Chapter 2

Literature Survey

Being an active research area for more than three decades, a very large research literature can be found on face recognition. Over the years both machine learning researchers and image processing/computer vision researchers have contributed a massive amount of knowledge to this field. Each year researchers around the world publish more and more novel techniques in face recognition. This chapter briefly summarizes methodologies and some of the most popular algorithms that have been successfully applied to face recognition.

2.1 Face Recognition Methods

Face recognition can be broadly divided into two methods as still image based methods and video based methods. In all these methods a face recognition system has to first detect and segment the face from its background and then process to extract features that are finally used to recognize the face. Since this thesis is considering only the still images based face recognition our focus is directed more towards still image based methods and algorithms. In the research literature [38] (as in Figure 2.1) following three methods of processing still image faces for recognition are described.

![Figure 2.1 Schematic diagram of face recognition methodologies](image_url)

Face recognition methods

- Holistic matching method
- Feature based method
- Hybrid method

Video based methods

- Still-image method
- Multimodal method
- Spatiotemporal method
• Holistic matching method: This is the most commonly used method and it considers the whole face region as a raw input to the recognition process. Facial components or segmentations of faces are not considered with the holistic method. Irrespective of the view point of the face and other attributes such as expressions holistic method takes the whole face (everything available) for recognition process.

• Feature based (structural) method: Geometric measures taken with respect to facial components such as eyes, nose and mouth are considered features inputs for recognition the face.

• Hybrid method: Use combination of holistic matching as well as feature based matching.

2.2 Popular Face Recognition Algorithms

Literature describes various techniques that are used for feature extraction and matching. Among these the most common for holistic face recognition include the following:

2.2.1 Principal Component Analysis (PCA)

One of the earliest and most successful applications of machine learning to face recognition was principal component analysis (PCA). Among them eigenfaces [1], [5], [21] is one the most popular and widely used face recognition methods. Eigenfaces have been applied to faces using all still image based recognition methods (holistic, feature based and hybrid).

With eigenfaces given a training set of N faces \{x_i | i = 1,2 ,… , N\} the average of faces \(\mu\) can be calculated as

\[
\mu = \frac{1}{N} \sum_{i=1}^{N} x_i
\]  

(2.1)

The covariance matrix \(\mathbf{C}\) can be calculated as

\[
\mathbf{C} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu) (x_i - \mu)^T
\]  

(2.2)

Next step is to find the eigenvectors \(u_i\) and eigenvalues \(\lambda_i\) from the covariance matrix \((\mathbf{C})\). Eigenvectors need to be normalized to 1. The result is a \(N^2\) of eigenvectors and eigenvalues. Among them the set of \(M \ll N^2\) highest eigenvectors are chosen.

When a new face \(x\) a given for testing its eigenface components are calculated by

\[
w_k = u_k^T(x - \mu)
\]  

(2.3)
which makes the $\Omega = [w_1, w_2, ..., w_M]$ a vector of weight for all eigenvectors. Using $\Omega$ with other face class $\{\Omega_k | k = 1, ..., N\}$ the Euclidean distances $\{\varepsilon_k | k = 1, ..., N\}$ can be calculated as

$$\varepsilon_k = \|\Omega - \Omega_k\|$$

(2.4)

to find the minimum value which would lead to the corresponding person to whom the new face belongs.

### 2.2.2 Fisherfaces

Another popular face recognition approach is the fisherfaces [1]. It uses fisher discriminate analysis (FDA). In FDA given a set of N sample images $\{x_1, x_2, \ldots, x_N\}$ of n dimension one propose that each belongs to one of C classes $\{X_1, X_2, \ldots, X_c\}$ and a dimension reduction to m < n with a linear transformation to $\{y_i | i = 1, 2, \ldots, N\}$ as

$$y_i = Wx_i$$

(2.5)

In this method we consider two scatter matrices. The between-class scatter metric $S_B$ can be calculated as

$$S_B = \sum_{i=1}^{C} N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

(2.6)

and the within class scatter metric $S_W$ is defined as

$$S_W = \sum_{i=1}^{C} \sum_{x_k \in X_i} (x_k - \mu)(x_k - \mu)^T$$

(2.7)

where $\mu_i$ is the mean image of class $X_i$, $N_i$ is the number of samples in the class $X_i$ and $\mu$ is the mean image of all samples.

The selection of weight $W$ should be such that the ratio of the between class scatter and the within class scatter is maximized. $W_{opt}$ is the optimal weight matrix for the problem.

$$W_{opt} = \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|} = [w_1, w_2, \ldots, w_m]$$

(2.8)

The resulting vector set $\{w_i | i = 1, 2, \ldots, m\}$ is the generalized eigenvector of $S_W$ and $S_B$ for the largest m eigenvalues $\{\lambda_i | i = 1, 2, \ldots, m\}$ such that

$$S_Bw_i = \lambda_i S_Ww_i, \quad i = 1, 2, \ldots, m$$

(2.9)
For a given new face image $x'$ its corresponding class can be found by $y = W_{opt}x'$. In a similar manner to the eigenfaces by taking the Euclidean distance the new face can be classified. According to [1] fisherfaces has proven to be better than eigenfaces. With the arrival of kernel methods both eigenfaces and fisherfaces have been extended as kernel eigenfaces and kernel fisherfaces,[37] which has proven to be much better in performance.

