2. Chapter 2

CHAPTER 2

BRIEF THEORY AND LITERATURE SURVEY

2. BRIEF THEORY AND LITERATURE SURVEY

2.1 Brief Theory

Usage of facial features to verify the identity of a person is one of the key factors in biometrics. Traditional face recognition systems were based on still images. Further, face recognition from video became popular. There are several challenging factors when a face is to be recognised from a video. Few of them can be exemplified as illumination factor, pose variation where videos can contain frames with varying pose which will make the system difficult for recognising face, occlusion when users are not cooperating with the camera which leads to loss of feature elements from a region of the image, blurring of the face as moving from one frame to another etc. Even with the above challenges in the system, video based face recognition is gaining popularity because of the abundant amount of information we get from the video. The spatio temporal information from video gives rise to increase research in this field. This leads to users using this biometric for surveillance purpose as well as for verifying the identity of a person to an access controlled area.

2.1.1 Architecture of Video Based Face Recognition System

Video based face recognition consists of three basic steps namely, (i) Pre-processing, (ii) Feature extraction, (iii) Classification and (iv) Evaluation. Different algorithms or methods can be used at each stage for a better performance.

Figure 2.1 gives the block diagram for a video-based face recognition system. As illustrated in the Figure 2.1, the face is detected and features are extracted from the detected face. The extracted features are classified with respect to the gallery set images and any of the distance measure method is used to find the closest match between the input image and the images stored in the database. Based on the nearest match a conclusion is made whether to accept the input image or to reject it.

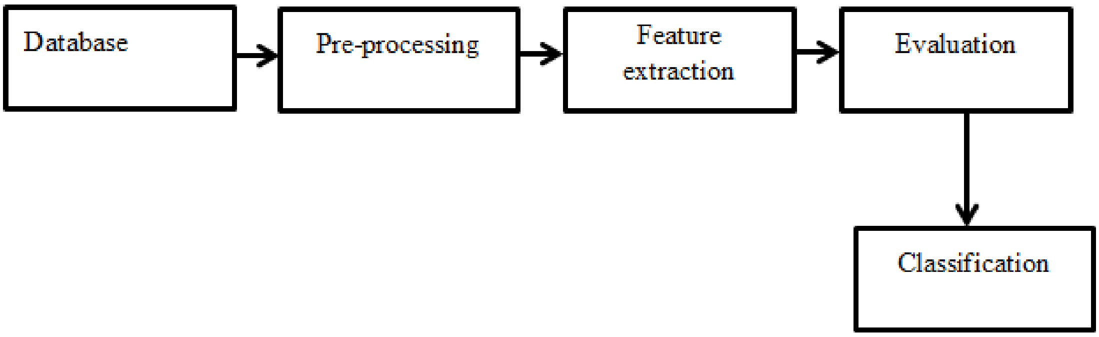

Figure 2.1: Architecture of video based face recognition.

2.1.2 Application of Video Based Face Recognition System

Numerous applications for face acknowledgment have been conceived. (Senior and Bolle 2002) described in their work about various applications of face recognition. Establishments so far are restricted in their capacity to handle posture, age and lighting varieties, yet as advances to handle these impacts are developing by researchers on this area.

Access control is one of the major applications with respect to face recognition. Face confirmation, coordinating a face against a solitary selected model, is well inside of the capacities of current Personal Computer equipment. Since PC cameras have gotten to be boundless, their utilization for face-based PC logon has ended up plausible; however take-up is by all accounts extremely restricted. Expanded convenience over secret key assurance is difficult to contend with today's to some degree untrustworthy and capricious frameworks, and for a couple of areas arrives inspiration to advance past the mixes of the password and physical security that ensure most venture computers.

Surveillance is another application with respect to video based face recognition. In this case video is the medium of choice as video provide a rich amount of information. Face recognition is considered to be the best biometric for video data. The advantage with face recognition is that user’s knowledge or participation is not required with this biometric.

2.1.3 Advantages and Disadvantages

The major advantage with video-based face recognition is that videos provide rich data information with respect to a subject in the sequence of frames and subsequently it gives more data for recognition. Compared to other biometric where client intercession is required with respect to the machine, in the case of video-based face recognition, subject’s knowledge or participation is not required with the framework.

Even though face recognition from video as a biometric has advantages over other biometrics, it suffers from various disadvantages that deteriorate the recognition rate. As subjects need not cooperate with the system in this biometric, a frame may consist of faces that are oriented at different angles, with different expressions, with varying illuminations and sometimes it is also possible that the face region is covered with accessories that make recognition a difficult task.

2.2 Literature Survey

2.2.1 Literature Survey on Video Based Face Recognition

Face recognition started with still faced images. The earlier studies in face recognition concentrated on the frontal view of the face. To recognize the face various methods were followed that are basically classified into feature based, holistic and neural network based approaches. Further, the study in face recognition shifted from still based images to images from a video sequence. From past numerous years, there has been a lot of work going ahead in expanding the recognition rate of faces in video sequences. Some of the works are discussed below.

(Xiaoming and Chen 2003) proposed an adaptive HMM that is used for recognising a face from the video sequence. HMM is used during training and testing procedure to learn the temporal dynamics of the faces. In this method, the feature vectors are reduced using PCA. The Eigen vectors are obtained while applying PCA. During recognition stage, feature vectors are obtained by projecting all the frames to Eigen space. This concept is used in the recognition process where once a test sequence is recognized as a subject, current sequence of HMM is used to update once measuring the confidence of the recognition rate. To test the efficiency of the proposed method, authors used Task database and MoBo database. Authors claim that they produced a better recognition rate with the proposed method.

