Automatic Recognition Of Facial Expressions In Color Spaces: A Survey

Rajesh.K. S
II M.Tech(CSE)
CMRIT, Hyderabad

S. M. Riyazoddin
Professor in CSE,
CMRIT, Hyderabad

Veena A Kumar
Assistant Professor in CSE,
Saintgits College of Engineering,

Abstract: *Since the early ninetees, automatic facial expression analysis has become an active research area that finds potential applications in areas such as more engaging human–computer interfaces, talking heads, image retrieval and human emotion analysis. Facial expressions reflect not only emotions, but other mental activities, social interaction and physiological signals. There have been several advances in the past few years in the areas of face detection and tracking, feature extraction mechanisms and the techniques used for expression classification. This paper is a survey of some of the published works since 2001 till date. In this survey, the most prominent automatic facial expression analysis methods and systems presented in the literature are discussed. Facial motion and deformation extraction approaches as well as classification methods are discussed with respect to issues such as face normalization, facial expression dynamics and intensity and also with regard to their robustness. The paper also has a special mention on the 3D facial expression databases.*

Keywords: *Facial expression recognition; Face detection; Feature extraction; Feature selection; Facial expression classification, 3D facial expression databases*

1. Facial Expression Recognition-Introduction

Facial expression plays a significant role for human beings to communicate their emotions and expressions. Automatic facial expression analysis is an interesting area of research in computer science and it is still a challenging area. Facial expressions signal a person's affective state, cognitive activity and personality. Humans can perform expression recognition with a remarkable robustness without conscious effort even under a variety of adverse conditions such as partially occluded faces, different appearances and poor illuminations. The advances in imaging technology and ever increasing computing power have opened up a significant research platform. One reason for this growing interest is due to a wide spectrum of possible applications in diverse areas such as human-computer interaction (HCI) systems, video conferencing, augmented reality etc. Automatic facial expression recognition is a difficult task due to its inherent subjective nature. It is additionally hampered by usual difficulties encountered in pattern recognition and computer vision research. The majority of the current state-of-the-art facial expression recognition systems are based on 2-D facial images/videos which offer good performance only for the data captured under controlled conditions. There is currently a paradigm shift towards usage of 3-D facial data to yield better recognition performance under varying illuminations and conditions. However, it requires more expensive data acquisition systems and complicated processing algorithms, which may also need a systematic evaluation of the existing methodologies and recent advances in the facial expression recognition expressions, starting from data acquisition and database creation to data processing algorithms and performance evaluation.

There are three sub problems while designing automatic facial expression recognition system, face detection, extraction of facial expression information and classification of the expression. A system that performs these operations in real time and more accurately would be crucial to achieve a human-like interaction between man and machine. The majority of work conducted in this area involves 2D imagery, inspite of the problems this presents due to inherent pose and illumination variations.

According to Pantic & Rothkrantz (2000), expressions shown on the face are produced by a combination of contraction activities made by facial

muscles, with most noticeable temporal deformation around nose, lips, eyelids and eyebrows as well as facial skin texture patterns. Typical facial expressions last for a few seconds, normally between 250 milliseconds and five seconds (Fasel & Luettin, 2003). According to the psychologists Ekman and Friesen (1971), there are six universal facial expressions, namely: anger, disgust, fear, happiness, sadness and surprise as shown in Figure 1. Expressions such as happiness can be accurately identified even if they are expressed by members of different ethnic groups. Others expressions are more difficult to recognize even if expressed by the same person.

In computer vision and pattern recognition, facial expression recognition (FER) is often confused with human emotion recognition. While facial expression recognition uses purely visual information to group facial expression into abstract classes, emotion recognition is based on many other physiological signals, such as voice, pose, gesture and gaze according to Fasel & Luettin (2003). It is noteworthy to mention that emotions (or in general person's mental state) are not the only cause of facial expressions. Facial expressions could also be manifestations of physiological activities and aid verbal and non-verbal communication. These, for example, may include physical symptoms of pain and tiredness or listener responses during verbal communication. Therefore emotion recognition requires not only interpretation of facial expression but also understanding of the full contextual information.

