CHAPTER 2

PROPOSED APPROACH

2.1 DEFINITION OF THE PROBLEM

From the earlier discussions it is observed that the face recognition performance is based on the relation between the dimension of the sample space and the number of samples in the training set. This relation is described by two terms, SSS and LSS. From the literature survey it is observed that the face recognition performance of KDCV and KCV method is superior to other methods in SSS and LSS case respectively. However, these methods are suffering from limitations which are mentioned in section 1.4. Thus the prime focus of this research work is to improve the recognition accuracy of KDCV and KCV methods for recognizing face images varying in facial expressions and illuminative conditions.

2.2 OBJECTIVES OF THE STUDY

The main objective of this research work is to increase the recognition accuracy of face recognition methods KDCV and KCV, for recognizing face images varying in facial expressions and illuminative conditions. In order to achieve this, the following work is proposed:

(i) Analyze the recognition performance of KDCV and KCV methods and observe its limitations which are mentioned in section 1.4.

(ii) Modify the followed between class scatter matrix and within class scatter matrix of KDCV method such that the modified between class and within class scatter matrices overcome the first two limitations of KDCV method.

(iii) Introduce the between class scatter matrix in KCV method to solve the limitations of KCV method.

(iv) KDCV and KCV methods follow the assumption that the training samples are labeled correctly and these methods follow the feature
extraction process once. If there is any mislabel in the training set, these methods fail to classify a given test sample and act like a weak classifier in this environment. To improve this, the classifiers are adjusted iteratively, using a boosting technique in favor of those instances misclassified by previous classifiers.

2.3 METHODOLOGY

2.3.1 Pair Wise Scatter Parameter

The between class scatter matrix is defined as

\[ S_b^o = \sum_{i=1}^{C} N_i (\mu_i^o - \mu^o)(\mu_i^o - \mu^o)^T. \tag{2.1} \]

It is popularly accepted that the between-class scatter matrix carries the discriminant information. However, the between class scatter matrix in Equation (2.1) measures the distribution of the mean of each class from the center. The larger value of \( S_b^o \) shows the classes are spread out more in the transformed space and thus easier to discriminate them. If some classes are much closer as compared to others, the between class scatter matrix in Equation (2.1) ignores the discriminatory information between the classes that are close to each other i.e., the overlapping between the classes is not considered. Another observation from this expression is that, it is not clear how the classes are discriminated from each other, pair wise. This is due to the between class scatter matrix in Equation (2.1) followed the assumption that each class is equally spaced with all other classes in the mapped space. These limitations can be solved by modifying the between class scatter matrix in Equation (2.1) such that the pair wise class discriminatory information is included in Equation (2.1). The importance of inclusion of pair wise class discriminatory information is described with an example which is shown in Appendix 1.
From Appendix 1, it is clear that the consideration of pair wise class discriminatory information in $S^\phi_{b}$ Equation (2.1) is important. By this the between class scatter matrix in Equation (2.1) is modified as

$$S^\phi_{b} = \sum_{i,j=1}^{C} (pd)_{i,j}^{\phi} N_i N_j (\mu_i^\phi - \mu_j^\phi)(\mu_i^\phi - \mu_j^\phi)^T$$

(2.2)

where \(\{pd\}_{i,j}^{\phi}\) is a set of weights and \((pd)_{i,j}^{\phi}\) is a non-negative weight assigned to class pair \((i,j)\) and it represents the discrimination of class \(i\) from class \(j\).

If there is uniform weight, the weighted between class scatter matrix $S^\phi_{b}$ in Equation (2.2) is exactly same as between-class scatter matrix $S^\phi_{b}$ in Equation (2.1). This is proved in Appendix 1. Therefore it turns out that the definition of weighted between-class scatter matrix $S^\phi_{b}$ is a generalization of between-class scatter matrix $S^\phi_{b}$. Thus, the modified between class scatter matrix in Equation (2.2) considers the discrimination of classes from each other pair wise i.e., it considers the overlapping region between the classes.

In the scatter space, it is necessary to consider the separability information of a class with respect to other classes in the transformed space. This separability information of a class is calculated using the pair wise scatter discriminatory information of a class with other classes. By using this concept, the separability information of a class \(i\) is defined as,

$$r_i^\phi = \sum_{j \neq i} w(pd)_{i,j}^{\phi} \quad \text{where } i, j = 1, \ldots, C$$

(2.3)

Thus the above expression shows the separability information $r_i^\phi$ of class \(i\) as the sum of pair wise scatter discriminatory information of class \(i\) with respect to other classes in the transformed space. Moreover it shows the class covariance structure. A larger value of $r_i^\phi$ indicates that class \(i\) has more overlapping with other classes. Moreover, it shows the worse separability of class \(i\).
2.3.2 BOOSTING TECHNIQUE

The KDCV and KCV methods follow the assumption that the training samples are labeled correctly. If there is any mislabel in the training set, these methods fail to classify a given test sample and act like a weak classifier in this case. The recognition performance of these methods is improved by using AdaBoost.M2 boosting algorithm proposed by Freund et al. [54]. The AdaBoost.M2 algorithm is described briefly as follows.

The AdaBoost.M2 algorithm is an extension of AdaBoost to the case where the label space $Y$ is finite. This algorithm requires more communication between the boosting algorithm and the weak learning algorithm. The advantage of this algorithm is that it gives the weak learner more flexibility in making its predictions. Moreover it enables the weak learner to make useful contributions to the accuracy of the final hypothesis. The hypothesis generated by the weak learner measures the degree to which it is believed that $Y$ is the correct label. Moreover, this method assigns a weight according to the discrimination of a sample to the correct label and is called as pseudo loss. Thus the pseudo loss measures the goodness of the weak hypothesis. The weak learner's goal is to minimize the expected pseudo loss for given distribution. By manipulating the distribution on training samples and pseudo loss this algorithm effectively forces the weak learner to focus not only on the hard samples but also on the incorrect class labels that are hardest to eliminate.

The procedure of AdaBoost.M2 algorithm used in this research work is, first assign mislabels distribution for samples in the training set. Then apply the weak learner on the training subset which consists of hardest samples in all classes. This constitutes a weak learner based feature extraction technique. Apply this technique on training set and build the weak hypothesis and measure pseudo loss. Then according to pseudo loss update the mislabel distribution of samples in training set. The final hypothesis output for a given test sample is the label $Y$ that maximizes the responses of weak hypothesis.