Temporal continuity in video helps to increase the recognition rate in the video sequence. In the work carried out by (Hadid and Pietikainen 2004) authors represent the faces from a video for recognition purpose. In this approach, authors compared the probe and gallery video to extract the exemplars. These exemplars are used as the training sample for appearance-based face recognition algorithm. The probabilistic voting strategy is used to identify the person in the video. In this work, authors used temporal HMM (Hidden Markov Model) for spatio temporal analysis of the video and PCA (Principle Component Analysis), LDA (Linear Discriminant Analysis) are used for still image based analysis. Experiments are performed on two databases MoBo and Honda/UCSD on both the above cases. Authors conclude that spatio temporal information did not use the spatial cue over temporal information. The result from the experiments shows that HMM gave a better recognition rate compared to PCA and LDA methods.

(Aggarwal et al. 2004) used Autoregressive and Moving Average (ARMS) model in his work on face recognition. ARMA model can take care of the change in appearance while modelling pose and expression. This experiment crops out the faces from each of the frames from the video as a pre-processing step. To track the face, nose tip is used as a point and KL tracker is used for tracking the nose tip. Edge based rough pose estimator is used to understand the pose estimation of the faces. Recognition of face by computing distance between the gallery and probe faces in the video is carried using the concept of subspace angles. Authors in this experiment used UCSD/Honda database. The result shows that the recognition performance is more than 90%.

(Stallkamp et al. 2007) developed a model to identify a face from real world data. They incorporated three weighting schemes DTM (Distance to Model), DT2ND (Distance to Second Closest) and DTM+DT2ND. DTM gives information on the similarity of the test sample to the training set image. If they are less similar then it will produce larger distances. DT2ND reduces the ambiguity in the classification. The combination of the two weighted schemes enhances the benefits of both the schemes. The result shows that this method could identify faces that are subjected to illumination, pose, expression and occlusion.

In the work carried out by (Connolly et al. 2012), a face recognition algorithm with an adaptive classification system (ACS) is proposed. This is a combination of ARTMAP (Adaptive Resonance Theory) neural network classifier, DPSO algorithm (Dynamic particle swarm optimization) and LTM (Long Term Memory) algorithm. The experimental result shows a higher classification rate compared when fuzzy ARTMAP used alone. It also shows that ACS parameters require more resources.

In video-based face recognition when face images are captured, they are affected by varied pose, illumination, blur. (Wong et al. 2011) proposed a method in which an ideal face is represented through probabilistic face modelling. The proposed algorithm started with pixel based image normalization. The normalized image is divided into patches and each patch is normalized to have zero mean and unit variance. From each such patch DCT (Discrete Cosine Transform) feature vector is extracted. Further, for each block probability of corresponding feature vector is calculated with the use of location-specific probabilistic model. Probability similarity score, Q (I) is calculated using the formulae,

$$\text{Q (I)}= \sum_{i=1}^{N} \log p(X_i|\mu_{i,}\textstyle\sum_i) \qquad (2.1)$$

Where I represent the image, N is the patches; x_i represents the feature vector, μ_i and $\sum_i$ represents the mean and covariance matrix of a normal distribution respectively. This probabilistic score will give the similarity of the face with the ideal face. Authors conducted the experiments on FERET and PIE datasets. The results show that the proposed method is capable of handling faces which are simultaneously affected by issues like pose variation, shadowing, blurred images because of motion and errors caused due to alignment.

Multiple face features are fused by (Choi et al. 2012) to recognise face. In this approach, weights of face features are computed. The experimental result says that the proposed method works well when the number of misaligned images is low in the dataset. (Majumdar and Ward 2012) proposed a work based on sparse classification approach. This method assumes that each test sample of a face image can be formed by a linear combination of training sample. MMV (Multiple Measurement Vectors) are used for classification. A modified iterating hard thresholding algorithm is proposed in this method to solve the optimization problem. Authors in this approach use VidTIMIT database. The comparative study is made on HMM and sparse classification approach. The result shows a better performance compared to the above two methods.

While recognizing a face from a sequence of video, selection of a good frame that contains the face is an important attribute. This will enhance the performance of the recognition rate. (Anantharajah et al. 2012) used different features on the face to normalize the face image. Haar feature based classifiers are used for detecting the eyes. Based on the detection of the eye position, the face is normalized. The following image features are extracted namely, the closeness of the mouth, sharpness, brightness, openness of the eyes, contrasts and face symmetry to prove the quality of an image for recognising. The experiment is performed on the Honda/UCSD video database and it is proved that selecting a good quality frame increases the recognition rate from the video.

(Chen et al. 2012) proposed a dictionary learning approach which can handle changes in illumination and pose. In this method, face images are cropped from the video and frames with the same pose and illumination are grouped into one partition. Sub dictionaries are generated and further combined to form sequence-specific dictionary. Further, the sequences are matched for similarity. Experiments are conducted on Multiple

Biometric Grand Challenge (MBGC), Face and ocular Challenge Series (FOCS) and Honda/UCSD datasets. The result shows the better performance of the algorithm when compared with the existing methods.

(Chen et al. 2013) proposed an approach for video to video face recognition. Face images extracted from a given video is partitioned where each partition contains faces with different pose and illumination conditions and form a dictionary of images. On testing, the same partition is used for the test video. Joint sparse representation makes the recognition of the matching face with minimum error rate. Authors claim that the proposed approach works better than many state-of-the-art face recognition algorithms.

