4

Facial Expression Recognition

Bogdan J. Matuszewski, Wei Quan and Lik-Kwan Shark
ADSIP Research Centre, University of Central Lancashire
UK

1. Introduction

Facial expressions are visible signs of 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 illumination. Over the last two decades, the advances in imaging technology and ever increasing computing power have opened up a possibility of automatic facial expression recognition and this has led to significant research efforts from the computer vision and pattern recognition communities. One reason for this growing interest is due to a wide spectrum of possible applications in diverse areas, such as more engaging human-computer interaction (HCI) systems, video conferencing, augmented reality. Additionally from the biometric perspective, automatic recognition of facial expressions has been investigated in the context of monitoring patients in the intensive care and neonatal units for signs of pain and anxiety, behavioural research, identifying level of concentration, and improving face recognition.
Automatic facial expression recognition is a difficult task due to its inherent subjective nature, which is additionally hampered by usual difficulties encountered in pattern recognition and computer vision research. The vast majority of the current state-of-the-art facial expression recognition systems are based on 2-D facial images or videos, which offer good performance only for the data captured under controlled conditions. As a result, there is currently a shift towards the use of 3-D facial data to yield better recognition performance. However, it requires more expensive data acquisition systems and sophisticated processing algorithms. The aim of this chapter is to provide an overview of the existing methodologies and recent advances in the facial expression recognition, as well as present a systematic description of the authors' work on the use of 3-D facial data for automatic recognition of facial expressions, starting from data acquisition and database creation to data processing algorithms and performance evaluation.

1.1 Facial expression

Facial expressions are generated ... skin texture" (Pantic & Rothkrantz, 2000)" should be replaced by "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 (Pantic & Rothkrantz, 2000). Typical facial expressions last for a few seconds, normally between 250 milliseconds and five seconds (Fasel & Luettin, 2003). According to psychologists Ekman and Friesen

(Ekman & Friesen, 1971), there are six universal facial expressions, namely: anger, disgust, fear, happiness, sadness, and surprise, as shown in Figure 1. Among these universal expressions, some, such as happiness, can be accurately identified even if they are expressed by members of different ethnic groups. Others are more difficult to recognise even when expressed by the same person.

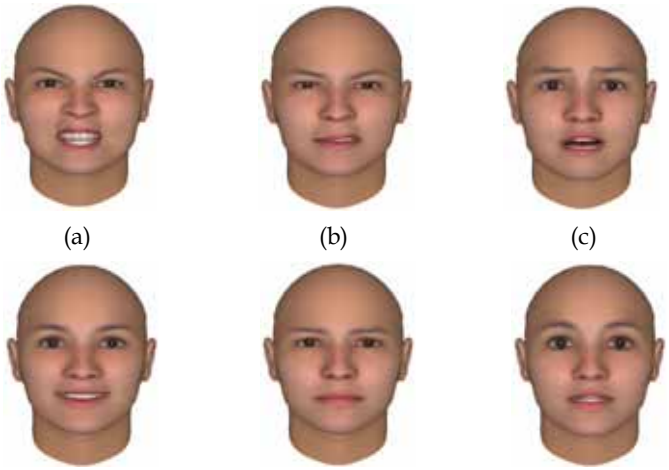


Fig. 1. Examples of six universal computer simulated expressions: (a) anger, (b) disgust, (c) fear, (d) happiness, (e) sadness, and (f) surprise (FaceGen, 2003).

In computer vision and pattern recognition, facial expression recognition is often confused with human emotion recognition. While facial expression recognition uses purely visual information to group facial articulation into abstract classes, emotion recognition is based on many other physiological signals, such as voice, pose, gesture and gaze (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.

1.2 Applications

Facial expression analysis can be traced back to the nineteenth century with Darwin's theory on the similarity of facial expressions across different cultures and between different species (Darwin, 1872). From then on, most of research was conducted by psychologists until 1978 with Suwa et al. (Suwa et al., 1978) presenting a preliminary investigation on automatic facial expression analysis from an image sequence. Currently automatic facial expression analysis and recognition have become an active research area associated with a wide range of applications, such as human-machine interaction, video conferencing, virtual reality, and biometrics.

Automatic recognition of facial expressions can act as a component of human-machine interface or perceptual interface (van Dam, 2000; Pentland, 2000). This kind of interface would be able to support the automated provision of services that require a good appreciation of the emotion from the user (Bartlett et al., 2003). For example, it could provide the user's intentions and feelings to a machine or robot to enable it to respond more appropriately during the service (Essa & Pentland, 1997). In the application of robot-assisted learning whereby a robot is used to teach the user by explaining the content of the lesson and question the user afterwards, understanding the human emotion will enable the robot to progress from one lesson to the next when the user is ready (Wimmer et al., 2008).
Video conferencing, tele-presence and tele-teaching require transmission of a large amount of data, and data compression is often needed in order to reduce the storage and bandwidth requirements. Facial expression analysis offers an approach to achieve data compression for video conferencing (Pearson, 1995; Aizawa & Huang, 1995). Using the video frames recorded by the cameras in front of the people attending the video conference, facial expressions and motions of each person can be estimated at the transmitting side as a set of parameters that describe the current appearance of each person, thereby reducing the amount of data to be transmitted. At the other side of the video conference, the set of parameters received is used to render the facial model to present the approximate appearance of each person (Eisert & Girod, 1998). The technique of automatic facial expression analysis used in video conferencing can also be applied to virtual reality to provide realistic synthesis of facial articulation (Morishima & Harashima, 1993).
In terms of biometrics, automatic expression recognition has been investigated in the context of monitoring patients in the intensive care and neonatal units (Brahnam et al., 2006) for signs of pain and anxiety, behavioural research on children's ability to learn emotions by interacting with adults in different social contexts (Pollak & Sinha, 2002), identifying level of concentration (Vural et al., 2007), for example detecting driver tiredness, and finally in aiding face recognition.

1.3 Challenges

Despite of the significant progress made in both research and application development, automatic facial expression recognition remains a particularly challenging problem (Wang & Yin, 2007; Bartlett et al., 2003). Broadly speaking, there are two major obstacles. One is related to a robust capture of facial expressions, and the other is associated with machine learning.
With facial data being the information source for the facial expression recognition task, the processing complexity and expression recognition performance are strongly dependent on the data capture technique used. Although simple and low cost, current expression recognition systems based on 2-D imaging are only able to achieve good recognition performance in constrained environments due to difficulties in handling large variations in illumination and view angle (Quan, 2009a). These difficulties arise from the fact that the human face is a three-dimensional surface rather than a two-dimensional pattern, resulting in its 2-D projection being sensitive to changes in lighting and head pose (Pantic & Rothkrantz, 2000). This has led to the increased use of 3-D facial data capture systems, since it is largely immune to changes in pose and illumination (Yin et al., 2006; Quan, 2009a). However, it needs to be stressed that the 3-D face acquisition does not solve all the problems as for example it does not help to alleviate issues associated with occlusions, where typical examples of facial occlusions include subjects wearing glasses, or having long hair.

The machine learning challenges are related to facial feature extraction and classification to achieve a high performance of the facial expression recognition. The extracted features should represent different types of facial expressions in a way which is not significantly affected by age, gender, or ethnic origin of the subject. The classification method must be capable to define appropriate rules in order to derive a specific type of facial expression from the facial features provided, even when the output from the preceding processing stages, such as facial data acquisition and facial feature extraction, is noisy or incomplete (Wimmer et al., 2008).

2. Facial expression recognition systems

In general, a typical facial expression recognition system, whether automatic or semi-automatic, consists of three main processing stages: acquisition, representation and classification. The general framework of a facial expression recognition system is illustrated in Figure 2 with each processing stage discussed in the following subsections. Facial data acquisition addresses how to detect, locate and acquire facial data in complex 2-D or 3-D scenes. Facial expression representation is concerned with the extraction of representative facial features for different types of expressions to give the required accuracy and robustness. Facial expression classification is tasked with finding a suitable classification algorithm to categorise facial expressions in terms of the facial features provided by the facial expression representation stage.

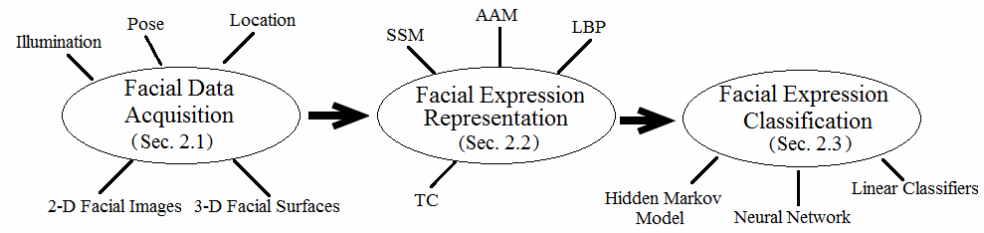


Fig. 2. A general framework of facial expression recognition system

2.1 Facial data acquisition

Facial data can be acquired in 2-D or 3-D, both in a static or dynamic mode. A static 2-D facial image represents a scene at a fixed moment in time and as such does not contain information about temporal evolution of the scene; a dynamic 2-D facial data set is a time ordered collection of images and therefore can provide information about expression evolution which can aid recognition. For faces appeared in complex scenes with cluttered backgrounds, face detection algorithms are required to locate the facial area in each image (Heisele et al., 2001; Jones & Viola, 2003), since most of the face expression analysis methods need the exact position of the face in order to extract facial features of interest (Lanitis et al., 1997; Hong et al., 1998; Steffens et al., 1998). Furthermore, the captured facial images usually need to be filtered so as to reduce the lighting variation as well as the specular reflection on eyes, teeth and skin (Black et al., 1998). Often for the subsequent feature extraction some assumptions need to be made about face appearance, these may include size, face orientation and intensity variations mentioned above. If the actual detected face does not

fulfil these restrictions, e.g. it is too small, it needs to be re-rendered. Such process is called face normalisation and is often applied after the face detection step and before the feature extraction stage (Flanelli et al., 2010). Interestingly some authors proposed to use 3-D cues for facial image normalisation even though the actual recognition process is based on 2-D information (Niese et al., 2007).