2. Background

The first known facial expression analysis was presented by Darwin in 1872 (Darwin, 1872). He presented the universality of human face expressions and the continuity in man and animals. He pointed out that there are specific inborn emotions which originated in serviceable associated habits. After about a century, Ekman and Friesen (1971) postulated six primary emotions that possess each a distinctive content together with a unique facial expression. These prototypic emotional displays are also referred to as basic emotions in the later

literatures. They seem to be universal across human cultures and are named happiness, sadness, fear, disgust, surprise and anger. They developed the Facial Action Coding System (FACS) for describing facial expressions which is appearance-based. FACS uses 44 action units (AUs) for the description of facial actions with regard to their location as well as their intensity. Individual expressions may be modeled by single action units or action unit combinations. FACS codes expression from static pictures.

In the nineties, more works emerged, starting from Cottrell et al. (1990) who described the use of a Multi-Layer Perceptron Neural Networks for processing face images. They presented a number of face images to the network and train it to perform various tasks such as coding, identification, gender recognition and expression recognition. During this procedure, face images are projected onto a subspace in the hidden layers of the network. This subspace is very similar to the eigenfaces space. However, an important difference is that in this case, the face subspace is defined according to the application for which the system is to be used. Correct identification rates of up to 97 percent were reported when the system was tested using a database of images from 11 individuals. Mase et al. (1991) then used dense optical flow to estimate the activity of 12 of the 44 facial muscles. The motion seen on the skin surface at each muscle location was compared to a pre-determined axis of motion along which each muscle expands and contracts, allowing estimates as to the activity of each muscle to be made. Recognition rates of 86% were reported. Subsequently, Matsuno et al. (1994) presented a method of facial expressions recognition using two dimensional physical models named *Potential Net* without using feature extraction. *Potential Net* is a physical model which consists of nodes connected by springs in two dimensional grid configurations. Recognition is achieved by comparing nodal displacement vectors of a net deformed by an input image with facial expression vectors. It recognized four kinds of facial expressions, happiness, anger, surprise and sad and the hit rate is about 90%.

Yacoob and Davis (1996) continued the research in optical flow computation to identify the direction of rigid and nonrigid motions that are caused by human facial expressions. The approach was based on qualitative tracking of principal regions of the face and flow computation at high intensity gradient points. Three stages of the approach are locating and tracking prominent facial features, using optical flow at these features to construct a mid-level representation that describes spatio-temporal actions and applying rules for classification of mid-level representation of actions into one of the six universal facial expressions. In the mean time, Rosenblum et al. (1996) proposed a radial basis function network to learn the correlation of facial feature motion patterns and human expressions. The network was trained to recognize "smile" and "surprise" expressions. Success rate was about 83% to 88%. The work explores the use of a connectionist learning architecture for identifying the motion patterns characteristic of facial expressions. Essa and Pentland (1997a) also described facial structure using optimal estimation optical flow method coupled with geometric, physical and motion-based dynamic models. It is an extension variance of FACS called FACS+, which is a more accurate representation of human facial expressions. Another work by Lanitis's group (Lanitis, 1997) used an active shape model to locate facial features and then used shape and texture information at these locations to classify expressions. A recognition rate of 74% was obtained for expression of the basic emotions and also neutral expressions. In the next year, Otsuka and Ohya (1998) proposed a method for spotting segments displaying facial expression from image sequences. The motion of the face is modeled by HMM in such a way that each state corresponds to the conditions of facial muscles, e.g., relaxed, contracting, apex and relaxing. Off-line and on-line experiments were carried out. Off-line experiment obtains the optimum threshold values and to evaluate relation of recognition rate and frame rate. On-line experiments were used to evaluate recognition performance for sequences of multiple instances of facial expressions. Experiments showed that the segments for the six basic expressions can be spotted accurately near real time. In 1999, Chandrasiri et al. (1999) proposed a facial expression space which is made by the same person's peak facial expression images based on multidimensional scaling and a method for deriving trajectory of a facial expression image sequence on it. The main advantage of this method is that it focuses on both temporal and spatial changes of personal facial expression.