(Rosales et al. 2013) in their proposed work uses Template Based Cross Correlation (TBCC) method. Templates are generated by extracting facial region and applying Optimal Adaptive Correlation (OAC) a normalized face image is obtained. The normalized face forms the template matrices of each subject. The test samples were resized to that of the training samples. TBCC is applied on all the testing samples to get the correlation value and to know the probability of test sample being same as the training sample. In this experiment RML and CK database are used for evaluation. The results show that the proposed method could recognise a face within a short period of time.

Video data consists of frames that vary in terms of pose, illumination and expression. In a method that is proposed by (Gong et al. 2013) a low dimensional subspace for each individual is constructed and key frames are extracted from each video sequence. All the key frames are further filtered by Difference of Gaussian (DoG) filters and changed to Local

Binary Patterns (LBP) representations. The experimental result shows a better performance over existing methods.

Recognising illumination and pose variant face from a video has been a challenging task. (Arandjelović and Roberto 2013) developed a photometric model combined with statistical model to solve faces subjected to illumination and used local appearance manifold structure and identity likelihood to achieve recognition of faces with head pose.

(Lu et al. 2013) used a discriminative multi-manifold analysis (DMMA) method by learning discriminative features from image patches. To achieve this, the authors partition the patches to form an image set to which manifold matching is applied. Reconstruction of face is performed.

(Beveridge et al. 2013) used a point to shoot approach in recognising face. Baseline algorithms LDA and LRPCA are used for this approach.

Videos provide more information compared to still faces. Considering this extra information provided by the video in face recognition (Bhatt et al. 2014) proposed an approach where in video signatures are generated by combining information available across frames of a video. The proposed algorithm generates ranked list for every frame in the video. Ranked lists across frames are optimized using clustering based re-ranking and fusion. The video signatures are compared using DCG measure which in turn uses ranking and relevance of images. Experiments are carried out on YouTube and MBGC V2 video databases. The results show that the proposed method outperforms the existing algorithms.

(Taigman et al. 2014) worked with the alignment and representation step of face recognition using neural network. Authors used a LFW dataset for analysing the result.

(Ding et al. 2015) applied the sequential scheme to streaming video data. Authors use Sequential Sample Consensus (SSC) for sequential sampling and updating. At each of the iterations, the face image from a frame in the video is compared against the samples from the training set. A temporary PMF is computed based on the evaluation of the consensus of the face image with the drawn samples. The priority PMF is updated for a newly detected face in the video. Authors for this experiment used Honda/UCSD and MoBo databases. The limitation of this algorithm is the difficulty in recognising faces when there is a large variation in poses. This algorithm is advantageous in terms of rate and time of recognition.

(Florian et al. 2015) presents a system FaceNet that maps face image to a Euclidean space where the similarity is checked with the distance. Authors claimed to have a recognition rate of 95.12% with you tube DB and 99.63 with LFW dataset.

From the above survey conducted on face recognition the major problems that deteriorates the recognition rate are found to be due to variation in pose and occlusion. There are studies going on even today applying various methods to solve this problem.

2.2.2 Literature Survey on Face Detection Algorithms

(Yang and Ahuja 1998), proposed a method using multiscale segmentation, colour and geometric information. Multiscale segmentation is used in this method to build up homogeneous regions that can deal with scale problems. Further, skin region is extracted using skin colour model.

These extracted skin regions are merged until the shape is approximately elliptical. Human face is assumed to be elliptical. The results based on this experiment show that the proposed method is able to detect human faces in colour images with different sizes, orientation and viewpoint.

A survey conducted on face detection algorithms by (Hjelmåsa and Low 2001) clearly states the different algorithms used in detecting a face. There are many methods to detect faces based on motion, colour and other universal information. Edges are one of the important features where head outline are traced to detect faces. Edge features within head outline are later analysed using shape and position information of the face. Edge detection techniques are also useful in detecting glasses in a face image. Gray information within face can also be used as features for detection. In this case, the property of facial features such as eyebrows, pupils and lips being darker than the surrounding regions help in detecting the face. Colour is another feature that helps in detecting a face. Compared to gray feature, colour is an important feature that is sensitive to the appearance of an object. One of the widely used colour models is the RGB representation. In this approach, colours are defined by the combination of Red, Green and Blue primary colours. The variation in skin appearance is large because of the luminance property even when faces are of different races, RGB colours are generally preferred to detect the face so that the effect of luminance can be filtered out. The normalised colours can be derived from the original RGB components as

$$r = \frac{R}{R+G+B} \tag{2.2}$$

$$g = \frac{G}{R+G+B} \tag{2.3}$$

$$b = \frac{B}{R+G+B} \tag{2.4}$$

By comparing the colour information of the pixel with respect to r and g values of the pixel, the possibility of the pixel belonging to the flesh region can be calculated. YIQ colour model can be used to detect skin by converting the RGB colours to YIQ representation. HSV and YCrCb are also effectively used in detecting the face. Colour segmentation is performed using skin colour thresholds where skin colour can be represented using histograms. Motion is another important feature that helps in locating moving object in a video. Frame difference analysis helps in achieving motion segmentation. Another approach is using moving image contours. Spatio temporal Gaussian filters can be used to detect moving boundaries of faces. The above low-level analysis methods alone is little confusing, a face can be detected by using feature analysis. One of the categories of feature analysis is feature searching and second constellations using face models. The second category of approach in face detection is image based approach. Most of the image based approaches use window scanning technique to detect a face. This algorithm is a method where input image is searched for possible face locations. Size of the scanning window, number of iterations may vary from algorithm to algorithm. This can be classified into (i) linear subspace (ii) neural network approach and (iii) statistical approach. In the case of linear subspace, principal component analysis (PCA), linear discriminant analysis (LDA), and factor analysis (FA) are some of the methods that can be used to detect a face.