A static 3-D facial data set is usually represented as a set of 3-D points or surface patches. Such data is normally captured by 3-D imaging systems (often called 3-D scanners). They scan a real-world object generating a geometric point cloud corresponding to samples taken from the observed 3-D surface. Apart from surface geometry, such 3-D scanners can often provide information about the corresponding 3-D point appearance e.g. colour. In general, there are two major types of 3-D scanners, contact and non-contact (Curless, 2000). Contact 3-D scanners, used mostly in manufacturing, are seldom used for facial expression analysis. Despite the fact that they provide accurate surface measurements, they require prohibitively long acquisition time. Non-contact scanners are much more suitable for 3-D facial data acquisition. They can be further divided into two broad categories of active and passive scanners.

Active scanners measure object surface by emitting a light pattern and detecting its reflection (Zhang & Huang, 2006). Active scanners commonly found in applications of facial data analysis use structured or random patterns which are projected on the face surface. The surface reconstruction is based on detected geometrical distortions of the pattern. Such scanners often include one or more cameras and a projector.

Passive 3-D scanners use the principles of multi-view geometry (Bernardini & Rushmeier, 2002) utilising information from multiple cameras placed around the face surface without emitting any kind of radiation. The most popular type of passive scanners uses two cameras to obtain two different views of the face surface. By analysing the position differences between the corresponding surface points seen by each camera, the distance of face surface point can be determined through triangulation. A passive scanner from the Dimensional Imaging which has been used by the authors for the research on facial expression analysis is shown in Figure 3. This 3-D scanner uses six digital cameras with three cameras on each side to capture six views of the face surface. Four of these images are used to reconstruct 3D facial geometry and the other two images provide the textural information for accurate 3D face rendering.

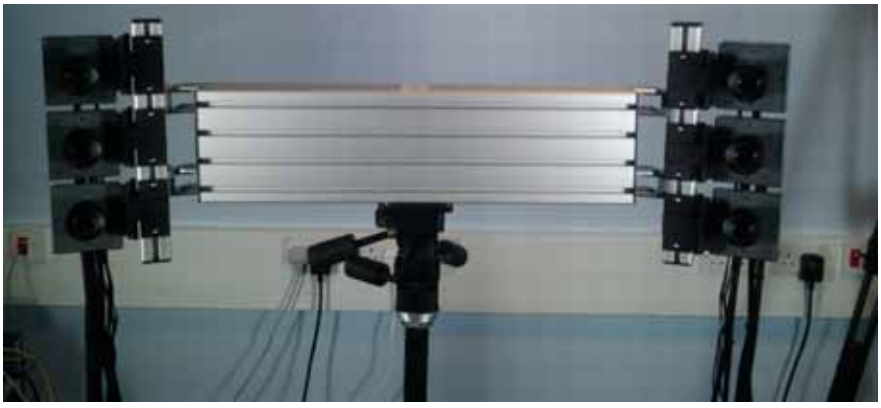


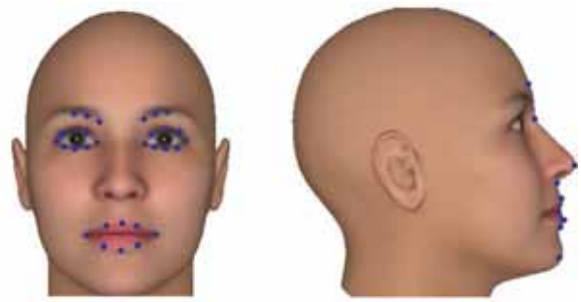
Fig. 3. A passive 3-D scanner from Dimensional Imaging.

Similar to the dynamic 2-D data, a dynamic 3-D facial data set is a time ordered collection of 3-D surfaces. This, among others, enables temporal tracking of facial points' motion in the 3-D space. The scanner shown in Figure 3 is capable of capturing dynamic data up to 60 frames per second (fps). Each second of the facial data with a resolution of 20,000 vertices captured at 60 fps requires around 10 Gigabytes of storage. This example illustrates one of the main disadvantages of 3-D dynamic scanners, namely: required availability of significant storage and computational resources.

2.2 Facial expression representations

Facial expression representation is essentially a feature extraction process, which converts the original facial data from a low-level 2-D pixel or 3-D vertex based representation, into a higher-level representation of the face in terms of its landmarks, spatial configuration, shape, appearance and/or motion. The extracted features usually reduce the dimensionality of the original input facial data (Park & Park, 2004) (a noticeable example to the contrary would be Haar or Gabor features calculated as an input for the AdaBoost training algorithm, where dimensionality of the feature vector could be higher than the dimensionality of the original data). Presented in the 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, since these areas show significant changes during facial articulation. Kobayashi and Hara (Kobayashi & Hara, 1997) proposed a geometric face model based on 30 facial characteristic points for the frontal face view as shown in Figure 4(a). Subsequently, the point-based model was extended to include 10 extra facial characteristic points on the side view of the face (Pantic & Rothkranthz, 2000) as shown in Figure 4(b). These points on the side view are selected from the peaks and valleys of the profile contours.



(a) (b)

Fig. 4. Point-based model: (a) 30 facial points selected from the frontal view, (b) 10 facial points selected from the side view.

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 colour analysis in the HSV (hue, saturation, value) colour space to extract the contours of local facial features, such as eyebrows, lips and nose. As the colour 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 colour 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 2-D images (Wang & Yin, 2007). This representation treats an intensity image as a 3-D 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). In addition, hill-labelled 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).

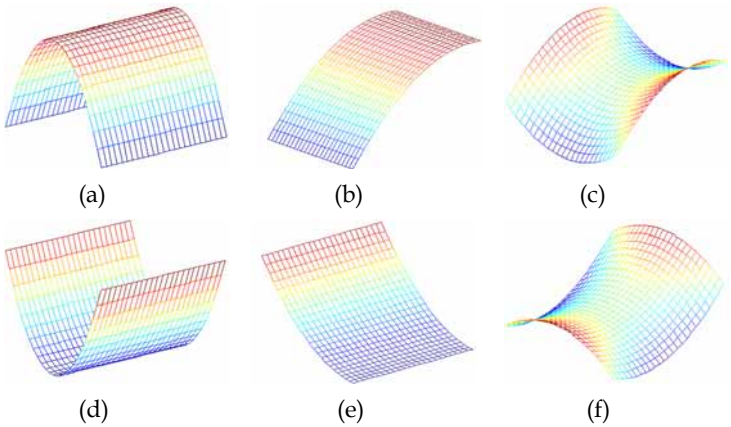


Fig. 5. A subset of the topographic labels: (a) ridge, (b) convex hill, (c) convex saddle hill, (d) ravine, (e) concave hill, and (f) concave saddle hill.

Similar to TC, the local binary patterns (LBP) have been also used to represent facial expressions in 2-D images (Liao et al., 2006). LBP as an operator was first proposed by Ojala et al. (Ojala et al., 2002) for texture description. An example of LBP calculation is illustrated in Figure 6. For a given pixel (shown in red in Figure 6), its value is subtracted from all the neighbouring pixels and the sign of the results is binary coded. After the clockwise grouping of the binary bits, starting from the top left pixel, the arranged binary string is converted to a decimal number as the final LBP result for that pixel. The LBP operator is an intensity invariant texture measure, which captures the directions of the intensity changes in an image. Using the LBP operator, each pixel of a facial image can be encoded by a LBP value which preserves the intensity difference with respect to its local neighbours. This encoded image can be used for the classification of facial expressions. Figure 7 shows a facial image and its corresponding LBP encoded image. The original LBP algorithm has also been

modified to encode the information of depth differences and applied to 3-D face representations (Huang et al., 2006).

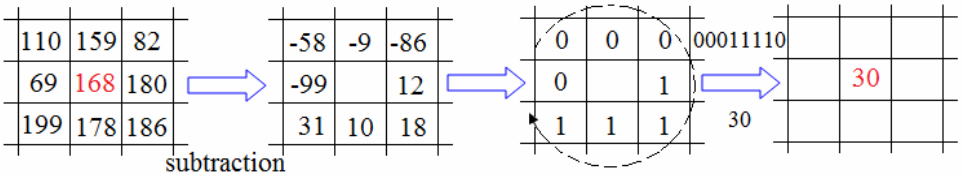


Fig. 6. An example of LBP calculation.

(a) (b)

Fig. 7. A facial image and its LBP encoded image: (a) original, and (b) encoded.

Among various shape based representations, the facial expression representation method based on the B-spline surface was proposed by Hoch (Hoch et al., 1994) for 3-D facial data. The B-spline surface is a parametric model which can efficiently represent 3-D faces using a small number of control points. Using B-splines in combination with the facial action coding system (FACS) (Ekman & Friesen, 1978), different facial expressions can be automatically created on the faces by moving the specific control points which correspond to the different action units of the FACS. The FACS is a human observer based system with the ability to detect subtle changes in facial features and it forms a widely accepted theoretical foundation for the control of facial animation (Hager et al., 2002). There are 46 action units defined by the FACS with each action unit representing a specific basic action performable on the human face.

Another often used shape based representation is the statistical shape model (SSM) (Cootes et al., 1995). The SSM can be built based on facial points, landmarks or other parameters (e.g. B-spline coefficients) describing face shape. Given a set of training faces with know correspondences between them, the SSM finds a mean representation of these faces, e.g. mean configuration of the facial landmarks, and their main modes of variation (so called eigen-faces). With such SSM each new face can be approximately represented as a linear superposition of the mean face and the weighted eigen-faces. The weights, calculated through projection of the observed face on the subspace spanned by the eigen-faces, can be subsequently used as an approximate representation of the face.

