CHAPTER 1

INTRODUCTION

In the domain of face recognition, numerous methods to reduce the dimensionality of the subspace in which the faces are presented have been reported. Recently, projection technique has emerged as a powerful method for dimensionality reduction. It represents a computationally simple and efficient method that preserves the structure of the data without significant distortion and has been applied on problems like machine learning and face recognition.

The dimensionality reduction is a process of taking a dataset with increased number of dimensions, and then creating a new dataset with less number of dimensions, while preserving the “structure” of the data. Among various alternate methods to reduce the dimension of facial data, the most widely reported are the principal component analysis (PCA) and the Linear Discriminant analysis (LDA). The projection scheme projects the data onto a random lower-dimensional subspace and yields results comparable to conventional dimensionality reduction methods; such as PCA. Also, projection technique is computationally less expensive.

The human motion analysis is a conventional approach to design and identify the video image sequence as static image sequence for detecting and analyzing human motion in real time. The human motion activities are recognized by a machine to interact intelligently and effortlessly with a human-inhabited environment. This approach is well established for identifying the images, speech, and video samples that are recognized from 2D images.

The motion estimation filter (MEF) is introduced to reduce the invariance in the 2D images. The performance evaluation of human motion system is estimated using the MEF algorithm. This algorithm is designed to track and recognize the visual constraints in the real world videos and with the combination of generative and discriminative models in a filtering framework can improve the normalization, selection and detection of visual attention. The MEF algorithm realizes a deterministic approach to track any 2D-features representable in a real time application.

1.1 PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) is a technique to analyze and dimensionally reduce the data, highlighting their similarities and differences. The task of facial recognition involves discriminating an image data into several classes. The input images are highly noisy (e.g. noise caused by differing lighting conditions, pose etc.), yet the input images are not completely random and in spite of their differences there are repetitive patterns which occur in any input signal (eyes, nose, mouth and the relative distance between these objects). These characteristic features are called eigenfaces.

In this thesis, the face Images are projected into a feature space (“Face Space”) that best encodes the variation among known face images. The face space is defined by the “eigenfaces”, which are the eigenvectors of the set of faces. They can be extracted from the original image data by Principal Component Analysis (PCA). The PCA is a statistical method for transforming an observed multidimensional random vector into its components that are statistically independent. A general approach to the PCA is to first solve the characteristic polynomial equation for all eigenvalue and then find their corresponding eigenvectors to produce principal components (PCs).

1.2 LINEAR DISCRIMINANT ANALYSIS (LDA)

Linear Discriminant Analysis (LDA) is efficient in encoding discriminatory information. LDA involves the grouping of similar classes of data that seeks to find directions along which the classes are best separated. It does so, by taking into consideration the scatter within and between classes. It is also more capable of distinguishing faces even in the presence of variations in images. Both the PCA and LDA are considered as linear space leaning algorithm and focus on the global structure of the Euclidean space. However, both PCA and LDA effectively focus on the Euclidean structure and fail to discover the underlying structure. These methods are appearance or feature based that search for certain global or local representations of a face.

1.3 FACE RECOGNITION

The Face recognition is a visual pattern recognition problem. A face is a three-dimensional object subject to varying illumination, pose, expression and needs to be identified based on its two-dimensional image (or three- dimensional images obtained by laser scan). A face recognition system generally consists of four modules - detection, alignment, feature extraction and matching. Localization and Normalization (face detection and alignment) are processing steps proceeding face recognition (facial feature extraction and matching) is performed.

The Face recognition is a task of automatically identifying or verifying a person from a stored image. An approach for face recognition classifies the human faces by simulating the behavior of human’s “glance” and face “similarity”. A new Emotional Back Propagation (EMBP) learning algorithm is developed based on two essential emotions such as anxiety and confidence. The pattern averaging method is employed for extracting features which mimics the human emotional judgments based on general impressions rather than the precise details of the objects.

The face recognition involves one-to-many matches that compare a query face image against all the template images in the database to determine the identity of the query face. The face detection involves the following criteria for authentication such as:

(i) Appearance-based and learning based approaches

(ii) Preprocessing

(iii) Neural networks and kernel-based methods

(iv) Dealing with head rotations and

(v) Performance evaluation.