From the survey conducted it is clear that Viola Jones face detection is the most used algorithm for real time application of face detection. This prompted to use the above method for detection of faces in the proposed work.

2.2.3 Literature Survey on Feature Extraction Algorithms

In order to find the efficient algorithm for feature extraction, we had to refer to the existing algorithms and study on the efficiency of each algorithm. Principle Component Analysis (PCA) algorithm is used in feature extraction in the work carried out by (Lee et al. 2003) to recognize the face. The basic principle that is followed by PCA algorithm is to convert the two-dimensional images to one-dimensional vector. This vector is decomposed into orthogonal principle components. Those features of the image that vary most from the rest of the image are selected for matching. We found that this method is sensitive to scale variation. Elastic Bunch Graph Matching discussed by (Bolme 2003) is a feature extraction algorithm that places small blocks of numbers over small areas of the image, multiplying and adding the blocks with the pixel values to produce numbers at various locations on the image. These locations are further adjusted to accommodate minor variations. The study on face recognition using this method proved that the recognition rate increased under pose and expression compared to PCA and LDA. Independent Component Analysis minimizes both second order and higher order dependencies in the input data and attempts to find the basis along which the data are statistically independent. This method is found to be not robust to partial occlusion. Linear Discriminant Analysis (LDA) finds the underlying vectors in the facial feature space that would maximize the variance between classes and minimize the variance within class (He et al. 2005). This algorithm suffers in recognition rate when the variation within a class is large. Local Binary Pattern (LBP) is another feature extraction method where the face image is divided into a grid of small non-overlapping regions where a histogram of the LBP for each region is constructed (Shan et al. 2009). The similarity of two images is then computed by summing the similarity of histograms from corresponding regions. This algorithm, the face recognition decreases when affected by illumination. Local Nonnegative Matrix Factorization (LNMF) is an extraction method where the face is broken down into components. This method works better when compared to Principle Component Analysis (PCA). Selective LNMF(S-LNMF) is a two stage process wherein LNMF is applied for all the regions and further selection is made. This method proved to be better when compared with LNMF for occluded regions. Discrete Cosine Transform (DCT) is an image compression technique where pixel intensity values of an image are expressed in terms of sum of cosine functions oscillating at different frequencies. Using DCT technique spatial domain image is converted into the frequency domain and reconstructed applying inverse DCT technique. The result proved to be of better recognition rate when compared to techniques like PCA. In the case of Discrete Wavelet Transform (DWT), the coefficients are described by a wavelet function. Wavelets have finite duration both in time and frequency. Wavelets do not exploit the regularity on edges. The features that are extracted using wavelets show better results. But they are weak in recognizing partially occluded face. (Candes and Donoho 2000) described Discrete Curvelets as well suited for representation of geometric feature like the facial structure. This approach is basically designed to represent edges and other singularities along the curves efficiently. As face consists of curves along the edges this method is found to act at a better rate in recognition (Donoho and Duncan 2000). (Mansoor 2005) describes in his work on compressing image data using curvelets. The author makes a comparison with wavelets to prove that curvelets are a better approach for image compression.

From the above analysis, it is clear that discrete curvelet works better in recognizing faces. The major reason being they are good at recognizing in curved areas and faces containing curved areas.

2.2.4 Literature Survey on Pose Variant Face Recognition

(Lee et al. 2003) in their work proposed a model that could recognize faces with a different pose from a video. Each of training set face is stored as a low dimensional appearance manifold in the ambient image space. Each of the pose manifolds is approximated by an affine plane. K-Means algorithm is used in this work to cluster the exemplars. Each of the clusters is represented as a plane that is constructed through Principle Component Analysis (PCA). The likelihood of information of an image comes from particular pose manifold and the transition to this frame from the previous frame. The experimental result shows that the proposed method is effective in recognizing faces from videos where faces are subjected to various head motion.

(Hu et al. 2004) created 3D face that is generated from a frontal face image. This is generated with neutral expression and normal illumination on the face image. With the 3D face image, realistic virtual faces are generated and used in recognition. This framework gives advantages when compared with the existing face recognition methods because in this proposed framework, only one frontal face is required for face recognition. The virtual faces that are synthesised help in recognition stage of the face. Author claims that the experimental result prove to be better in the accuracy of recognition.

(Zhou et al. 2005) defends that face recognition under pose variation is identified as a challenging problem and the change in visual appearance of the face with respect to a face is varied. The existing algorithms does not provide satisfactory recognition rate when there is a variation in pose. This is a great challenge to face recognition algorithms. In the proposed approach, authors used an image based method for face recognition across different poses. Even though the work is done with given set of poses, generalization on poses is not implemented in this work. Pose variation is handled using light field in this method. In this approach, faces sampled from a specific set of poses are only handled whereas the future work the authors are working on to handle faces with arbitrary poses.

(Jiang et al. 2005) in their proposed framework provided a solution to the problem of recognising faces from different poses. The authors followed the approach of reconstructing faces from 2D to 3D. Frontal face is used to locate the feature points. The feature points considered are the contour points of the face, eyes, nose and the mouth. After extracting the feature points, a 3D face is constructed depending on the feature points. A 2D face image is projected onto the 3D face shape using texture mapping. Samples are synthesized based on this 3D model. Further face recognition is carried out on the varied sample images. The experiments show an increased accuracy in the recognition rate. According to this model, 3D face is constructed from frontal face images only which authors claim to enhance by using pose variant faces in their future work.

