M., & Shen, L. S. (2004). Optimal sampling of Gabor features for face recognition. *Pattern Recognition Letters*, 25, 267-276.

Marta, P., Cassia, M. & Chiara,T. (2006). The development of configural face processing: the face inversion effect in preschool-aged children, Annual *meeting of the XVth Biennial International Conference on Infant Studies*, Jun 19, Westin Miyako, Kyoto, Japan.

Mikolajczyk and Schmid C. (2002). Performance evaluation of local descriptors, *IEEE Transaction on Pattern Analysis and Machine Intelligence PAMI*, 27(10), 31-47.

Martınez A. M. (2002) Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transaction on Pattern Analysis and Machine Intelligence PAMI*, vol. 24, no. 6, pp. 748–763.

Ojala, T., Pietikainen, M. & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987.

Olshausen, B., Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381, 607–609.

Pentland, A., Moghaddam, B. & Starner, T. (1994). View-Based and modular eigenspaces for face recognition," *Proceedings IEEE Conference of Compute Vision and Pattern Recognition*, pp. 84–91.

Phillips, P.J., Moon, H., Rizvi, S.A. & Rauss, P.J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104.

Rao, R., Ballard, D. (1995). An active vision architecture based on iconic representations, *Artificial Intelligence,* 78,461–505.

Sarfraz, M.S. and Hellwich, O. (2008). Statistical Appearance Models for Automatic Pose Invariant Face Recognition, *Proceedings of 8th IEEE Int. conference on Face and Gesture Recognition 'FG',* IEEE computer Society , September 2008, Holland.

Sarfraz, Muhammad Saquib (2008). Towards Automatic Face Recognition in Unconstrained Scenarios", *PhD Dissertation*, urn:nbn:de:kobv:83-opus-20689.

Sarfraz, M.S., Hellwich, O. (2009)" Probabilistic Learning for Fully Automatic Face Recognition across Pose", *Image and Vision Computing*, Elsevier, doi: 10.1016/j.imavis.2009.07.008.

Schwaninger, A., Wallraven, C., Cunningham, D. W. & Chiller-Glaus, S. (2006). Processing of identity and emotion in faces: a psychophysical, physiological and computational perspective. *Progress in Brain Research* 156, 321-343.

Schiele, B., Crowley, J. (2000). Recognition without correspondence using multidimensional receptive field histograms, *International Journal on Computer Vision*, 36 31–52.

Shiguang, S., Wen, G., Chang,Y., Cao,B., Yang, P. (2004). Review the Strength of Gabor features for Face Recognition from the Angle of its Robustness to Misalignment, *Proceedings of International conference on Pattern Recognition ICPR*.

Turk, M. & Pentland, A. (1991). Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86.

Tan, K. & Chen, S. (2005). Adaptively weighted sub-pattern PCA for face recognition, *Neurocomputing*, 64, pp. 505–511.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, (7), 682–687.

Wang, Y. J., Chua, C. S., & Ho, Y. K. (2002). Facial feature detection and face recognition from 2D and 3D images. *Pattern Recognition Letters*, 23, 1191-1202.

Webb, A.R. (2002). *Statistical pattern recognition*, 2nd edition, John Willey & Sons.

Wiskott, L., Fellous, J., Krüger, N., Malsburg, C. (1997) Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 775–779.

Wu, H. Y., Yoshida Y., & Shioyama, T. (2002). Optimal Gabor filters for high speed face identification. *Proceedings of International Conference on pattern Recognition*, pp. 107-110.

Zhang, Lingyun and Garrison W. Cottrell (2005). Holistic processing develops because it is good. *Proceedings of the 27th Annual Cognitive Science Conference*, Italy.

Jie Zou, Qiang Ji, and George Nagy (2007). A Comparative Study of Local Matching Approach for Face Recognition, *IEEE Transaction on image processing*, VOL. 16, NO.10.

Zhang, W., Shan, S., Gao, W., Chen, X. & Zhang, H. (2005). Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," *Proceedings International Conference of Computer Vision ICCV*, pp. 786–791.