More variances of recognition approaches emerge after the year 2000, and some are tested with video sequences (Bourel, 2002; Pardas, 2002; Barlett et al, 2003). Shin et al. (2000) extract pleasure/displeasure and arousal dimension emotion features using hybrid approach. The hybrid approach used feature clustering and dynamic linking to extract sparse local features from edges on expression images. The expressions are happiness, surprise, sadness, disgust, fear, satisfaction, comfort, distress, tiredness and worry. They concluded that the arousal-sleep dimension may depend on the personal internal state more than pleasure-displeasure dimension, that is to say, the relative importance of dimension can have an effect on facial expression recognition on the two dimensional structure of emotion. Fasel and Lüttin (2000) described a system that adopts holistic approach that recognizes asymmetric FACS Action Unit activities and intensities without the use of markers. Facial expression extraction is achieved by difference images that are projected into a sub-space using either PCA or ICA, followed by nearest neighbor classification. Recognition rates are between 74 to 83%. Bourel (2002) investigated the representation of facial expression based on spatially-localised geometric facial model coupled to a state-based model of facial motion. The system consists of a feature point tracker, a geometric facial feature extractor, a state-based feature extractor and a classifier. The feature extraction process uses 12 facial feature points. The spatio-temporal features are then created to form a geometric parameter. The state-based model then transforms the geometric parameter into 3 possible states: 'increase', 'stable', 'decrease'. The classifier makes use of k-nearest neighbour approach. Another work on video sequence recognition is by Pardas (2002) who described a system that recognizes emotion based on MPEG4 facial animation parameters. The system is based on HMM. They defined a four-state HMM. Each state of each emotion models the observed FAPs using a probability function. Kim et al. (2003)

then proposed a method to construct a personalized fuzzy neural networks classifier based on histogram-based feature selection. Recognition rate is reported to be in the range of 91.6% to 98.0%. The system proposed in (Bartlett et al., 2003) detects frontal faces in the video stream and classifies them in seven classes in real time: neutral, anger, disgust, fear, joy, sadness, and surprise. An expression recognizer receives image regions produced by a face detector and then a Gabor representation of the facial image region is formed to be later processed by a bank of SVMs classifiers.

Ji (2005) based on FACS, developed a system that adopted a dynamic and probabilistic framework based on combining Dynamic Bayesian Networks (DBM) with FACS for modeling the dynamic and stochastic behaviors of spontaneous facial expressions. The three major components of the system are facial motion measurement, facial expression representation and facial expression recognition. Wu et al. (2005) modeled uncertainty in facial expressions space for facial expression recognition using fuzzy integral. The fuzzy measure is constructed in each facial expression space. They adopted Active Appearance Models (AAM) to extract facial key points and classify based on shape feature vector. Fuzzy C-Means (FCM) was used to build a set of classifiers. The recognition rates were found to be 83.2% and 91.6% on JAFFE and FGnet databases respectively. Yeasin et al. (2005) compared the performances of linear and non-linear data projection techniques in classifying six universal facial expressions. The three data projection techniques are Principal Component Analysis (PCA), Non-negative Matrix Factorization (NMF) and Local Linear Embedding (LLE). The system developed by (Anderson and McOwan, 2006) characterized monochrome frontal views of facial expression with the ability to operate in cluttered and dynamic scenes, recognizing six emotions universally associated with unique facial expressions, namely happiness, sadness, disgust, surprise, fear and anger. Faces are located using a spatial ratio template tracker algorithm. Optical flow of face is subsequently determined using a real-time implementation of gradient model. The expression recognition system then averages facial velocity information. The motion signatures produced are then classified using Support Vector