(Chai et al. 2007) used LLR (Locally Linear Regression) method to generate the virtual front view of a face from a non-frontal face image. A linear mapping between the frontal face image and its non-frontal part is an assumption made in this study. A regression based solution is provided formulating the linear mapping. LLR is used to predict the accuracy. LLR is applied to each patch in the face. Virtual frontal view is created through the combination of all patches. The proposed method shows that it can recognise faces with different poses.

(Dai et al. 2009) proposed an algorithm where in during the training stage, a single frontal image is the input to the recognition system. From this frontal face, 3D face model is constructed. The resultant 3D face model is rotated to generate the faces with different poses. In the recognition stage, the video sequence finds the virtual face and self-PCA is performed. Experimental results show a good recognition rate of faces with different poses.

(Wright and Hua 2009) proposed a multi scale local descriptor based face representation is proposed in this work. The proposed system takes an input image that contains a face. Face in the image is detected using Viola-Jones face detector. Eyes are detected from the face and are used to perform geometric rectification. Smooth variation due to illumination is removed using photometric rectification. Features are computed from the resultant image. These features with the location fed onto a randomized decision trees. The final representation is a histogram and distance metric is used for the task of recognition.

(Zang and Gayo 2009) carried an extensive review on pose variant face recognition techniques. According to the survey conducted in face recognition, pose variation is identified as one of the major problem in recognition rate of faces as faces are captured in controlled situations. Many face recognition techniques are limited to controlled environment and are unrealistic when applied to real world scenarios. To solve this problem, many methods like 3D morphable model, Eigen light field, illumination cone model etc... Where used still none of the above are free from limitation of recognition. In this review, authors have categorised the face recognition technique in to three, namely general algorithms, recognition using 2D techniques and recognition using 3D techniques. Authors' analysis is that a small orientation of faces affect if holistic approach is applied on faces for recognition and with large rotation of faces, these approaches gave a smaller recognition rate. In 2D techniques, multiple images of a face in different orientation or a panoramic view of the faces were collected to perform recognition on different poses. An alternative to collecting multiple views of face was replaced by synthesizing virtual views of faces.

(Li et al. 2013), approached the problem of face variation in face recognition is through probabilistic elastic matching method. Local features are extracted from image patches. Features and locations are augmented and Gaussian Mixture Model (GMM) is trained that captures spatial appearance distribution of all the face images in the training samples. For matching and recognition, SVM is used. Further, the authors in this work proposed a joint Bayesian adaptation algorithm to adapt GMM improve the recognition rate with respect to different poses. The result shows that this method outperforms the state-of-the-art when experimented on Labelled Face in the Wild (LIW) and the YouTube video databases.

(Yi et al. 2013) propose an algorithm with 3D sensor for recognising faces under special conditions. It fills holes with respect to the depth data. A dictionary is created by learning from training data. Gabor filters are used for feature extraction. Experiment is conducted on FERET and PIE datasets. Experiment claimed to have a recognition rate of 96.7%.

(Lu et al. 2013) used a discriminative multi-manifold analysis (DMMA) method by learning discriminative features from image patches. To achieve this, the authors partition the patches to form an image set to which manifold matching is applied.

(Liao et al. 2013) developed an align-free approach based on multi key point descriptors. Gabor ternary pattern is developed for robust face recognition. Experiment is analysed for databases like AR, FRGCv2.0, LFW and pubFig. Analysis of the result is made on basic algorithms PCA, LDA and LBP.

2.2.5 Literature Survey on Face Recognition on Occluded Faces

(Martínez 2002) to solve the challenge of face recognition, occlusion the face is divided into k local regions and analysed. Gaussian distribution is used to model the error. Mean feature vector and covariance matrix is calculated for each region. Probability of a given match is calculated as sum of all Mahalanobis distances. AR face database is used for experimentation.

(Lin et al. 2007) developed a model that is based on the work done by Viola and Jones for detecting face in an image. To detect frontal faces and occluded faces, a main cascade and eight occlusion cascades are built. The rectangular feature computes the difference between sums of pixel intensity in adjacent regions. Bhattacharya coefficient is computed between the weighted positive and negative histograms. The proposed work is tested on MIT-CBCL and AR databases. The method is found to be effective in detecting occluded faces.

(Zhang et al. 2007) proposed a method based on Kullback-Leiber Divergence – Local Gabor binary patterns (KLD-LGBP). KLD features applied between LGBP features of the local region and that of the non-occluded local region to understand the probability of occlusion. This probability is considered as the weight of the local region during matching phase. Experiments are conducted on AR database and the result shows the effectiveness of the above method.

Facial occlusion is one of the main challenges of face recognition. (Rama et al. 2008) detected and recognised occluded faces using PCA, LPCA and CPCA. These three strategies are compared and checked for partially occluded faces.

(Han et al. 2008) proposed a model where people in motion are segmented using combination of Gaussian Mixture model (GMM) and Dynamic Markov Random Fields (DMRF) algorithms. DMRF based object segmentation is done on these feature to extract a better shape. In the video sequence, tracking of the moving bodies is performed using Mean Shift Algorithm. The occlusion problem in video is addressed using horizontal histograms of the human body shapes using non- linear regression algorithm. This particular model helps in automatically locate the human body during occlusion. Silhoutte-related image features are used in this method except that natural boundaries are not used from the body shape received. Projection histograms so obtained are modelled by non-linear regression algorithm. The regression technique generates the relative maximum points on the smooth curve, this relative maximum points corresponds to one visual person in the occlusion. Experimental result of the proposed model in this work shows that the performance of the proposed approach is equal to that of particle filter approach but the computing speed is much better compared to the particle filter approach.