Using the SSM constructed, a 3-D face in the training data set is represented by a shape space vector (SSV) of much smaller dimensionality than the dimensionality of the original data space vector (Blake & Isard, 1998). To produce the SSV for an unseen new face, a modified iterative closest point (ICP) method is often incorporated with the least-squares

projection to register the SSM to the new face (Cootes & Taylor, 1996). For 3-D facial expression representation, it has been proved experimentally by the authors (Quan et al., 2007a; Quan et al., 2007b) that the SSV is able to encode efficiently different facial expressions.

A combination of shape and appearance based representations yields the active appearance model (AAM), which could be classified as another statistical model and has been used for facial expression representation (Hong et al., 2006; Edwards et al., 1998). It models the shape as well as grey levels textures and it is mainly used for 2-D facial images in facial expression representation applications. The model is built using a set of training facial images with corresponding landmarks selected manually and localised around the prominent facial features. In order to build the AAM, the selected landmarks representing the shape of each training face are aligned into a common coordinate system with each training facial image normalised to the same size. Subsequently, principal component analysis (PCA) is applied to the aligned landmarks and the normalised facial images to yield a vector of parameters which represents both the shape and grey level variations of the training facial images (Cootes et al., 1998).

The 3-D morphable model is an extension of the AAM for 3-D data. Instead of using the manually selected landmarks, the 3-D morphable model uses all the data points of 3-D facial scans to represent the geometrical information (Blanz & Vetter, 2003). Optic flow is often used for establishing the dense correspondences of points between each training face and a reference face in texture coordinate space. This model has been used for representing the facial expressions which are embedded in the 3-D faces (Ramanathan et al., 2006).

Finally, a motion based representation is the local parameterised model, which is calculated using the optical flow of image sequences (Black & Yacoob, 1997). Two parametric flow models are available for image motion, namely: affine and affine-plus-curvature. They not only model non-rigid facial motion, but also provide a description about the motion of facial features. Given the region of a face and locations of facial features of interest, such as eyes, eye brows and mouth, the motion of the face region between two frames in the image sequence is first estimated using the affine model. Subsequently the estimated motion is used to register the two images via warping, and the relative motions of the facial features are computed using the affine-plus-curvature model. Using the estimated motions, the region of the face and the locations of the facial features are predicted for the next frame, and the process is repeated for the whole image sequence. The parameters of estimated motions provide a description of the underlying facial motions and can be used to classify facial expressions. The region of the face and the locations of the facial features can be manually selected in the initial processing stage and automatically tracked thereafter.

2.3 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. As there is a great variety of classification algorithms reported in literature, this section will 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

neighbour classifiers, Fisher's linear discriminant (also know as linear discriminant analysis), support vector machines, artificial neural networks, AdaBoost, random forests, and hidden Markov models.

Nearest neighbour 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 in case 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.

3. Expression recognition from 3-D facial data

As discussed in the previous section, automatic recognition of facial expressions can be achieved through various approaches by using different facial expression representation methods and feature classification algorithms. In this section, more details are given for one of the techniques previously proposed by the authors. The method uses the shape space vector (SSV) as the feature for facial expression representation, which is derived from the statistical shape model (SSM) and has shown promising results in facial expression recognition (Quan et al., 2009b).

3.1 Statistical shape models

The basic principle of the SSM is to exploit a redundancy in structural regularity of the given shapes, thereby enabling a shape to be described with fewer parameters, i.e., reducing the dimensionality of the shapes presented in the original spatial domain. In order to achieve this dimensionality reduction, the principal component analysis (PCA) is usually used. Given a set of M 3-D facial surfaces, $\{\mathbf{Q}_i, i \in [1, M]\}$, where $\mathbf{Q}_i$ denotes a column vector representing the set of N corresponding vertices in the i-th face with $\mathbf{Q}_i \in R^{3N}$, the first step of the PCA is to construct the mean vector, denoted by $\overline{\mathbf{Q}}$, and calculated from all the available data.

$$\overline{\mathbf{Q}} = \frac{1}{M} \sum_{i-1}^{M} \mathbf{Q}_i \tag{1}$$

Let $\mathbf{C}$ be defined as a $3N \times 3N$ covariance matrix calculated from the training data set.

$$\mathbf{C} = \frac{1}{M} \sum_{i-1}^{M} (\mathbf{Q}_i - \overline{\mathbf{Q}})(\mathbf{Q}_i - \overline{\mathbf{Q}})^T \tag{2}$$

By building matrix $\mathbf{X}$ of "centred" shape vectors, with $(\mathbf{Q}_i - \overline{\mathbf{Q}})$ as the i-th column of matrix $\mathbf{X}$, covariance matrix $\mathbf{C}$ can be calculated as

$$\mathbf{C} = \mathbf{X}\mathbf{X}^T \tag{3}$$

Since the number of faces, M, in the training data set is usually much smaller than the number of data points on a face, the eigenvectors $\mathbf{u}_i'$ and eigenvalues λ_i' of $M \times M$ matrix $\mathbf{X}^T\mathbf{X}$ are calculated first, and the corresponding eigenvectors $\mathbf{u}_i$ and eigenvalues λ_i of the $\mathbf{X}\mathbf{X}^T$ are subsequently calculated from $\mathbf{u}_i = \mathbf{X}\mathbf{u}_i' / \|\mathbf{X}\mathbf{u}_i'\|$ and $\lambda_i = \lambda_i'$ (Vrtovec et al., 2004).
By using these eigenvalues and eigenvectors, the data points on any 3-D face in the training data set can be approximately represented using a linear model of the form

$$\hat{\mathbf{Q}} = \mathbf{W}\mathbf{b} + \overline{\mathbf{Q}} \tag{4}$$

where $\mathbf{W} = [\mathbf{u}_1, \ldots, \mathbf{u}_i, \ldots, \mathbf{u}_K]$ is a $3N \times K$ so called Shape Matrix of K eigenvectors, or "modes of variation", associated with the largest eigenvalues; and $\mathbf{b} = [b_1, \ldots, b_i, \ldots, b_K]^T$ is the shape space vector (SSV), which controls the contribution of each eigenvector, $\mathbf{u}_i$, in the approximated surface $\hat{\mathbf{Q}}$ (Cootes et al., 1995).
Projection of a facial surface denoted by $\mathbf{Q}$ on to the shape space using the shape matrix is then given by

$$\mathbf{b} = \mathbf{W}^T(\mathbf{Q} - \overline{\mathbf{Q}}) \tag{5}$$

Based on the SSM built using 450 randomly selected 3-D faces from the BU-3DFE database (see Section 4), Figure 8 shows variability of the first three elements of the SSV. The faces in each row have been obtained by adding to the mean face the corresponding eigenvector multiplied by a varying weight.

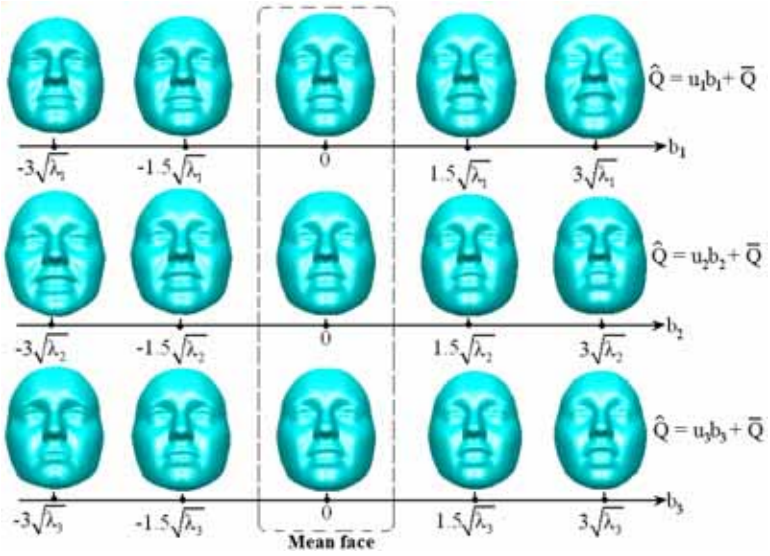


Fig. 8. Facial articulations controlled by the first three elements of the SSV.

In order to achieve the facial expression recognition using the extracted SSV, two processing stages are required, namely: model building and model fitting. The aim of the model building is to construct a 3-D SSM from a set of training data, whereas the model fitting is to estimate the SSV describing the new input 3-D faces so the SSM matches the new face.

3.2 Model building

A fundamental problem when building an SSM lies in correct determination of point correspondences between 3-D faces in the training data set. This is critical, because incorrect correspondences can either introduce too much variations or lead to invalid cases of the model (Cootes et al., 1995). The dense point correspondence estimation can be achieved in three steps: (i) facial landmark determination, (ii) thin-plate spline warping, and (iii) closest points matching. The first step is to identify the sparse facial landmarks on the training faces and a selected reference face; the second step is to warp the reference face to each of the training faces using the thin-plate spline transformation; and the last step is to use the warped reference face to estimate dense point correspondences for each of the training faces.

3.2.1 Facial landmark determination

This step is to identify the sparse facial landmarks which are around areas sensitive to facial expressions. These landmarks can be either extracted automatically or selected manually. Although automated identification of facial landmarks could be fast, without requiring user input, it is likely to fail at some point and has low accuracy due to data uncertainties. On the other hand, manual selection of facial landmarks can be used to handle data uncertainties, but it is tedious, time-consuming and user dependent.

A number of authors have proposed an automated identification method by using a template with annotated landmarks. Through registration of the template with a face, the landmarks pre-selected in the template can be propagated to the face (Frangi et al., 2002; Rueckert et al., 2003). Optical flow is another technique that can be used to identify corresponding landmarks between 3-D facial surfaces through their 2-D texture matching (Blanz & Vetter, 1999).

For the results presented later in this section the BU-3DFE database was used to construct the SSM. Each face in that database has 83 manually selected landmarks and these landmarks were used to establish a dense point correspondence. Figure 9 shows an example of faces in the BU-3DFE database with the selected facial landmarks.