1.3.1 Neural Network based Image Recognition

The use of neural networks for image recognition has been developed based on back propagation learning algorithm. A supervised neural network structure can be used to investigate the effect of the added emotional factors on the learning and decision making capabilities of the neural network. The neural network provides a remedy for any lack of uncertainty in the data. It has the advantage of including heuristics judgments into the optimization process.

1.4 FACE DETECTION

Face Detection is a concept that includes many sub-problems. Some systems detect and locate faces at the same time; others first perform a detection routine and then, if positive, try to locate the face. In this research, the face detection algorithms share the following common steps:

(i) Data dimension reduction scheme to achieve an admissible response time.

(ii) Pre-processing to adapt the input image to the algorithm prerequisites.

(iii) Extracting facial features or measurements.

(iv) These will then be weighted, evaluated or compared to decide the presence and location of a face.

(v) Finally, the use of algorithms that have a learning routine and they include scalability to their models.

Face detection is, therefore, a two class problem and can be considered as a simplified face recognition problem. Face recognition has to classify a given face, and there can be as many classes as candidates. Consequently, many face detection methods are very similar to face recognition algorithms. i.e., the techniques used in face detection are often applicable to face recognition. The face alignment seeks to deform a face model to match it with the features of the image of a face by optimizing an appropriate cost function. The new face model is aligned by maximizing a score function. (learnt from training data) and imposed to be concave. Experimental study shows the superiority to other learning paradigms and demonstrates that this model exceeds the alignment performance of the state-of-the-art schemes.

1.4.1 Face Detection for Video

The face detection scheme segments the face areas from the background. In the case of video, the detected faces may need to be tracked using a face tracking component. Face alignment is aimed at achieving more accurate localization and at normalizing faces, whereas face detection provides coarse estimates of the location and scale of each face. Facial components and facial outline are located; based on the location points, the input face image is normalized with respect to geometrical properties, such as size and pose, using geometrical transforms or morphing. The face is further normalized with respect to photometrical properties such as illumination and gray scale. After a face is normalized, feature extraction is performed to provide an effective information that is useful for distinguishing between faces of different persons and stable with respect to the geometrical and photometrical variations.

For face matching, the extracted feature vector of the input face is matched with those of enrolled faces in the database. Whenever a match is found with sufficient confidence an output is obtained. The face attributes (global) considered in this work are:

(i) Pose of the face

(ii) Facial expression

(iii) Presence of added objects and

(iv) Image condition.

1.4.2 DCT One Dimensional and Two Dimensional

The use of Karhumen – Loeve Transform (KLT) has been reported in previous works, as KLT is optimal in terms of compactness of representation. However, KLT requires more pre-processing stages and hence, as an alternative DCT is preferred in this research. Also, DCT closely approximates KLT in the context of information packing.

1.4.2.1 One-Dimensional DCT

The most common DCT definition of a 1-D sequence of length N is

$$C(u) = \propto (u) \sum_{x=0}^{N-1} f(x) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \qquad \ldots\ldots \quad (1.1)$$

for $u = 0,1,2,\ldots,N-1$. Similarly, the inverse transformation is defined as

$$f(x) = \sum_{x=0}^{N-1} \propto (u) c(u) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \qquad \ldots\ldots \quad (1.2)$$

for $x = 0,1,2,\ldots,N-1$. In both equations (1.1) and (1.2) α *(u)* is defined as

$$\propto (u) = \begin{cases} \sqrt{\frac{1}{N}} & for \quad u = 0 \\ \sqrt{\frac{2}{N}} & for \quad u \neq 0 \end{cases} \qquad \ldots\ldots \quad (1.3)$$

It is clear from Equation (1.1) that for

$$u = 0, c(u = 0) = \sqrt{\frac{1}{N}} \sum_{X=0}^{N-1} f(x) \qquad \ldots..$$

(1.4)

Thus, the first transform coefficient is the average value of the sample sequence. In literature, this value is referred to as the DC Coefficient. All other transform coefficients are called the AC Coefficients.