Face recognition from video is an emerging research field with face occlusion as a major challenge. To solve this problem of face occlusion and increase the face recognition rate (Hu et al. 2009) proposed a novel method to recognize face based on face patches. To carry out this experiment, as the first stage face patches are cropped from video frame by frame. These face patches are matched to an overall face model and stitched together. The face is reconstructed by accumulating patches and the reconstructed face is used for recognition. Face recognition is done using sparse representation of the reconstructed face. The experimental result shows that the recognition rate is 81%.

(Zhou et al. 2009) in their work integrated Markov Random Field for spatial continuity and sparse representation of the test image with respect to the training set of images. The proposed work is claimed to identify the occluded region and excludes those region from sparse representation.

(Chiang and Chen 2011) proposed a novel approach in recognising partially occluded and damaged face image. As a first step in the proposed method, the face is normalised. Every face is reshaped to uniform face shape to ignore the face variation if any. In this approach authors used PCA algorithm to build a person specific Eigen space for each identity. PCA is used as it can retain the statistical personal characters of an image. On the normalised face define a set of landmarks that corresponds to set facial points. In this approach AAM is used for locating the landmarks. Authors proposed a face recovery method as an iterative process that incrementally restores the face textures that are been lost. In this method once the system enters the iterative state, the face textures are reconstructed and unreliable pixels are identified. Further lost textures are restored. Reconstructed face structure is compared with the restored face structure pixels that are occluded is identified. The texture of the occluded pixels is restored with the help of the reliable pixels. The experimental result shows a better performance.

(Bindu and RaviKumar 2011), proposed a method for recovering lost region using inpainting. In this method authors concentrated on the recovery of face if the occlusion is on the boundary of the face region. The authors proposed a method where image is captured and face regions are detected and further localised using skin illumination compression algorithm. Aki-Toussaint heuristic approach is applied as a pre-processing step. The bounds of the face region are determined using convex hull. With the determined bound of the face, localized and detected face regions are overlapped. Finally the face is reconstructed using inpainting technique. Authors claim that the result is found to be positive.

(Rui et al. 2011) worked on improving recognition rate of faces that are occluded by facial accessories. In this work, authors have considered sunglasses and scarf as the facial accessories. They developed a method where the presence of sunglass and scarf are detected using Gabor wavelets, Principle Component Analysis (PCA) and Support vector Machines (SVM). Once face with occlusion of accessories is detected, the recognition is performed from the non-occluded region. This is done using block-based local binary patterns. The authors used AR database for experimentation and the proposed method shows significant improvement in the recognition when compared with the existing methods on recognising faces with partial occlusion.

(Eishita et al. 2012) proposed a method for tracking objects in a video when they are occluded. Shape feature combined with color and texture gives extra information on the object to be tracked. This work proposes a model to track objects after occlusion. The algorithm proposed here works with the shape information of an object when there is an occlusion affecting the moving object.

(Li et al. 2013) proposed a morphological graph model that describes the morphological structure of the occlusion. Incorporating the errors in occluded part and non-occluded part, authors proposed structured sparse error coding for face recognition from occlusion. The experiment proved to be better for high level occlusion when compared with existing methods.

(Yang et al. 2013) used Gabor feature based robust representation and classification (GRRC) scheme to recognise face. Gabor feature is used for computing Gabor occlusion dictionary. Experimental results show a high accuracy rate in the proposed method.

(Alyuz et al. 2013) developed a 3-D face recognition system that is robust to occlusions. Missing data is handled using subspace analysis technique. Non occluded patches are utilised for construction. For classification, a masked projection is proposed that use the subspace analysis technique.

(Zhao et al. 2014) proposed a method that is based on the idea of sparse representation. Sparse representation is used as this forms minimum reconstruction error from the target template. In this work, authors propose sparse representation to detect occlusion during the tracking of the video sequence. Here the face is divided into patches. Spatial distribution of the samples that are employed by the sparse representation is analysed in this technique, which determines the occlusion rate in each of the patches that is divided in the target face. The experimentation results prove that the proposed algorithm performs better when compared to state-of-the-art methods.

(Zhang et al. 2014) proposed a part based visual tracking system to detect occlusion. In the proposed work, a part matching tracker is established that has properties to explore the spatial-temporal locality constraint property for part matching. During occlusion, the part of the object that is visible gives reliable information for tracking. Part matching approach is used to get confidence scores of individual parts. The approach followed here matches local parts from frames by considering information like low rank and sparse structure. The experiments conducted based on this approach shows an increase in performance.

(Su et al. 2015) combined the information of original image ad reconstructed image for face recognition. Authors also concluded in this work that the recognition rate is comparatively better while non-occluded region of a face is considered when compared with the reconstructed face.

Literature review conducted on occlusion based face recognition gives a clear idea that none of the approach gives a 100% recognition rate with occluded faces.

2.2.6 Literature Survey on Inpainting Algorithms

Texture synthesis is a technique of growing a new image outward from an initial seed one pixel at a time. One of the applications of texture synthesis is occlusion fill in. (Efros and Leung 1999) proposed a non-parametric method for texture synthesis. In this method, texture is assumed as a Markov random field. A square window around the pixel is considered as the neighbourhood of a pixel. The sample image is queried and similar neighbourhoods are found. Even though this approach gives good result for wide range of textures, sometimes there is a tendency of the textures to move to wrong part and accumulate garbage. This problem arises when the sample contains too many Texel or Texel with different illuminations.