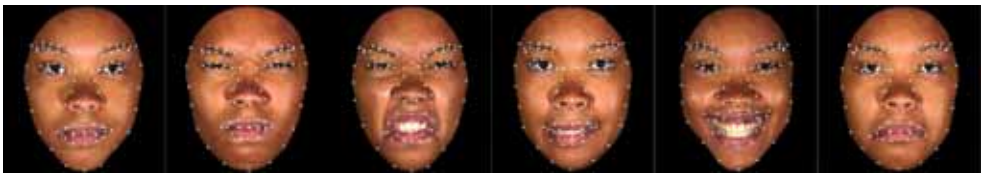


Fig. 9. BU-3DFE faces with selected facial landmarks (Yin et al., 2006).

3.2.2 Thin-plate spline warping

Using the sparse facial landmarks, the non-rigid transformations between the reference and training faces are estimated so that the reference face can be warped or deformed to match the shape of each training face. The thin-plate spline transformation model is used to achieve the non-rigid transformation (Bookstein, 1989), since the thin-plate spline is an effective tool for modeling deformations by combining global and local transformations and it has been applied successfully in several computer vision applications (Johnson & Christensen, 2002). In this step, the coefficients of the thin-plate spline transformation model are first computed using the sparse corresponding facial landmarks in two facial surfaces, and the transformation with the computed coefficients are applied to all the points of one facial surface so as to deform it to match to the other facial surface.

3.2.3 Closest point matching

After thin-plate spline warping, the reference face is deformed to match each of the training faces. Since the pose and shape of the deformed reference face are very close to the associated training face, the correspondences of all points between them can be determined using the Euclidean distance metric (Besl & McKay, 1992).

Let $\boldsymbol{P} = \{\mathbf{p}_i, i \in [1, N]\}$ be the point set of one of the training faces. The point $\hat{\mathbf{p}}(\mathbf{q})$ on that training face estimated to be closest to any given point $\mathbf{q}$ in the reference face is given by

$$\hat{\mathbf{p}}(\mathbf{q}) = \arg\min_{\mathbf{p} \in \boldsymbol{P}} d(\mathbf{p}, \mathbf{q}) \tag{6}$$

where $d(\mathbf{p}, \mathbf{q})$ is the Euclidean distance between points $\mathbf{p}$ and $\mathbf{q}$. When the correspondences between points on the reference face and each of the training faces are found, the whole training data set is aligned and brought into correspondence. With the aligned training data set, the SSM can be built directly using PCA as discussed in Section 3.1. Figure 10 shows examples on how the reference face is deformed to match five, randomly selected, training faces. These deformed reference faces are subsequently used to select points on the training faces which are in correspondence, across all the faces in the training database. It is seen that the shapes of deformed reference faces are close enough to the training faces to allow for selection of dense correspondences.

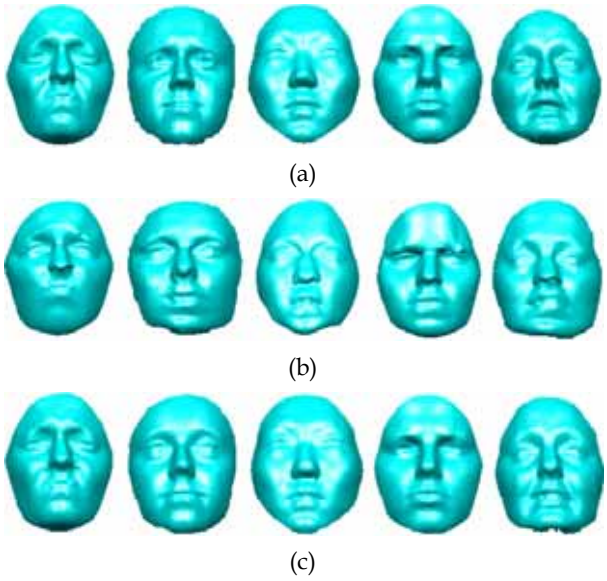


Fig. 10. Alignment of training images using thin-plate spline warping: (a) training faces, (b) reference faces after the thin-plate spline warping, and (c) training faces after re-selection of points which are in correspondence across all faces in the training database

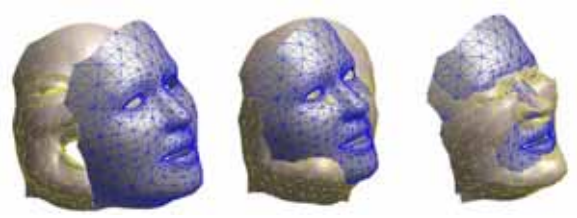
3.3 Model fitting

The model fitting process is treated as 3-D surface registration, which consists of initial model fitting followed by refined model fitting. While the former provides a global

alignment, using similarity transformation, between the new input face and the SSM model constructed, the latter refines the alignment by iteratively deforming the SSM model to match the input face.

3.3.1 Initial model fitting

The initial model fitting stage is achieved by using the iterative closest point (ICP) method with similarity transformation (Besl & McKay, 1992), and is implemented in two steps.
The first step is to roughly align the mean face, of the built SSM, and the input face based on their estimated centroids. This alignment does not have to be very accurate. An example of such alignment is shown in Figure 11(b), where the blue surface indicates the mean face of the SSM and the yellow surface represents the input face.



(a) (b) (c)

Fig. 11. An example of initial model fitting: (a) starting poses, (b) after rough alignment, and (c) after initial model fitting.

In the second step this rough alignment is iteratively refined by alternately estimating point correspondences and finding the best similarity transformation that minimises a cost function between the corresponding points. In the specific implementation of the algorithm described here the cost function is defined as

$$E = \frac{1}{N} \sum_{i-1}^{N} \left\| \mathbf{q}_i - (s\mathbf{R}\mathbf{p}_i + \mathbf{t}) \right\|^2 \tag{7}$$

where N is the number of point correspondence pairs between the two surfaces, $\mathbf{q}_i$ and $\mathbf{p}_i$ are respectively corresponding points on the model and new data surfaces, $\mathbf{R}$ is a 3x3 rotation matrix, $\mathbf{t}$ is a 3x1 translation vector, and s is a scaling factor. For the given point correspondence the pose parameters minimising cost function E can be found in a closed form (see (Umeyama, 1991) for more details).
The iterations in the second step could be terminated either when: (i) the alignment error between the two surfaces is below a fixed threshold, (ii) the change of the alignment error between the two surfaces in two successive iterations is below a fixed threshold, or (iii) the maximum number of iterations is reached. An option, which is often used, is to combine the second and third conditions.
Furthermore, since the search of the dense point correspondences is one of the most time-consuming tasks, a multi-resolution correspondence search has been developed to reduce the computation time. It starts searching for the point correspondences with a coarse point density, and gradually increases the point density in a series of finer resolutions as the registration error is reduced. To make this process more robust, random sampling is used to

sub-sample the original surface data to obtain a sequence of lower resolution data sets, with each sub-sampling reducing the number of the surface points by a quarter. At the beginning of the correspondence search, the lowest resolution data set is used, which allows for estimation of the large coarse motion of the model to match the input face. As the point density increases with the subsequent use of the finer data sets, it gradually restricts the amount of motion allowed and produces finer and finer alignment. Hence, the use of the multi-resolution correspondence search not only saves computation time, but also improves robustness of the alignment by avoiding local minima. Figure 11(c) shows an example result produced by the initial model fitting stage.

3.3.2 Refined model fitting

After the initial model fitting stage, the SSM and the input face are globally aligned. In the next stage, the SSM is iteratively deformed to better match the input face. This is achieved in the refined model fitting stage that consists of two main intertwined iterative steps, namely: shape and correspondence updates.

Based on the correspondence established in the initial model fitting stage, the new face denoted here by vector $\mathbf{P}^{(k)}$, with k representing the iteration index, and built by concatenating all surface vertices $\mathbf{p}_i^{(k)}$ into a single vector, is projected onto the shape space to produce the first estimate of the SSV, $\mathbf{b}^{(k)}$ and the first shape update, $\hat{\mathbf{Q}}^{(k)}$, as described in equations (8-9).

$$\mathbf{b}^{(k)} = \mathbf{W}^T\left(\mathbf{P}^{(k)} - \bar{\mathbf{Q}}\right) \tag{8}$$

$$\hat{\mathbf{Q}}^{(k)} = \mathbf{W}\mathbf{b}^{(k)} + \bar{\mathbf{Q}} \tag{9}$$

In the subsequent correspondence update stage, as described in equations (10-12), the data vertices, $\mathbf{p}_i^{(k)}$, are matched against vertices $\mathbf{q}_i^{(k)}$ from the updated shape, $\hat{\mathbf{Q}}^{(k)}$. This includes similarity transformation of the data vertices to produce new vertices position $\tilde{\mathbf{p}}_i^{(k+1)}$, and updating of the correspondence, represented by re-indexing of the data vertices, $\mathbf{p}_{j(i)}^{(k+1)}$.

$$\left\{\hat{\mathbf{R}},\ \hat{\mathbf{T}},\ \hat{s}\right\} = \arg\min_{\{\mathbf{R},\mathbf{T},s\}}\left(\sum_i \left\| \hat{\mathbf{q}}_i^{(k)} - s\mathbf{R}\mathbf{p}_i^{(k)} - \mathbf{T} \right\|^2\right) \tag{10}$$

$$\tilde{\mathbf{p}}_i^{(k+1)} = \hat{s}\hat{\mathbf{R}}\mathbf{p}_i^{(k)} + \hat{\mathbf{T}} \tag{11}$$

$$\tilde{\mathbf{p}}_i^{(k+1)} \rightarrow \mathbf{p}_{j(i)}^{(k+1)} \tag{12}$$

This process iterates till either the change of the alignment error between the updated shape and the transformed data vertices is below a fixed threshold, or the maximum number of iterations is reached. The final result of the described procedure is the shape state vector $\mathbf{b}^{(K)}$, where K denotes the last iteration, representing surface deformations between the model mean face and the input facial data.