1.4.2.2 Two-Dimensional DCT

The 2-D DCT is a direct extension of the 1-Dimensional DCT and is given by

$$c(u, v) = \alpha(u)\alpha(v) \sum_{X=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right] \ldots (1.5)$$

For u, v = 0, 1, 2,…, N −1 and $\alpha(u)$ and $\alpha(v)$ are defined in (1.3).

$$f(x, y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \alpha(u)\alpha(v) c(u, v) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right] \ldots (1.6)$$

The inverse transform is defined as for x, y = 0, 1, 2…, N −1. The 2-D basis functions can be generated by multiplying the horizontally oriented 1-D basis functions with vertically oriented set of the same functions.

1.4.3 Human Motion Analysis using DCT

Human motion analysis is receiving increasing research attention. This interest is motivated by a wide spectrum of applications. In this research, a process is described for detecting moving targets and extracting boundaries. The skeletonization schemes are suitable for detecting and analyzing human motion in real time. Also, the method requires a minimum of image-based information only and is therefore efficient. Extremal points (like head, hands and legs) are extracted their tracking, based on an n*n block of DCTs (Discrete Cosine Transforms) coefficient is described. Subsequently, the false tracked extremal points such as occluded extremal points are corrected.

Human motion analysis use static image sequences for detecting and analyzing human motion in real time from video image sequence. This research work detects the features such as hands, head and feet for tracking and fit them into some apriori human model. A method for human detection uses an adaptive background model with robustness against

(i) Long –term changes

(ii) illumination changes and

(iii) Repetitive clutter motion in the scene.

There are two main drawbacks of these systems: (i) Completely human specific and (ii) require a great deal of image-based information in order to work effectively. For generic video applications, it may be necessary to derive motion analysis tools that are not constrained to human models, but are applicable to other types of targets, or even to classify targets into different types. In some real video applications, such as outdoor surveillance, it is unlikely that there will be enough “pixels on target” to adequately apply these methods. A robust system can make broad assumptions about target motion from small amounts of image data. Therefore, the requirement is a fast and robust system which can make broad assumptions about target motion from small amounts of image data.

The focus of this work is to investigate the dimensionality reduction offered by polar coordinates and perform an artificial intelligent system based face recognition using back propagation neural network. Experiments show that projecting the data onto a random lower-dimensional subspace yields better results and give an acceptable face recognition rate.

1.4.4 Image Alignment

The image alignment or fitting is essentially an image registration problem, where a face model needs to be deformed to match the image of the face, so that the natural facial features are aligned with the model. The dramatic variations of facial appearance due to shape, pose, illumination, expression, occlusions and image resolution make this a challenging problem. The Active Shape Model (ASM) is one of the early approaches that attempt to fit the data with a model that can deform in ways consistent with a training set.

The Active Appearance Model (AAM) is a popular extension of the ASM. The particular structure of the resulting classifier allows mapping the original problem to a ranking problem, because it implies the learning of a function, i.e. the alignment score function, which can be interpreted as a ranking function (able to order instances corresponding to different degrees of alignment of the Boosted Rank Model (BRM)), and which is meant to be concave.

During the fitting phase, the AAM is aligned in such a way that the data can be reproduced in the least mean square error sense. However, the alignment performance degrades quickly if either the AAM is trained on a large dataset or it is fitted to unseen subjects or both. This problem can be tackled by proposing the Boosted Appearance Model (BAM) which uses a shape representation similar to the AAM, whereas the appearance is given by a set of discriminative features, trained to form a boosted classifier which is able to distinguish between correct and incorrect alignments. In this thesis, a hybrid classifier is trained for pair of images wrapped from different landmarks, informs that which of the two corresponds to a better alignment.

1.5 PROBLEM STATEMENT

Several problems exist in human motion analysis and this includes;

(i) Unable to provide an optimal method of dimensionality reduction to achieve higher recognition rate.

(ii) Human motion tracking cannot be done (from the training set) by combining the eigenfaces alone.

(iii) The weighting factors need to be more adaptive to achieve better results.

(iv) Less scalability exists for detecting the human motion.