Machines. The best recognition rate is 81.82%. Zeng et al. (2006) classified emotional and non emotional facial expressions occurred in a realistic human conversation setting-Adult Attachment Interview (AAI). The AAI is a semistructured inter-view used to characterize individuals' current state of mind with respect to past parent-child experiences. Piecewise Bezier Volume Deformation (PBVD) was used to track face. They applied kernel whitening to map the data to a spherical symmetrical cluster. Then Support Vector Data Description (SVDD) was applied to directly fit a boundary with minimal volume around the target data. Experimental results suggested the system generalize better than using PCA and single Gaussian approaches. Xiang et al. (2007) utilized fourier transform, fuzzy C means to generate a spatio-temporal model for each expression type. Unknown input expressions are matched to the models using Hausdorff distance to compute dissimilarity values for classification. The recognition rate was found to be 88.8% with expression sequences.

In general, a facial expression recognizer comprises of 3 stages, namely feature extraction, features selection, and classification. Feature extraction involves general manipulation of the image. The raw image is processed to provide a region of interest (human face without hairs and background) for the second stage to select meaningful features. Some noise reduction, clustering, labelling or cropping may be done in this stage. The first stage is unnecessary for some ready data which are taken off the shelf from online database.

Feature selection is an important module. Without good features, the effort made in the classification stage would be in vain. Fasel and Luettin (2003) provides a detailed survey on facial expression, they classified the feature extraction methods into several groups. Some of the methods were highlighted like Gabor wavelets, Active Appearance Model, Dense Flow Fields, Motion and Deformable Models, Principal Component Analysis, High Gradient Components and etc. They group the facial features into 2 types: intransient facial features and transient facial features. Intransient features are like eyes, eyebrow and mouth which are always present in the face. Transient features include wrinkles and bulges.

Neural network is a popular choice for classification. Most of them fall under supervised learning. The other classification methods underscored by Pantic and Rothkrantz (2000) are Expert System Rules, Discriminant functions by Cohn et al. (1998), Spatio-temporal motion energy templates by Essa and Pentland (1997b), Thresholded motion parameters by (Black & Yacoob, 1997), and Adaptive Facial Emotion Tree Structures by Wong and Cho (2006). A latest work by Seyed Mehdi Lajevardi, *Member, IEEE*, and Hong Ren Wu (2012) proposed the use of LDA classifier for classification. In the past two decades, research has focused on how to make face recognition systems fully automated by tackling problems, such as, localization of a face in a given image or video clip and extraction of features such as eyes, mouth, etc. Meanwhile, significant advances have been made in the design of feature extractions and classifiers for successful face recognition. Both the appearance-based holistic approaches and feature-based methods have the strength and weaknesses. Compared to holistic approaches, feature-based methods are less sensitive to variations in illumination and viewpoint and to inaccuracy in face localization. Several survey papers offer more insights on facial expression systems which can be found in (Andrea et al., 2007; Kevin et al., 2006; Zhao et al., 2003; Fasel & Luettin, 2003; Lalonde & Li, 1995).

3. Face Detection

The first step in the Facial Expression Recognition system is face detection, i.e., identification of all regions in the scene that contain a human face. The problem of finding faces should be solved regardless of clutter, occlusions and variations in head pose and lighting conditions. The presence of non-rigid movements due to facial expression and a high degree of variability in facial size, color and texture make this problem even more difficult. Numerous techniques have been developed for face detection in still images (Yang et al., 2002; Li & Jian, 2005). The most common face detector was proposed by Viola and Jones (2004) to detect the face region from either a still image or a video stream. This detector consists of a cascade of classifiers trained by AdaBoost. Each classifier employs integral image filter, also called "box filters", which are reminiscent of Haar Basis functions, and can be computed very fast at any

location and scale. This is essential to the speed of the detector. For each stage in the cascade, a subset of features is chosen using a feature selection procedure based on AdaBoost.