From equation 8, it is seen that the size of the SSV is fixed which in turn determines the number of the eigenvectors from the shape matrix to be used. Although using a fixed size of the SSV can usually provide a reasonable final result of matching if a good initial alignment is achieved, it may mislead the minimisation of the cost function towards a local minimum in the refined model fitting stage. This can be explained by the fact that when a large size of the SSV is used, the SSM has a high degree of freedom enabling it to iteratively deform to a shape representing a local minimum, if it happens that the SSM was instantiated in the basing of this minimum. On the other hand, the SSM with a small size of the SSV has a low degree of freedom which constraints the deformation preventing it from converging to shapes associated with the local minima of the cost function, thereby limiting the ability of the SSM to accurately represent the data (Quan et al., 2010a). Figure 12 shows some examples of model matching failures caused by the size of the SSV being too large or too small.

(a) (b) (c) (d)

Fig. 12. Failed examples of model refinement by using a fixed size of SSV: (a) model, (b) input face, (c) using large size of SSV, and (d) using small size of SSV.

In order to solve the problem caused by the fixed size of SSV, the multi-level model deformation approach could be employed by using an adaptive size for the SSV. In the beginning of the refined model fitting stage, the SSV has a small size. Although it may results in a large registration error, it allows the algorithm to provide a rough approximation of the data. When the registration error is decreased, the size of SSV is gradually increased to provide more shape flexibility and allow the model to match the data. To implement this adaptive approach the fixed sized shape space matrix $\mathbf{W}$ in equations (8-9) needs to be replaced with $\mathbf{W}_k$, where index k signifies that the size of the matrix (number of columns) is changing (increasing) with the iteration index. A reasonable setting for the adaptive SSV is to start with

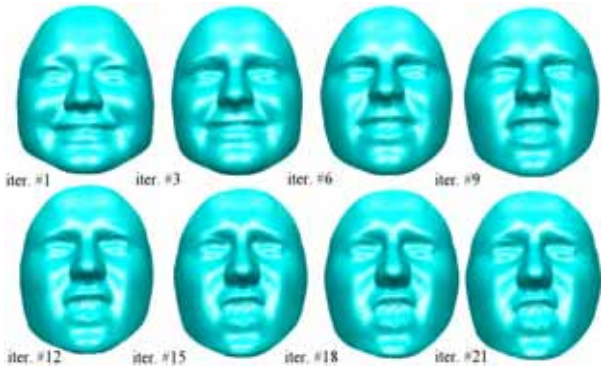


Fig. 13. Example of intermediate results obtained during iteration of refined model fitting with adaptive size of SSV.

an SSV size that enables the SSM to contain around 50% shape variance of the training data set (this corresponds to five eigenvectors in the case of 450 BU-3DFE training faces), to increase the SSV size by one in every two iterations during the refined model fitting stage, and to terminate the increase when the SSV size enables the SSM to contain over 95% shape variance. Some intermediate results taken from 21 iterations performed using the described procedure are shown in Figure 13. In that figure, the model and the input face used are the same as those shown in Figure 12. It can be seen that the multi-level model deformation approach not only provides a smooth transition between iterations during the model refinement stage but also enables the model to match the appropriate shape accordingly.

4. Facial expression databases

In order to evaluate and benchmark facial expression analysis algorithms, standardised data sets are needed to enable a meaningful comparison. Based on the type of facial data used by an algorithm, the facial expression databases can be categorised 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 is 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 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).

4.1 2-D image facial expression databases

CMU-PIE initial database was created at the Carnegie Mellon University in 2000. The database contains 41,368 images of 68 people, and the facial images taken from each person cover 4 different expressions as well as 13 different poses and 43 different illumination conditions (Sim et al., 2002). Due to the shortcomings of the initial version of CMU-PIE database, such as a limited number of subjects and facial expressions captured, the Multi-PIE database has been developed recently as an expansion of the CMU-PIE database (Gross et al., 2010). The Multi-PIE includes more than 750,000 images from 337 subjects, which were captured under 15 view points and 19 illumination conditions. MMI database includes hundreds of facial images and video recordings acquired from subjects of different age, gender and ethnic origin. This database is continuously updated with acted and spontaneous facial behaviour (Pantic et al., 2005), and scored according to the facial action coding system (FACS) (Ekman & Friesen, 1978). JAFFE database contains 213 images of 6 universal facial expressions plus the neutral expression (Lyons et al., 1999). This database was created with a help of 10 Japanese female models. Examples of the six universal expressions from that database are shown in Figure 14.

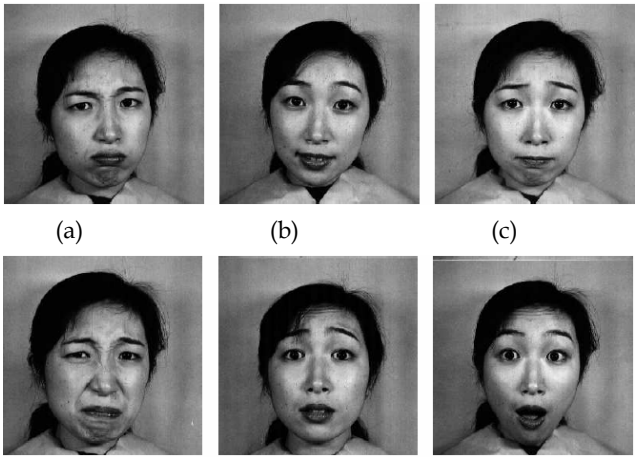


Fig. 14. Examples of the six universal facial expressions from the JAFFE database: (a) anger, (b) happiness, (c) sadness, (d) disgust, (e) fear, and (f) surprise.

4.2 2-D video facial expression databases

Cohn-Kanade AU Coded database is one of the most comprehensive 2-D video databases. Its initial release consisted of 486 video sequences from 97 subjects who were university students (Kanade et al., 2000). They ranged in age from 18 to 30 years old, with a variety of ethnic origins, including African-American, Asian and Hispanic. The peak expression for each sequence is coded according to the FACS (Ekman & Friesen, 1978). The second version of the database is an expansion of its initial release, which includes both posed and spontaneous facial expressions, with increased number of video sequences and subjects (Lucey et al., 2010). The third version of the database is planned to be published in 2011, and will contain, apart from the frontal view, additional synchronised recordings taken at 30 degrees angle. MPI database was developed at the Max Planck Institute for Biological Cybernetics (Pilz et al., 2006). The database contains video sequences of four different expressions: anger, disgust, surprise and gratefulness. Each expression was record from five different views simultaneously as shown in Figure 15. DaFEx database includes 1,008 short videos containing 6 universal facial expressions and the neutral expression from 8

Fig. 15. Examples of the MPI database showing five views: (a) left 45 degree, (b) left 22 degree, (c) front, (d) right 22 degree, (e) right 45 degree.

professional actors (Battocchi et al., 2005). Each universal facial expression was performed at three intensity levels. FG-NET database has 399 video sequences which were gathered from 18 individuals (Wallhoff, 2006). The emphasis in that database is put on recording a spontaneous behaviour of the subjects with emotions induced by showing participants suitable selected video clips. Similar to the DaFEx database, it covers 6 universal facial expressions plus the neutral expression.

4.3 3-D static facial expression databases

BU-3DFE 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. Figure 16 shows some examples from the BU-3DFE database.

Fig. 16. Examples from the BU-3DFE database (Yin et al., 2006).

ZJU-3DFED database is a static 3-D facial expression database, which was 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.

4.4 3-D dynamic facial expression databases

BU-4DFE database (Yin et al., 2008) is a 3-D dynamic facial expression database and 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). A few 3-D temporal samples from one of the BU-4DFE sequences are shown in Figure 17.

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 help from 10 graduates from 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

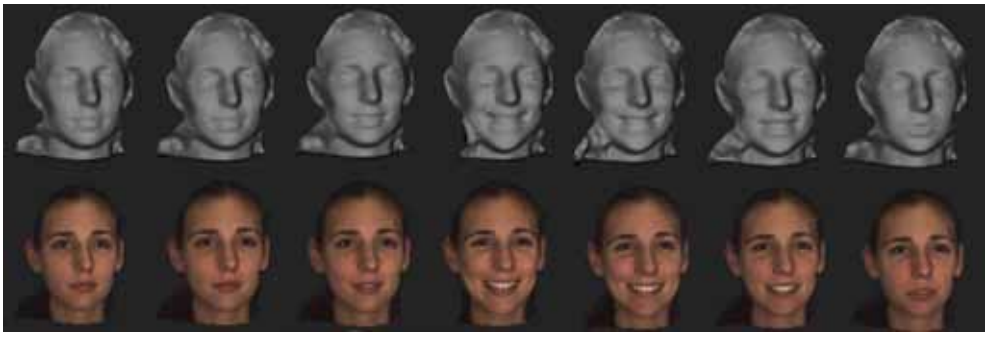


Fig. 17. Temporal 3-D samples from one of the sequences in the BU-4DFE database (Yin et al., 2008).