(v) Image block matching, gradient constraints, phase conservation or energy models are bottlenecks.

1.5.1 Challenges

(i) Real – time implementation of high frame –rate sequence.

(ii) Models work fine for low motion velocities, they fail when estimate fast motion.

(iii) System resource requirements need to be optimal.

1.5.2 Requirements

(i) Difficult task to evaluate the segmentation quality of the algorithm using real images since pixels are not labelled according to the object they belong to.

(ii) Features moving coherently (with the same speed and direction), being good candidates for a moving rigid body.

1.6 PROBLEM SOLUTION

(i) A proper choice of the projection matrix is essential to produce the optimal dimension of the feature vector of the original data. The projection technique is an optimal method of dimensionality reduction to obtain a higher recognition rate.

(ii) In this proposed work, the Principal Component Analysis (PCA) approach is used to transform the original image of the training set.

(iii) The neural network based image recognition is identified to investigate the effects of the emotional factors of learning and decision making capabilities. The neural methods are very sensitive to the weighting factors that are tuned properly to achieve good results. The novelty of this research work is its lesser dependency on weighting factors.

1.7 OBJECTIVES OF THIS RESEARCH

- Recover 2 –D motion from video sequences.
- Efficiently segment moving objects using a sparse map of features from the visual field.
- Optimizing the number of surveillance cameras that leads to a strong demand for automatic processing methods from their output.

1.8 MOTIVATION

The goal of this thesis has been (i) to come up with new human motion detection scheme that is best suited for real-time implementation. These criteria are mainly required for optimizing the computing power and storage space. In order to accomplish this goal, a preprocessing method that depends on skin segmentation morphological opening and labeling of connected regions is selected. (ii) The edge detection algorithm is made more suited for use in real-time by using novel transformation based image analysis techniques.

The correlation between the values of the different coefficients among different patterns of various classes is used to construct the classification boundaries. The collection of invariants has a significant advantage over other geometrical moments for face detection. In addition, the comparative evaluation of the large set of face detection approaches presented in the literature is carried out.

To summarize, the aim of this thesis is to investigate the performance of a novel face detection system and is developed in three stages; (i) Pre-processing (ii) Feature extraction and (iii) Learning based classification.

Stage 1: Skin segmentation, Morphological opening and region labeling for preprocessing the input image to extract the candidate skin regions that are likely to represent the faces.

Stage 2: Features of the candidate regions are extracted. Several experiments have been done to investigate the performance of feature extraction methods and the edge detection method is used for optimal edge detection.

Stage 3: A self organizing map neural network for classifying the candidate regions as either face or non-face is done. The aim is to achieve a high detection rate, good performance, reduce the computational complexity of the proposed scheme with respect to processing time and memory requirements. To satisfy such criteria, the statistical principal component analysis is used to select a subset of the features space to construct the optimal feature vector.

1.9 METHODOLOGY FOR SOLVING THE PROBLEM

In the broad area of optimization method, the face detection algorithm forms a wide trade-off between local and global searching strategy. This algorithm has a standard solution approach for multi dimensional optimization problems and the face model defines a fitness function and has to set some parameters which present a novel feature selection search procedure. The face detection algorithm utilizes both the local importance of features and overall performance of subsets to search through the feature space for optimal solutions as shown in Table 1.1.

Table 1.1 Face Recognition Systems

Face Image Acquisition

Face Image Enhancement	Grey scale conversion Noise Reduction Edge Detection
Feature Extraction & Feature Selection	
Face Detection Algorithm	

Face detection and recognition has emerged as an active area of research in fields such as security system, videoconferencing and identification. As security deserves prime concern in today’s networked world, face recognition can be used as a preliminary step of personal identity verification, facial expression extraction, gender classification, advanced human and computer interaction. Face recognition is considered to be a complex task due to enormous changes produced on face by illumination, facial expression, size, orientation, accessories on face and aging effects. Usually, face recognition systems accomplish the task through face detection, facial feature extraction and face recognition. Generally, face identification technique can be considered as image based or feature based.