4. Facial Expression Representations

Facial expression representation is basically a feature extraction process, which converts the original facial data from a low-level 2D pixel or 3D vertex based representation, into a higher-level representation of the face in terms of its landmarks, spatial configurations, shape, appearance and motion. The extracted features usually reduce the dimensionality of the original input facial data according to Park & Park (2004). Following are a number of popular facial expression representations.

A landmark based representation uses facial characteristic points, which are located around specific facial areas such as edges of eyes, nose, eyebrows and mouth. These areas show significant changes during facial articulation. Kobayashi and Hara (1997) proposed a geometric face model based on 30 facial characteristic points for the frontal face view. Afterwards, Pantic & Rothkranthz has proposed an extension to the point-based model to include 10 extra facial characteristic points on the side view of the face (2000). These points on the side view are selected from the peaks and valleys of the profile contours.

The localised geometric model could be classified as a representation based on spatial configuration derived from facial images (Saxena et al., 2004). The method utilises a facial feature extraction method, which is based on classical edge detectors combined with color analysis in the hue, saturation, value (HSV) color space to extract the contours of local facial features, such as eyebrows, lips and nose. As the color of the pixels representing lips, eyes and eyebrows differ significantly from those representing skin, the contours of these features can be easily extracted from the hue color component. After facial feature extraction, a feature vector built from feature measurements, such as the brows distance, mouth height, mouth width etc., is created.

Another representation based on spatial configuration is topographic context (TC) that has been used as a descriptor for facial expressions in 2D images (Wang & Yin, 2007). This representation treats an intensity image as a 3D terrain surface with the height of the terrain at pixel (x,y) represented by its image grey scale intensity $I(x,y)$. Such image interpretation enables topographic analysis of the associated surface to be carried out leading to a topographic label, calculated based on a local surface shape, being assign to each pixel location. Resulting TC feature is an image of such labels assigned to all facial pixels of the original image. Topographic labels include peak, ridge, saddle, hill, flat, ravine and pit. In total, there are 12 types of topographic labels (Trier et al., 1997). Additionally, hill-labeled pixels can be divided into concave hill, convex hill, saddle hill (that can be further classified as a concave saddle hill or a convex saddle hill) and slope hill and saddle-labelled pixels can be divided into ridge saddle or ravine saddle. For a facial image, the TC-based expression representation requires only six topographic labels shown in Figure 5 (Yin et al., 2004).

After the presence of a face has been detected in the observed scene, the next step is to extract the information about the displayed facial signals. Most of the existing facial expression analyzers are directed toward 2D spatiotemporal facial feature extraction. The feature extractor then adopted Gabor wavelet feature extraction.

Gabor wavelet is a popular choice because of its capability to approximate mammals' visual cortex. The primary cortex of human brain interprets visual signals. It consists of neurons, which respond differently to different stimuli attributes. The receptive field of cortical cell consists of a central ON region is surrounded by 2 OFF regions, each region elongates along a preferred orientation (Daugman, 1985). According to Jones and Palmer (1987), these receptive fields can be reproduced fairly well using Daugman's Gabor function.

In the latest work by Seyed Mehdi Lajevardi, *Member, IEEE*, and Hong Ren Wu(2012), Gabor Filters along with 2-D FFT(fast Fourier Transform) have been used for Feature Extraction.

5. Feature Selection

Feature selection plays an important role in the automatic facial expression recognition system. The objectives of feature selection include noise reduction, regularization, relevance detection and reduction of computational effort. It is likely a learned image filters, such as, ICA, PCA and LFA which are based on unsupervised learning from the statistics of large image databases. There are two main models of feature selection, the filter model and the wrapper model. The filter model is generic, which is not optimized for any specific classifier. It is modularized and may sacrifice classification accuracy. The wrapper model is always tailored to a specific classifier and it may lead to better accuracy as a result. The strength of the wrapper model is that it differentiates irrelevance, strong and weak relevance and it improves performance significantly on some datasets. One weakness of the wrapper model is that calling the induction algorithm repeatedly may cause over fitting. Another weakness is that its computational cost is expensive. So the filter model is more preferable in Facial Expression Recognition related works.