(Bertalmio et al. 2000) introduced an algorithm to inpaint still images. At the beginning, the whole of original image is subjected to diffusion smoothing to reduce noise followed by inpainting procedure. The algorithm fills the region with surrounding information. In this approach, from the contours of the missing region, the filling is prolonged inside. The information is propagated from outside the missing region inwardly.

(Oliveira et al. 2001) proposed an algorithm that is designed to inpaint a small occluded region. It involves a procedure of diffusing the pixel information from the boundary of the occluded region inwards. As with different iterations, the inpainting procedure progresses from the boundary of the occluded region to the interior of the occluded region. The experimental result shows fast filling of the occluded region.

(Drori et al. 2003) designed a method for image filling by fragments. The image is completed by a composition of fragments under spatial transformation. The visible part of the image helps to infer the unknown parts. An image fragment at each step is selected from the similar and frequent area. The experimental result shows a good rate in filling the missing region if rich fragments are available. This method fails if the occluded region is the boundary of an image as well as in case of intersection of two images.

(Loke and Ranganath 2009) in their work used a guidance vector field from training data. Missing region is recovered by solving poisson equation using learned guidance vector field along with dirichlet boundary conditions. The best training set is selected by finding the correlation between the neighbouring patches of the input image and training image. A reconstructed gradient of region of interest is calculated. This is used as the guidance vector field for inpainting the missing region. This is the input to the PCA model proposed and the region is reconstructed. The result shows that the computational time is reduced and missing regions are better represented which leads to good inpainting result.

(Mahajan and Vaidya 2012) conducted a detailed study on different inpainting techniques. According to the survey conducted by them diffusion based inpainting is the first approach to digital inpainting. In this method image information from the known region is filled in to the unknown region at pixel level. This algorithm is based on the theory partial differential equation (PDE). This is found to be good method for filling in small area. The disadvantage with respect to this technique is that this creates a blur which is noticeable while filling in large regions. The second category according to the survey is exemplar based inpainting. This method iteratively synthesises the missing region by using the most similar patches from the known region. This algorithm is found to be good for large missing regions. Authors classified the different approaches of inpainting as (i) texture synthesis based (ii) PDE based (iii) exemplar based (iv) hybrid inpainting (v) semi-automatic and fast inpainting. In the case with texture synthesis, new image pixels are synthesised from initial seed. The objective of this method is to generate patterns of textures that is similar to a sample pattern. This is done so that the newly produced texture will retain the statistical properties of the root texture. There are three categories: (i) statistical (parametric), pixel-based and patch-based (non-parametric). Partial Differential Equation (PDE) is an iterative algorithm. This method is based on continuing the geometric and photometric information that are available at the border of the missing area into the area. This algorithm produces good result if the missing area is a small one. Exemplar based inpainting is proved to be very effective method of filling in the missing region. This approach will copy the section of best matching patch from the known area and with certain metrics the similarity is measured and copied into the target area. This is proved to be an effective method to fill in a large missing area. The hybrid approach discussed in this paper puts forward the idea that it is a combination of texture synthesis and PDE based inpainting for filling in the missing region. In this case the region is broken down into two parts, structure region and texture region. These parts are filled with edge propagating and texture synthesis algorithms. Semi-automatic inpainting technique requires user assistance to help in structure completion. In this case user manually decides the object boundary from the known region. This is followed by patch based texture synthesis to generate that area.

(Chung et al. 2013) used particle swarm optimisation to inpaint the higher priority region in a missing image. (Hassan et al. 2013) used mean shift algorithm for image segmentation. Further classification is performed. The segmented images are subjected to exemplar inpainting algorithm. (Bhatewara et al. 2013) segmented the foreground and background of each frame in the video. The inpainting algorithm fills in the missing regions in the foreground. Further the background from the background model is copied and remaining holes are filled. (Muddala et al. 2014) used local depth information to classify the foreground and background. The inpainting is performed only on the background data as foreground information is excluded. (Patil and Deokate 2015) used wavelet transform for restoring the lost image as a process of inpainting. Authors claim that wavelets show a significant performance in the process of restoring the missing region in an image. (Cai and Kim 2015) worked with non-hybrid image inpainting technique. The authors have enhanced the exemplar based inpainting keeping in mind the idea of taking the minimal sub image to be refined. (Mugrey et al. 2015) uses a clustering method based on locally linear embedding (LLE). With the clustered patches, guided inpainting algorithm is developed based on exemplar inpainting.

2.2.7 Literature survey on reconstructing occluded faces using inpainting

(Mo et al. 2004) used sequence of weak similarity measures to select the nearest neighbours. The similarity criteria used are: (i) skin texture (ii) location of eye, nose and mouth (iii) the size of eye, nose and mouth. (Zhuang et al. 2009) worked on repairing the occluded or facial images with the introduction of inpainting technique. Exemplar based image inpainting algorithm is used to fill in the missing region of a face. In this approach the face with occlusion is searched for a similar match using the edge features of the occlusion area and the face area. Now the occluded area is patch inpainted. To get a similar face feature extraction is followed by distance measuring is performed. Exemplar based inpainting is used for filling in the missing region followed by patch guidance based on the edge and feature points of the face.

(Bindu and RaviKumar 2013) proposed a method for inpainting heavily occluded face regions. The assumption they have is the input image has only 10% of visibility where only the eyes are visible in a face. As a first step, skin illumination compensation algorithm is used to localise the face region. Upright-Speeded Up Robust Features (U-SURF) is used to extract features. The advantage of using U-SURF is that they are invariant to different scale in the image. The width of one eye that is visible is repeated five times horizontally along the eye line. The face width obtained and ground truth features are compared to the subjects in the database. The experimental result shows a recognition rate of 98%.