### 2.2.3 Neural Networks

Neural networks [9], a well-known classification algorithm have also been applied to classification and recognition of faces [33]. Inspired by the properties of biological neuron networks and their learning ability many researches over the years developed many models of artificial neural network. These models have been successfully applied to many areas in pattern recognition, function approximations and data processing.

The basic model of a neural network is to have a collection of neurons, most of the time arranged as layers with connection between them such that signals can pass between them. As shown in the Figure 2.2 weights ($w_{ij}$) and bias ($b$) are set as parameters between nodes that can be learned in the process of training of the neural model based on the data. Among the neural network models applied to pattern recognition back-propagation and radial function networks are the most.

![Figure 2.2 Basic model of an artificial neural network](image)

With a training set of faces given as input and specifying the relevant value for each face at the output nodes, a neural network can be trained in online or batch modes for binary or multiclass classifications of faces. Once the weight and bias values of nodes are trained based on the optimizations strategy of the neural network can indicate the person for a given new face as input.

Application of Neural networks to problems like face recognition with high dimensions of data frequently suffers from curse of dimensionality. This happens due to the large number of
weight parameters due to connections between many nodes in the network. This leads to longer training times and lot of computational power. Some results on face recognition with neural networks and comparisons with other methods are available in [33].

2.2.4 Support Vector Machines
Among the most important developments in the last decade in the field of machine learning was the growth of kernel methods. Vapnik’s pioneering work on statistical learning theory [35], [36] and support vector machines gave birth to a very powerful methodology that uses a specific set of functions known as kernels that can be used to develop robust classification algorithms. A detailed description of support vector machines is given in the next chapter.

Support vector machines have been applied successfully in many areas in pattern recognition problem domain. These include bioinformatics, speech recognition, text classification and object recognition. With its simplicity and ability to represent data as vectors that can even be extended infinity while being powerful enough to work with large data sets quickly and support effective multi class classifying, SVM is one of the most popular tools available for the machine learning community. The plethora of research literature available on this topic also proves its vast growing popularity.

It also can be very successfully applied to face recognition [10], [12] and other closely related areas in facial analysis such as facial expression recognition and face detection. Many experiments conducted on several standard databases have indicated that SVM is more accurate than other popular methods such as eigenfaces [10], fisherfaces and neural networks [33]. Both holistic and feature based face recognition experiments are found in the literature. In [10] experimental results show that component based recognition outperform the holistic method.

Apart from the classification accuracy, the simplicity of the SVM algorithm and its ability to generate sparse solutions with lower training time and higher computational efficiency has made it a more prominent than other older classification methods neural networks. But SVM does not perform well with common problems of face recognition such as illuminations and expressions. Also SVM lack the ability to optimize using hyperparameters for kernels and at certain occasions both accuracy and computation time varies with the kernel parameters.

2.2.5 Gaussian Process
Another kernel based approach is Gaussian process [11],[15]. Due to its ability to provide much sparse solutions compared to SVM many researchers became interested in Gaussian process which led to the development of several interesting algorithms. An early and closely related
technique is the relevance vector machines developed by Tipping [34]. Relevance vector machine is one of the successful sparse Bayesian classification algorithms based on kernel methods. In the very limited literature of applications of RVM especially in facial analysis we found that in [6] results show that RVM performance is quite close to that of SVM even though its less accurate.

Even though there are several interesting Gaussian process based algorithms still some open problems are listed to be investigated. One such open problem proposed in NIPS’05 GP workshop [27] was the “good application” for Gaussian process to be applied. The basic theoretical construction of Gaussian process models is available in the next chapter.

2.3 Informative Vector Machines
One of the recent developments in the Gaussian process based classification algorithms is informative vector machines [17] – [20],[29],[30] which is mainly contributed by Ralph Herbrich, Neil Lawrence and Mathias Seeger. This probabilistic algorithm can optimize kernels parameters which lead to sparse much sparse solutions than SVM similar to RVM. Another special feature about is that it has more flexibility in choosing kernel functions and can employ automatic relevance determination (ARD) kernels. It also differs from other Gaussian process model since it works with an active data set that is selected based on entropy changes.

No previous research of face recognition with IVM is available but it has been successfully applied to digit classification [32] as well as text categorization [30]. In both these experiments IVM has produced sparse solutions than SVM and shown comparable accuracy. Though IVM seems like a promising method it has some limitations and problems. IVM selects its active data points using a greedy selection. This selection process sometimes relies on random selection which makes the selection process not optimal. Also the appropriate number of data points and the optimal stopping criteria cannot be specified other than by experiments.

2.4 Summary
As listed above in the broad literature of face recognition many algorithms are available that are capable of effective performance. But our research methodology described in rest of chapters is holistic face recognition using informative vector machines. The reason that we use kernel methods is its simplicity, flexibility, efficiency and accuracy compared to other algorithms. One major motivation to do research with a Gaussian process classifier is the lack of previous research on face recognition and to investigate whether face recognition a “good application” for Gaussian
process. As mentioned above SVM has proven to be a robust classifier for faces and informative vector machines has proven to have comparable performance in other recognition applications. The ability of IVM to generate sparse solutions than SVM deserves a broader investigation since it can be useful in improving storage and computational efficiency in real world applications. We also intend to investigate the problems related to IVM as mentioned above and try to propose improvements.