6. Facial Expression Classification

In the context of facial expression analysis, classification is a process of assigning observed data to one of predefined facial expression categories. The specific design of this process is dependent on the type of the observation (e.g. static or dynamic), adopted data representation (type of the feature vector used to represent the data) and last but not the least the classification algorithm itself. A great variety of classification algorithms have been reported for facial expressions in literature, We focus only on those, which are the most often used, or which have been recently proposed in the context of facial expression recognition. From that perspective, some of the most frequently used classification methods include Nearest Neighbor Classifiers, Fisher's Linear Discriminant (also known as Linear Discriminant Analysis), support vector machines, artificial neural networks, AdaBoost, random forests and Hidden Markov models.

Nearest Neighbor Classifier (NNC) is one of the simplest classification methods, which classifies

objects based on closest training examples in the feature space. It can achieve consistently high performance without a prior assumption about the distribution from which the training data is drawn. Although there is no explicit training step in the algorithm, the classifier requires access to all training examples and the classification is computationally expensive when compared to other classification methods. The NNC assigns a class based on the smallest distances between the test data and the data in the training database, calculated in the feature space. A number of different distance measures have been used, including Euclidian and weighted Euclidian (Md. Sohail and Bhattacharya, 2007), or more recently geodesic distance for features defined on a manifold (Yousefi et al., 2010).

Linear discriminant analysis (LDA) finds linear decision boundaries in the underlying feature space that best discriminate among classes, i.e., maximise the between-class scatter while minimise the within-class scatter (Fisher, 1936). A quadratic discriminant classifier (Bishop, 2006) uses quadratic decision boundary and can be seen, in the context of Bayesian formulation with normal conditional distributions, as a generalisation of a linear classifier incase when class conditional distributions have different covariance matrices.

In recent years, one of the most widely used classification algorithms are support vector machines (SVM) which performs classification by constructing a set of hyperplanes that optimally separate the data into different categories (Huang et al., 2006). The selected hyperplanes maximise the margin between training samples from different classes. One of the most important advantages of the SVM classifiers is that they use sparse representation (only a small number of training examples need to be maintained for classification) and are inherently suitable for use with kernels enabling nonlinear decision boundary between classes.

Other popular methods are the artificial neural networks. The key element of these methods is the structure of the information processing system, which is composed of a large number of highly interconnected processing elements working together to solve specific problems (Padgett et al., 1996).

AdaBoost (Adaptive Boosting) is an example of so called boosting classifiers which combine a number of weak classifiers/learners to construct a strong classifier. Since its introduction (Freund and Schapire, 1997), AdaBoost is enjoying a growing popularity. A useful property of these algorithms is their ability to select an optimal set of features during training. As results AdaBoost is often used in combination with other classification techniques where the role of the AdaBoost algorithm is to select optimal features which are subsequently used for classification by another algorithm (e.g. SVM). In the context of facial expression recognition Littlewort (Littlewort et al., 2005) used the AdaBoost to select best Gabor features calculated for 2D video which have been subsequently used within SVM classifier. Similarly in (Ji and Idrissi, 2009) authors used a similar combination of AdaBoost for feature selection and SVM for classification with LBP calculated for 2D images. In (Whitehill et al., 2009) authors used a boosting algorithm (in that case GentleBoost) and the SVM classification algorithm with different features including Gabor filters, Haar features, edge orientation histograms, and LBP for detection of smile in 2D stills and videos. They demonstrated that when trained on real-life images it is possible to obtain human like smile recognition accuracy. Maalej (Maalej et al., 2010) successfully demonstrated the use of Adaboost and SVM, utilising different kernels, with feature vector defined as geodesic distances between corresponding surface's patches selected in the input 3D static data.