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. Results of this validation are summarised in the next section. Figure 18

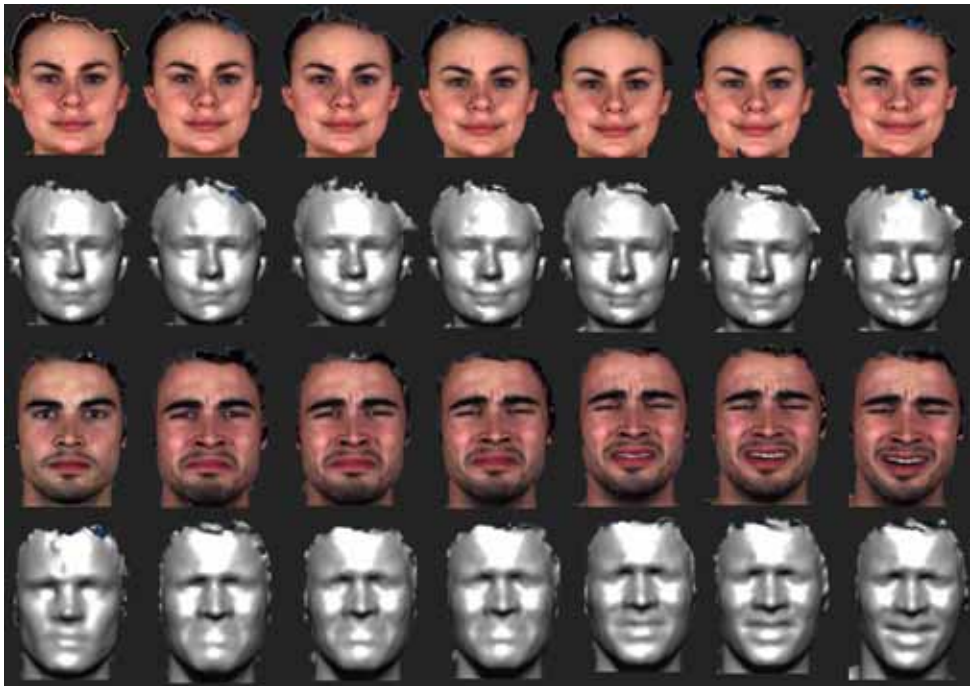


Fig. 18. Two examples from the ADSIPmark1 database showing happiness expression in normal intensity level (top two rows) and sadness expression in extreme intensity level (bottom two rows).

shows a couple of examples from the ADSIPmark1 database. That 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. A couple of examples of this currently extended ADSIP database are shown in Figure 19.

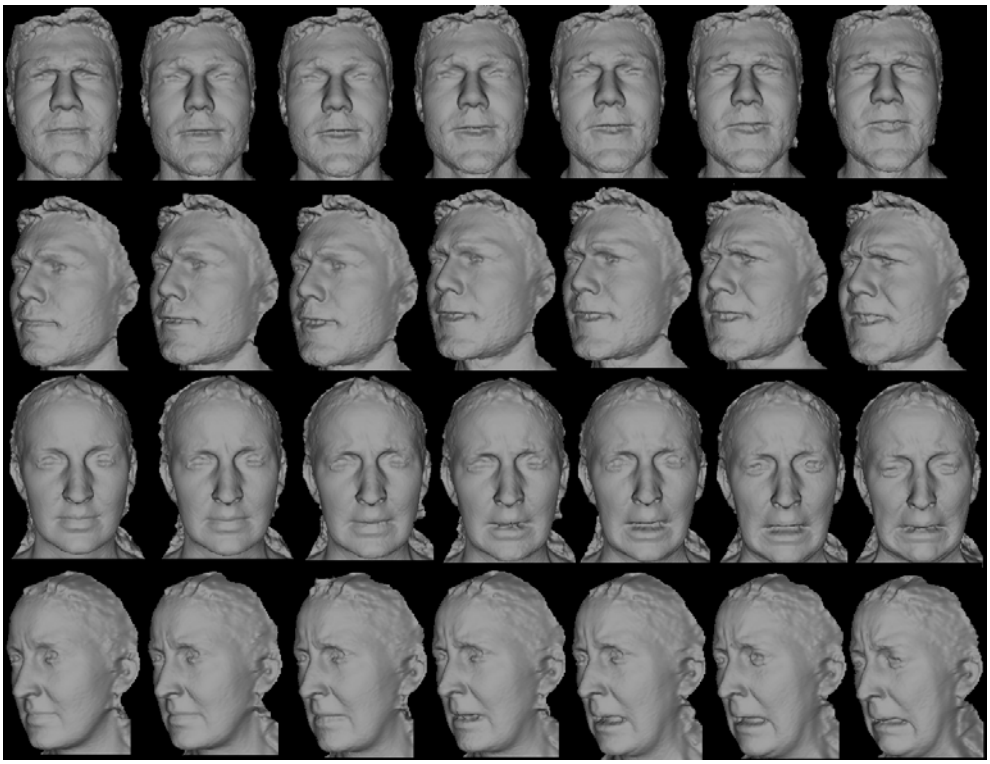


Fig. 19. Examples from the extended ADSIP database with anger expression in normal intensity level (top two rows with two different views) and fear expression in extreme intensity level (bottom two rows with two different views).

4.5 Database validation

Facial expressions are very subjective in nature. In other words, some of expressions are difficult to interpret and classify even for human observers who are normally considered as the "best classifier" for facial expressions. In order to validate the quality of a 3-D facial expression database, human observers have to be involved to see whether a particular facial expression, performed by a subject in response to a given instruction, is executed in a way which is consistent with the human perception of that expression. Use of human observers in validation of a facial database enables the assumed ground truth to be benchmarked, as it

provides a performance target for facial expression algorithms. It is expected that the performance of the best automatic facial expression recognition systems should be comparable with the human observers (Black & Yacoob, 1997; Wang & Yin, 2007).
Using the video clips recorded simultaneously with dynamic 3-D facial scan, the first part of the ADSIP database was assessed by 10 invited observers. They were the staff and students at the University of Central Lancashire. A bespoke computer program was designed to present the recorded video clips and collect the confidence ratings given by each of the observers. Participants were shown one video clip at a time and were asked to enter their confidence ratings against seven categories for expressions: anger, disgust, fear, happiness, pain, sadness and surprise, with the confidence ratings selected from the range of 0 to 100% for each category. To reflect possible confusions from observers about an expression for a given video clip, ratings could be distributed over the various expression categories as long as scores added up to 100%. Table 1 presents the confidence scores for each expression averaged over all video clips scored by the all observes. It can be seen that happiness expressions were given near perfect confidence scores, and anger, pain and fear were the worst rated with fear scored below 50%. Also, the 'normal' intensity level was somewhat better rated than 'mild', and 'extreme' was also somewhat better than 'normal'. Table 2 shows the confidence confusion matrix for the seven expressions. It can be seen that the observers were again very confident about recognising the happiness expression whereas the fear expression was often confused with the surprise expression (Frowd et al., 2009).

Intensity	Anger (%)	Disgust (%)	Fear (%)	Happiness (%)	Sadness (%)	Surprise (%)	Pain (%)	**Mean** (%)
Mild	47.5	51.5	43.3	90.3	72.9	72.9	57.4	57.9
Normal	56.6	78.3	41.5	94.3	75.6	75.6	62.0	65.7
Extreme	61.4	80.7	48.4	96.0	74.0	74.0	75.7	70.3
Mean (%)	55.2	70.2	44.4	93.5	74.2	74.2	65.0	64.6

Table 1. Mean confidence scores for seven expressions.

Input/Output	Anger (%)	Disgust (%)	Fear (%)	Happiness (%)	Sadness (%)	Surprise (%)	Pain (%)
Anger	**55.39**	26.03	5.19	0.00	5.13	5.31	2.94
Disgust	7.70	**68.86**	5.22	0.00	8.47	4.59	5.16
Fear	3.80	9.02	**46.90**	0.00	7.13	23.90	9.26
Happiness	0.27	0.98	0.71	**92.95**	1.15	2.35	1.59
Sadness	4.07	5.87	3.63	0.71	**74.15**	3.22	8.33
Surprise	0.60	7.54	21.84	1.04	2.46	**64.64**	1.88
Pain	4.94	9.45	9.46	2.30	18.96	3.85	**51.04**

Table 2. Confidence confusion matrix for the human observers.

5. Evaluation of expression recognition using BU-3DFE database

In order to characterise the performance of the SSV based representation for facial expression recognition, its effectiveness is demonstrated here in two ways. At first, the

distribution of the low-dimensional SSV-based features, extracted from the faces with various expressions is visualised, thereby showing the potential of the SSV-based features for facial expression analysis and recognition. Secondly, standard classification methods were used with the SSV-based features to quantify their discriminative characteristics.

5.1 Visual illustration of expression separability

Since it is difficult to visualise SSV-based features in a space with more than three dimensions, only the first three elements of the SSV are used to reveal its clustering characteristics and discriminative powers. An SSM was built using 450 BU-3DFE faces which were randomly selected from the database. The SSV-based features extracted from another 450 BU-3DFE faces which cover 18 individuals with six universal expressions were chosen for the visual illustration in 3-D shape space. It can be found that the SSV-based feature exhibits good expression separability even in a low-dimensional space, especially for those expression such as "anger vs. disgust", "disgust vs. surprise" and "fear vs. happiness". Examples of the expressions are given in Figure 20, where the SSV-based features representing these expressions are seen to form relatively well defined clusters in the low-dimensional shape space. Although some parts of the clusters slightly intersect with each other, the clusters can be identified easily.

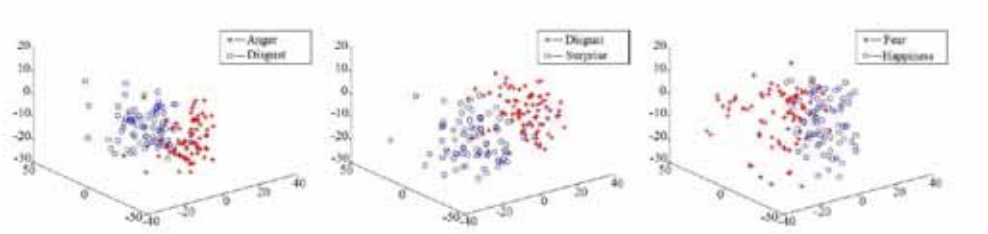


Fig. 20. Visualisation of expression separability in the low-dimensional SSV feature space: (a) anger vs. disgust, (b) disgust vs. surprise, and (c) fear vs. happiness.