2.2.8 Literature Survey on Exemplar Inpainting Technique

(Drori et al. 2003) synthesized missing image from an original image. In this work, the occluded region is iteratively by image fragments using the principle of self-similarity. The fragment of image is selected from most similar set of fragments. The results show a good performance in filling but the computational time is long compared to existing methods.

Exemplar based image inpainting works better when compared to PDE and texture synthesis methods (Criminisi et al. 2004). They proposed an algorithm that combines the advantages of PDE and texture synthesis to inpaint the missing region of an image. This algorithm is on the basis of isophote driven image sampling. The best matched patch from the source region is propagated inward to fill the missing region. The orientation of the isophote is automatically preserved. Given an input image with a target missing region, the size of the template window is fixed to 9×9 pixels. According to this algorithm each pixel has a colour value and a confidence value. The pixels at the contours are given priority which determines the order of filling. Computing the priority of the patches and propagating the structure and texture information. Sum of square distance (SSD) is used in computing the most similar patch.

(Cheng et al. 2005) proposed a robust algorithm for exemplar image inpainting, which identifies a dropping effect of the confidence value to zero with respect to Criminisi's algorithm. The authors modified the priority function from a multiplicative form to additive form as

$$P(p) = C(p)D(p) \tag{2.5}$$

Where C (p) is the confidence term and D (p) is the data term. Criminisi algorithm works well for linear structures. In order accept it for the curved structures, the authors propose a generalized function RC(p) to smoothen the curve of the confidence term C(p) to match with the data term with ω as the regularizing factor for the case with curves is given as,

$$RC(p) = (1 - \omega) * C(p) + \ \omega \text{ where } 0 \leq \omega \leq 1 \tag{2.6}$$

Authors claim that the above priority function avoids noises and is robust to the phenomena of filling the missing region in an image.

There has been research going on improving this approach that helps in more accurate fill of the occluded region with the patch obtained from the available region of an image. One of the drawbacks of Criminisi's approach is that of high time cost. This is reduced using a search strategy in the work carried out by (Chen et al. 2007). In this work the authors claim that most of the similar patch that matches with the target patches lies around the target patch and they propose a concept of window. A window length is denoted manually for each window. Window length and actual window size is calculated as

$$Window\ length = steplength * patchsize \tag{2.7}$$

$(2 * windowlength + 1) * (2 * windowlength + 1)$ respectively.

This change in defining the window size forces the search for the similar patch in the local area and reduces the cost of computation. Along with this instead of using SSD (Sum of Square Distance), L1 Euclidean Distance Norm is used to compute the similarity of the patches. Once the most similar patch is chosen, the pixels are filled in the target region using colour transfer method instead of direct fill as in the case with Criminisi's approach. Mean and standard deviation for both the target patch Ψ_p and the most matched patch from the source region, Ψ_q is computed. The mean is subtracted from the pixels selected and these are scaled by respective standard deviation. To get the transferred image patch, the average computed for the target images is added. The authors claim the effectiveness of the proposed method through experiments.

An attempt is made to improve the maintenance of linear edges during the filling of the pixels in the target area in (Martínez-Noriega et al. 2012). In Criminisi algorithm, the confidence term and the data term are used in identifying the patch to be filled, where as in the proposed work authors modified the priority function as,

$$P(p) = C(p) exp\left(\frac{D(p)}{2\sigma^2}\right) \tag{2.8}$$

In addition to this, to select similar patch from the contour region and the known area, a new metric is chosen in this method as a combination of SSD (Sum of Square Distance) and Hellinger distance. This metric proposed by the authors in this work, carries the whole information from the candidate patches. Authors claim that this metric helps in selecting more visually pleasant patches for filling in the missing region.

(Patel et al. 2014) proposed a modified exemplar image inpainting algorithm. In this method, authors feel that measuring similarity of two patches based only on colour is not enough to diffuse the structure into the target region that is to be filled. Hence a new function, image gradient is added to this function to get the value from

$$G = G(\Psi\mathrm{p} - \Psi\mathrm{q}) \tag{2.9}$$

Where, G defines each pixel’s gradient value in both the patches. Including the gradient value into the function, the similarity function is based on the difference in colour and the gradient value of the pixels. Magnitude and direction of the gradient provides information like how quickly the colour changes and in which direction it is changing. Experimental result with this modification in the exemplar based image inpainting algorithm shows enhancement in the performance of the algorithm compared to existing methods.

2.3 Objective

Video based face recognition is one of the biometrics that can be utilized for security and since this technique does not require the collaboration with the user, it is broadly utilized nowadays as a part of open and in addition private spots. Despite the fact that face is considered over different biometrics, posture variety remains a noteworthy test in acknowledgment.

Features extracted from a frontal face are larger and nearly exact when contrasted with the side perspective of a face. As frontal face contains more information, perceiving a frontal face picture is a simple errand when contrasted with face pictures with various postures. At the point when face acknowledgment is conveyed with still pictures, user adjusts himself with the camera keeping in mind the end goal to catch the frontal face. This is not the situation with video. At the point when faces are caught from a video camera or a surveillance camera, human faces in every frame might have boundless orientation and positions which prompts assortment of difficulties to analysts in recognition of faces. This issue has lead scientists to broadly think about on the issue as for posture variety and partial occlusion to enhance the recognition rate.

The objective of this thesis work is to improve the recognition rate of faces from video sequence. With the aim improving the recognition rate two problems of face recognition are considered

- Increase the recognition rate of faces invariant of the pose.
- Increase the recognition rate when faces are partially occluded.