More recently random forest (Breiman, 2001) classification techniques have gained significant popularity in the computer vision community. In (Flanelli et al., 2010) authors used random forest with trees constructed from a set of 3D patches randomly sampled from the normalised face in 2D video. The decision rule used in each tree node was based on the features calculated from a bank of log-Gabor filters and estimated optical flow vectors.

Hidden Markov Models (HMM) are able to capture dependence in a sequential data and therefore are

often the method of choice for classification of spatio-temporal data. As such they have been also used for classification of facial expressions from 2D (Cohen et al., 2003) and 3D (Sun & Yin, 2008) video sequences.

In the latest work by Seyed Mehdi Lajevardi, *Member, IEEE*, and Hong Ren Wu(2012), LDA classifier has been used for Feature Extraction.

7. Facial Expression Databases

In order to evaluate and benchmark facial expression analysis algorithms, standardized data sets are required to enable a meaningful comparison. Based on the type of the facial data used by an algorithm, the facial expression databases can be categorized into 2-D image, 2-D video, 3-D static and 3-D dynamic. Since facial expressions have been studied for a long time using 2-D data, there are a large number of 2-D image and 2-D video databases available. Some of the most popular 2-D image databases include CMU-PIE database (Sim et al., 2002), Multi-PIE database (Gross et al., 2010), MMI database (Pantic et al., 2005) and JAFFE database (Lyons et al., 1999). The commonly used 2-D video databases are Cohn-Kanade AU-Coded database (Kanade et al., 2000), MPI database (Pilz et al., 2006), DaFEx database (Battocchi et al., 2005) and FG-NET database (Wallhoff, 2006). Due to the difficulties associated with both 2-D image and 2-D video based facial expression analysis in terms of handling large pose variation and subtle facial articulation, there is recently a shift towards the 3-D based facial expression analysis, however this is currently supported by a rather a limited number of 3-D facial expression databases. These databases include BU-3DFE (Yin et al., 2006) and ZJU-3DFED (Wang et al., 2006b). With the advances in 3-D imaging systems and computing technology, 3-D dynamic facial expression databases are beginning to emerge as an extension of the 3-D static databases. Currently the only available databases with dynamic 3-D facial expressions are ADSIP database (Frowd et al., 2009) and BU-3DFE database (Yin et al., 2008). The 3-D facial expression databases are explained in detail below.

7.1 3-D Facial Expression Databases

The static BU-3DFE facial expression database was developed at the Binghamton University for the purpose of 3-D facial expression analysis (Yin et al., 2006). The database contains 100 subjects, with ages ranging from 18 to 70 years old, with a variety of ethnic origins including White, Black, East-Asian, Middle-East Asian, Indian and Hispanic. Each subject performed seven expressions, which include neutral and six universal facial expressions at four intensity levels. With 25 3-D facial scans containing different expressions for each subject, there is a total of 2,500 facial scans in the database. Each 3-D facial scan in the BU-3DFE database contains 13,000 to 21,000 polygons with 8,711 to 9,325 vertices.

ZJU-3DFED database is a static 3-D facial expression database developed in the Zhe Jiang University (Wang et al., 2006b). Compared to other 3-D facial expression databases, the size of ZJU-3DFED is relatively small. It contains 360 facial models from 40 subjects. For each subject, there are 9 scans with four different kinds of expressions.

The 3-D dynamic facial expression database BU-4DFE database (Yin et al., 2008) is an extension of the BU-3DFE database, to enable the analysis of the facial articulation using dynamic 3-D data. The 3D facial expressions are captured at 25 Frames per Second (FPS), and the database includes 606 3D facial expression sequences captured from 101 subjects. For each subject, there are six sequences corresponding to six universal facial expressions (anger, disgust, happiness, fear, sadness, and surprise).