5.2 Expression recognition

To better evaluate the discriminative characteristics of the vectors in the shape space, quantitative results of facial expression recognition are shown in this section. Several standard classification methods were employed in the experiments. They are linear discriminant analysis (LDA), quadratic discriminant classifier (QDC), and nearest neighbour classifier (NNC) (Duda, 2001; Nabney, 2004). 900 faces from the BU-3DFE database were used for testing, which were divided into six subsets with each subset containing 150 randomly selected faces. During the experiment, one of the subsets was selected as the test subset while the remaining subsets were used for learning. Such experiment was repeated six times, with the different subset selected as the test subset each time. Table 3 shows the averaged results for expression recognition achieved by the three classifiers. It can be seen that the LDA achieved the highest recognition rate of 81.89%.

Table 4 shows the confusion matrix of the LDA classifier. It can be seen that the anger, happiness, sadness and surprise expressions are all classified with above 80% accuracy, whereas the fear expression is only classified correctly for around 73%. This is consistent with the validation results for the ADSIP database discussed in Section 4.5, which showed that the fear expression is often confused with other expressions by the human observers.

Classifiers	LDA	QDC	NNC
Recognition rate	**81.89**%	80.11%	78.53%

Table 3. Average recognition rate of faces from BU-3DFE database using the SSV-based features.

Input/Output	Anger (%)	Disgust (%)	Fear (%)	Happiness (%)	Sadness (%)	Surprise (%)
Anger	**82.64**	3.48	4.17	3.47	4.86	1.39
Disgust	7.64	**78.47**	3.48	5.56	2.08	2.78
Fear	4.27	3.57	**72.59**	12.50	5.66	1.39
Happiness	2.78	5.56	8.33	**83.33**	0.00	0.00
Sadness	4.17	3.47	11.11	0.00	**81.25**	0.00
Surprise	0.00	0.00	4.17	2.78	0.00	**93.06**

Table 4. Confusion matrix of the LDA classifier for BU-3DFE database.

6. Evaluation of expression recognition using ADSIP database

The performance evaluation using the ADSIP database is an extension of that based on the BU-3DFE database, which aims to check the stability of the proposed methodology with a statistical shape model built from one database and tested using a different database. Furthermore, it is also used to assess the performance of the proposed algorithm against human observers.

As explained in Section 5.2, the statistical shape model was built based on 450 faces randomly selected from the BU-3DFE database. One hundred static 3-D faces were randomly chosen as the testing faces from the 210 dynamic 3-D facial sequences in the ADSIP database, with each one occurring approximately at the maximum of the expression. The selected test set contained 10 subjects with six different facial expressions at various intensities. For each testing face represented by its SSV, the three classifiers LDA, QDC and NNC, were used for the facial expression classification.

Table 5 shows the average results obtained from expression recognition using the three classifiers. It can be seen that the LDA again achieved the highest recognition rate of 72.84%. Although this result is worse than the result obtained using the test faces from the BU-3DFE database, it is around 2.5% higher than the mean recognition rate achieved by the human observers for those 'extreme' expressions in the same ADSIP database shown in Table 1, and it is much better than the human observers' mean recognition rates for the other two intensity levels of expressions.

Classifiers	LDA	QDC	NNC
Recognition rate	**72.84**%	69.31%	70.54%

Table 5. Average recognition rate of the SSV-based feature from ADSIP database.

The confusion matrix of the LDA is shown in Table 6. It can be seen that the happiness is the most recognisable expression and followed by the surprise expression. Fear and sadness are the expressions which are often confused with others.

Input/Output	Anger (%)	Disgust (%)	Fear (%)	Happiness (%)	Sadness (%)	Surprise (%)
Anger	**73.08**	11.54	0.00	1.92	13.46	0.00
Disgust	11.54	**71.15**	9.62	1.92	1.92	3.85
Fear	1.92	11.54	**55.77**	17.31	13.46	0.00
Happiness	2.08	0.00	10.83	**87.08**	0.00	0.00
Sadness	22.92	0.00	10.42	0.00	**66.67**	0.00
Surprise	0.00	6.25	8.33	0.00	2.08	**83.33**

Table 6. Confusion matrix of the LDA classifier for ADSIP database.

7. Discussion and conclusions

This chapter has introduced 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 normalisation, feature extraction and subsequent classification. The available facial expression databases were also introduced to provide complete information about available options for algorithm validation. To make these ideas clearer one of the algorithm, using shape space vector (SSV) calculated for 3D static facial data, was described in detail. The algorithm consists of model building and model fitting. In the model building stage, a statistical shape model (SSM) is constructed from a set of training data. In the model fitting stage, the built SSM is aligned and matched to the input faces through iterative estimation of pose and shape parameters, and such estimated shape parameters (embedded in the SSV) are used as a feature vector for facial expression recognition.

In order to examine the facial expression recognition performance offered by the SSV representation method, two recently developed 3-D facial expression databases were used. The evaluation results based on the BU-3DFE database show that the SSV based representation method is capable of simulation and interpretation of 3-D human facial expressions. The test preformed with the ADSIP database indicates that the method can be used with data collected using different acquisition protocols. More significantly, the recognition rates for different facial expressions obtained using the SSV based representation are shown to be similar to those obtained using the human observers.

The temporal information plays a vital role in the facial expression recognition. An analysis of a dynamic facial sequence instead of a single image is likely to improve the accuracy and robustness of the facial expression recognition systems (Sun & Yin, 2008). To reflect this new trend in facial data analysis the chapter includes short information about available 3D dynamic datasets.

8. References

Aizawa, K. & Huang, T. S. (1995). Model-based image coding: advanced video coding techniques for very low bit-rate applications. *Proceedings of IEEE*, Vol.83(2), pp. 259-271.

Battocchi, A.; Pianesi F. & Goren-Bar, D. (2005). DaFEx: database of facial expressions. Intelligent Technologies for Interactive Entertainment, *Lecture Notes in Computer Science*, Vol.3814(2005), pp. 303-306.

Bartlett, M. S.; Littlewort, G.; Fasel, I. & Movellan J. R. (2003). Real time face detection and facial expression recognition: development and applications to human computer interaction. *CVPR workshop for HCI*.

Bernardini, F. & Rushmeier, H. E. (2002). The 3D model acquisition pipeline. *Computer Graphics Forum*, Vol.21(2), pp. 149-172.

Besl, P. J. & McKay, N. D. (1992). A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.14(2), pp. 239-256.

Bishop, C., M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Black, M. J.; Fleet, D. J. & Yacoob, Y. (1998). A framework for modeling appearnce change in image sequence. *6th International Conference on Computer Vision*.

Black, M. J. & Yacoob, Y. (1997). Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Journal of Computer Vision*, Vol.25(1), pp. 23-48.

Blake, A. & Isard, M. (1998). *Active Contours*. Springer-Verlag Berlin and Heidelberg.

Blanz, V. & Vetter, T. (2003). Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.25(9), pp. 1063-1074.

Blanz, V. & Vetter, T. (1999). A morphable model for the synthesis of 3-D faces. *SIGGRAPH*, pp. 187-194.

Bookstein, F. L. (1989). Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.11(6), pp. 567-585.

Brahnam, S.; Chuang, C.; Shih, F. Y. & Slack M. R. (2006). Machine recognition and representation of neonatal facial displays of acute pain. *Artificial Intelligence in Medicine*, Vol.36(3), pp. 211-222.

Breiman, L. (2001). Random forests. *Machine Learning*, Vol. 45(1), pp. 5-32.

Cohen, I.; Sebe, N.; Garg, A.; Chen, L. & Huang T. S. (2003). Facial expression recognition from video sequences: temporal and static modelling. *Computer Vision and Image Understanding*, Vol. 91, pp. 160-187.

Cootes, T. F.; Taylor, C. J.; Cooper, D. H. & Graham, J. (1995). Active shape models – their training and application. *Computer Vision and Image Understanding*, Vol.61(1), pp. 38-59.

Cootes, T. F. & Taylor, C. J. (1996). Data driven refinement of active shape model search. *British Machine Vision Conference*, pp. 383-392.

Cootes, T. F.; Edwards, G. J. & Taylor, C. J. (1998). Active appearance models. *European Conference on Computer Vision*, pp. 484-498.

Curless, B. (2000). From range scans to 3D models. *ACM SIGGRAPH Computer Graphics*, Vol.33(4), pp. 39-41.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. J. Murray, London

Dimensional Imaging (2010). http://www.di3d.com.

Duda, R. O. (2001). *Pattern Classification*, second edition, John Wiley & Sons Inc.

Edwards, G. J.; Cootes, T. F. & Taylor, C. J. (1998). Face recognition using active appearance models. *5th European Conference on Computer Vision*, pp. 581-595.

Ekman, P. & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Personality and Social Psychology*, Vol.17(2), pp. 124-129.

Ekman, P. & Friesen, W. V. (1978). *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press.

Eisert, P. & Girod, B. (1998). Analyzing facial expressions for virtual conferencing. *IEEE Computer Graphics & Applications*, Vol.18(5), pp. 70-78.

Essa, I. A. & Pentland, A. P. (1997). Coding, analysis, interpretation and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19, pp. 757-763.

FaceGen (2003). http://www.facegen.com. Singular Inversions.

Fasel, B. & Luettin, J. (2003). Automatic facial expression analysis: a survey. *Pattern Recognition*, Vol.36(1), pp. 259-275.

Feldmar, J. & Ayache, N. (1996). Rigid, affine and locally affine registration of free-form surfaces. *International Journal of Computer Vision*, Vol.18(2), pp. 99-119.

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, Vol.7, pp. 179-188.

Flanelli, G.; Yao, A.; Noel, P.-L.; Gall, J. & Van Gool, L. (2010). Hough forest-based facial expression recognition from video sequences. *International Workshop on Sign, Gesture and Activity (SGA) 2010, in conjunction with ECCV 2010*.

Frangi, A.; Ruechert, D.; Schnabel, J. & Niessen, W. (2002). Automatic construction of multiple-object three-dimensional statistical shape models: application to cardia modeling. *IEEE Transaction on Medical Imaging*, Vol.21(9), pp. 1151-1166.

Freund, Y. & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Science*, Vol.55, pp. 119-139.