The image based methods uses predefined standard face patterns whereas feature based techniques concentrate on extracted features such as distance between eyes, skin color, eye socket depth etc. The majority of the face alignment model is based on ASM, AAM are incorporates with a generative template model for each landmark, whereas the same model is discriminative. These models greatly improve the efficiency of the image alignment.

The Rank Boost learning is used to provide a relative similarity measure between a given shape and a reference shape. Ranking a fixed number of predefined wrappings of an image and then combine the first few top ranked to perform shape detection is that typically includes representation of mean shape, superimposed image shape and wrapped image.

1.10 PRINCIPLES BEHIND DIMENSIONALITY REDUCTION

The number of bits required to represent the information in data (video) can be minimized by removing the redundancy present in it. There are three types of redundancies:

(i) Spatial redundancy, which is due to the correlation or dependence between neighboring pixel values.

(ii) Spectral redundancy, which is due to the correlation between different color planes or spectral bands.

(iii) Temporal redundancy, which is present because of correlation between different frames in video.

This research aims to reduce the number of bits required to represent video by removing the spatial and spectral redundancies as much as possible. Data redundancy is of central issue in digital video processing. If n_1 and n_2 denote the number of information carrying units in original and compressed video respectively, then the compression ratio CR can be defined as

$$CR= n_1/n_2$$

And relative data redundancy RD of the original video can be defined as

$$RD=1-(1/CR);$$

Three possibilities arise here:

1) If $n_1=n_2$, then CR=1 and hence RD=0 which implies that original video does not contain any redundancy between the pixels.

2) If $n_1>>n_2$, then CR→∞ and hence RD→1 which implies considerable amount of redundancy in the original video.

3) If $n_1<<n_2$, then CR→0 and hence RD→-∞ which indicates that the lower-dimension video contains more data than the original.

1.11 MERITS OF THE PROPOSED WORK

Edge detection is a well developed field on its own within image processing. Region boundaries and edges are closely related with each other. The edge detection is used as the base of another segmentation technique for identifying the edge. This segmentation algorithm is computed more efficiently and implemented using finite scheme.

The background subtraction technique is used for analyzing the motion of human targets and extracts the broad internal motion features of a target. The Gaussian components allow an efficient representation of a large variety of pdf. Thus, Gaussian mixture models (GMM) are commonly employed in image segmentation tasks. Modeling the probability density function (pdf) of pixel attributes with finite mixture models (FMM) is also a natural way to cluster data because it automatically provides a grouping based on the components of the mixture that generated them. The parameters of the FMM model with Gaussian components can be estimated very efficiently through maximum likelihood (ML) estimation using the Expectation – Maximization (EM) algorithm.

It is difficult to give a universal definition of the term 'cluster'. The clustering algorithm for image segmentation has two parameters namely:

(i) Number of clusters and

(ii) Exponential weighting factor over the membership functions.

Clustering methods are used for data exploration and to provide prototypes for use in supervised classifiers. Methods that operate both on dissimilarity matrices and measurements on individuals are described, each implicitly imposing its own structure on the data. Mixtures explicitly model the data structure. Cluster analysis is the grouping of individuals in a population in order to discover structure in the data.

The clustering based image segmentation methods rely on arranging the data into groups having common characteristics. The results of a cluster analysis may produce identifiable structure that can be used to generate hypotheses (to be tested on a separate data set) to account for the observed data.

1.12 BENEFITS OF THE RESEARCH WORK

(i) Perform a comprehensive study on analyzing human activities from short term to long term and from simple to complicated activities in surveillance video activities.

(ii) Detection is another essential task of surveillance video analysis and is used to identify the temporal position in a video.

(iii) An intelligent visual surveillance system based on efficient and robust activity analysis.

1.13 THESIS ORGANIZATION

This thesis is organized as follows:

Chapter 1 Introduction, Objective, Motivation and Problem statement and Solution

Chapter 2 Literature Survey.

Chapter 3 Hardware Architecture of human motion detection.

Chapter 4 Dimensionality reduction in feature extraction.

Chapter 5 Design and Implementation of human motion estimation filter.

Chapter 6 Experimental Results & Discussions.

Chapter 7 Summary and Conclusion.