ADSIP database is a 3-D dynamic facial expression database created at the University of Central Lancashire (Frowd et al., 2009). The first release of the database (ADSIPmark1) was completed in 2008 with the help from 10 graduates of the School of Performing Arts. The use of actors and trainee actors enables capture of fairly representative and accurate facial expressions (Nusseck et al., 2008). Each subject performed seven expressions: anger, disgust, happiness, fear, sadness, surprise and pain, at three intensity levels (mild, normal and extreme). Therefore, there is a total of 210 3D facial sequences

in that database. Each sequence was captured at 24 fps and lasts for around three seconds. Additionally each 3D sequence is accompanied by a standard video recording captured in parallel with the 3D sequence. This database is unique in the sense that it has been independently validated as all the recordings in the database have been assessed by 10 independent observers. These observers assigned a score against each type of the expression for all the recordings. Each score represented how confident observers were about each sequence depicting each type of the expression. The ADSIPmark1 database is being gradually expanded. The new acquisitions are captured at 60 fps. Furthermore, some additional facial articulations with synchronised audio recording are captured, with each subject reading a number of predefined special phrases typically used for the assessment of neurological patients (Quan et al., 2010a; Quan et al., 2010b). The final objective of the ADSIP database is to contain 3-D dynamic facial data of over 100 control subjects and additional 100 subjects with different facial disfunctions.

The latest work by Seyed Mehdi Lajevardi, *Member, IEEE*, and Hong Ren Wu(2012), "facial expression recognition in perceptual color space" has been tested
.

8. Conclusion

This paper has discussed several concepts related to automatic facial expression recognition. Although these have included description of general issues relevant to such problem, the main emphasis has been on a review of the recent developments in the corresponding processing pipeline including data acquisition, face normalization, feature extraction and subsequent classification. The available 3-D facial expression databases were also introduced to provide complete information about available options for algorithm validation. This survey paper throws light on the available practices of FER.

References:

[1]. A. Samal and P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey," *Pattern Recognition,* vol. 25, no. 1, pp. 65-77, 1992.

[2]. B. Fasel and J. Luttin, "Automatic Facial Expression Analysis: a survey," *Pattern Recognition,* vol. 36, no. 1, pp. 259-275, 2003.

[3]. M. Pantic, L.J.M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art", *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol.22, no. 12, pp. 1424-1445, Dec. 2000.

[4]. Z. Zeng, M. Pantic, G.I. Roisman and T.S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 31, no. 1, pp. 39–58, Jan. 2009.

[5]. Vinay Kumar Bettadapura, "Face Expression Recognition and Analysis:The State of the Art".College of Computing, Georgia Institute of Technology,2012.

[6]. P. Ekman, E. T. Rolls, D. I. Perrett, and H. D. Ellis, "Facial expressions of emotion: An old controversy and new findings discussion," *Phil.Trans. Royal Soc. London Ser. B, Biol. Sci.*, vol. 335, no. 1273, pp. 63–69, 1992.

[7]. S. M. Lajevardi and Z. M. Hussain, "Emotion recognition from color facial images based on multilinear image analysis and Log-Gabor filters," in *Proc. 25th Int. Conf. Imag. Vis. Comput.*, Queenstown, New Zealand, Dec. 2010, pp. 10–14.

[8]. C. J. Young, R. Y. Man, and K. N. Plataniotis, "Color face recognition for degraded face images," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 5, pp. 1217–1230, Oct. 2009.

[9]. M. Thomas, C. Kambhamettu, and S. Kumar, "Face recognition using a color subspace LDA approach," in *Proc. 20th IEEE Int. Conf. Tools Artif. Intell.*, Dayton, OH, Nov. 2008, pp. 231–235.

[10]. S. M. Lajevardi and Z. M. Hussain, "Automatic facial expression recognition: Feature extraction and selection," *Signal, Image Video Process.*, vol. 6, no. 1, pp. 159–169, 2012.

[11]. Y. Lijun, C. Xiaochen, S. Yi, T. Worm, and M. Reale, "A high-resolution 3-D dynamic facial expression database," in *Proc. 3rd Int. Conf. Face Gesture Recognit.*, Amsterdam, The Netherlands, Sep. 2008, pp. 1–6.