Frowd, C. D.; Matuszewski, B. J.; Shark, L.-K. & Quan, W. (2009). Towards a comprehensive 3D dynamic facial expression database. *9th WSEAS International Conference on Multimedia, Internet and Video Technology (MIV'09)*, Budapest, Hungary.

Gross, R.; Matthews I.; Cohn, J.; Kanade, T. & Baker, S. (2010). Multi-PIE. *Image and Vision Computing*, Vol.28(5), pp. 807 – 813.

Hager, J. C.; Ekman, P. & Friesen, W. V. (2002). *Facial Action Coding System*. Salt Lake City, UT.

Heisele, B.; Serre, T.; Pontil, M. & Poggio, T. (2001). Component-based face detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 657-662.

Hoch, M.; Fleischmann, G. & Girod, B. (1994). Modeling and animation of facial expression based on B-splines. *Visual Computer*, pp. 87-95.

Hong, H.; Neven, H. & Von der Malsburg, C. (1998). Online facial expression recognition based on personlized galleries. *2nd International Conference on Automatic Face and Gesture Recognition*, pp. 354-359.

Hong, T.; Lee, Y.-B.; Kim, Y.-G. & Hagbae, K. (2006). Facial expression recognition using active appearance model. *Advance in Neural Network, Lecture Notes in Computer Science*, Vol.3972, pp. 69-76.

Huang, T. M.; Kecman, V. & Kopriva, I. (2006). *Kernel Based Algorithms for Mining Huge Data Sets, Supervised, Semi-supervised, and Unsupervised Learning*. Springer-Verlag, Berlin.

Isgro, F.; Trucco, E.; Kauff, P. & Schreer, O. (2004). Three-dimensional image processing in the future of immersive media. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.14(3), pp. 288-303.

Ji, Y. & Idrissi, K. (2009). Facial expression recognition by automatic facial parts position detection with boosted-LBP. *5th International Conference on Signal Image Technology and Internet Based Systems*, Marrakech, Marocco, pp. 28-35.

Johnson, H. & Christensen, G. (2002). Consistent landmark and intensity based image registration. *IEEE Transactions on Medical Imaging*, Vol.21(5), pp. 450-461.

Jones, M. & Viola, P. (2003). Face recognition using boosted local feature. *International Conference on Computer Vision*.

Kanade, T.; Cohn, J. F. & Tian, Y. (2000). Comprehensive database for facial expression analysis. *IEEE International Conference on Automatic Face and Gesture Recognition*.

Kobayashi, H. & Hara, F. (1997). Facial interaction between animated 3D face robot and human beings. *International Conference on Systems*, *Man*, *Cybernetics*, pp. 3732-3737.

Lanitis, A.; Taylor, C. J. & Cootes, T. F. (1997). Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19(7), pp. 743-756.

Liao, S.; Fan, W.; Chung, A. & Yeung, D.-Y. (2006). Facial expression recognition using advanced local binary patterns, Tsallis entropies and global appearance feature. *IEEE International Conference on Image Processing*, pp. 665-668.

Lisetti, C. L. & Schiano, D. J. (2000). Automatic facial expression interpretation: where human-computer interaction, artificial intelligence and cognitive science intersect. *Pragmatic and Cognition*, Vol.8(1), pp. 185-235.

Littlewort, G.; Bartlett, M. S.; Fasel, I.; Susskind, J. & Movellan, J. (2005). Dynamics of facial expression extracted automatically from video. *Image and Vision Computing*, Vol.24, pp. 615-625.

Lyons, M. J.; Budynek, J. & Akamatsu, S. (1999). Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.12, pp. 1357-1362.

Lucey, P.; Cohn, J., F.; Kanade, T.; Saragih, J.; Ambadar, Z. & Matthews, I. (2010). The extended Cohn-Kanade Dataset (CK+): a complete dataset for action unit and emotion-specified expression. *Computer Vision and Pattern Recognition Workshops (CVPRW)*, San Francisco, pages 94-101.

Maalej, A.; Amor, B. B.; Daoudi, M.; Srivastava, A. & Berretti, S. (2010). Local 3D shape analysis for facial expression recognition. *20th International Conference on Pattern Recognition ICPR*, pp. 4129-4132.

Md. Sohail, A. S. & Bhattacharya, P. (2007). Classification of facial expression using k-nearest neigbhour classifier. *Lecture Note in Computer Science*, Vol.4418(2007), pp. 555-566.

Morishima, S. & Harashima, H. (1993). Facial expression synthesis based on natrual voice for virtual face-to-face communication with machine. *Virtual Reality Annual International Symposium*.

Niese, R.; Al-Hamadi, A. & Michaelis, B. (2007). A novel method for 3D face detection and normalization. *Journal of Multimedia*, Vol.2(5), pp. 1-12.

Nabney, I. T. (2004). *NETLAB: Algorithms for Pattern Recognition*, Springer-Verlag London.

Nusseck, M.; Cunningham, D. W.; Wallraven, C. & Bulthoff, H. H. (2008). The contribution of different facial regions to the recognition of conversational expressions. *Journal of Vision*, Vol.8, pp. 1-23.

Ojala, T.; Pietikainen, M. & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24(7), pp. 971-987.

Padgett, C.; Cottrell, G. & Adolphs, R. (1996). Categorical perception in facial emotion classification. *18th Annual Cognitive Science Conference*, pp. 249-253.

Pantic, M. & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: the state of art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.22(12), pp. 1424-1445.

Pantic, M.; Valstar, M. F.; Rademaker, R. & Maat, L. (2005). Web-based database for facial expression analysis. *Proc. IEEE International Conference on Multimedia and Expo (ICME'05)*, Amsterdam, Netherlands.

Park, H. & Park, J. (2004). Analysis and recognition of facial expression based on point-wise motion energy. *Lecture Notes in Computer Sciences*, Vol.3212(2004), pp. 700-708.

Pearson, D. E. (1995). Developments in model-based video coding. *Proceedings of IEEE*, Vol.83(6), pp. 892-906.

Pentland, A. (2000). Looking at people: sensing for ubiquitous and wearable computing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.22(1), pp. 107-119.

Pilz, K.; Thornton, I. M. & Bulthoff, H. H. (2006). A search advantage for faces learned in motion. *Experimental Brain Research*, Vol.171(4), pp. 436-447.

Pollak, S. D. & Sinha, P. (2002). Effects of early experience on children's recognition of facial displays of emotion. *Developmental Psychology*, Vol.38(5), pp. 784-791.

Quan, W. ; Matuszewski, B. J. ; Shark, L.-K. & Ait-Boudaoud, D. (2007a). Low dimensional surface parameterisation with applications in biometrics. *IEEE 4th International Conference on Biommedical Visualisation (MediViz07)*, pp. 15-20.

Quan, W.; Matuszewski, B. J.; Shark, L.-K. & Ait-Boudaoud, D. (2007b). 3-D facial expression representation using B-spline statistical shape model, *BMVC Vision*, *Video and Graphics Workshop*, Warwick.

Quan, W. ; Matuszewski, B. J. ; Shark, L.-K. & Ait-Boudaoud, D. (2008). 3-D facial expression representation using statistical shape models. *BMVA Symposium on 3D Video Analysis*, *Display and Applications*, Royal Academy of Engineering, London.

Quan, W. (2009a). 3-D facial expression representation using statistical shape models. PhD Thesis, University of Central Lancashire.

Quan, W.; Matuszewski, B. J.; Shark, L.-K. & Ait-Boudaoud, D. (2009b). 3-D facial expression biometrics using statistical shape models. *Special Issue on Recent Advances in Biometric Systems: A Signal Processing Perspective*, *EURASIP Journal on Advances in Signal Processing*, Vol.2009, Article ID 261542, pp. 1-17.

Quan, W.; Matuszewski, B. J. & Shark L.-K. (2010a). A statistical shape model for deformable surface registration. *International Conference on Computer Vision Theory and Applications*, Angers, France.

Quan, W.; Matuszewski, B. J. & Shark L.-K. (2010b). Improved 3-D facial representation through statistical shape model. *IEEE International Conference on Image Processing (2010 ICIP)*, pp. 2433-2436.

Ramanathan, S.; Kassim, A.; Venkatesh, Y. V. & Wah, W. S. (2006). Human facial expression recognition using a 3D morphable model. *IEEE International Conference on Image Processing*.

Ruechert, D.; Frangi, A. F. & Schnabel, J. A. (2003). Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, Vol.22(8), pp. 1014-1025.

Saxena, A.; Anand, A. & Mukerjee, A. (2004). Robust facial expression recognition using spatially localized geometric model. *International Conference on Systemic, Cybernetics and Informatics*, pp. 124-129.

Steffens, J.; Elagin, E. & Neven, H. (1998). PresonSpotter – fast and robust system for human detection tracking and recognition. *2nd International Conference on Face and Gestuure Recognition*, pp. 516-521.

Sim, T.; Baker, S. & Bsat, M. (2002). The CMU pose, illumination, and expression (PIE) database. *5th IEEE International Conference on Automatic Face and Gesture Recognition*.

Sun, Y. & Yin, L. (2008). Facial expression recognition based on 3D dynamic range model sequences. *10th European Conference on Computer Vision (ECCV08)*.

Suwa, M.; Sugie, N. & Fujimora, K. (1978). A preliminary note on pattern recognition of human emotional expression. *4th International Joint Conference on Pattern Recognition*, pp. 408-410.

Trier, O. D.; Taxt, T. & Jain, A. (1997). Recognition of digits in hydrographic maps: binary vs. topographic analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19(4), pp. 399-404.

Umeyama, S. (1991). Least-square estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.13(4), pp. 376-380.

Vural, E.; Cetin, M.; Littlewort, G.; Bartlett, M. & Movellan, J. (2007). Drowsy driver detection through facial movement analysis. *Lecture Note in Computer Science*, No.4796, pp. 6-18.

van Dam, A. (2000). Beyond WIMP, *IEEE Computer Graphics and Applications*, Vol.20(1), pp. 50-51.

Vrtovec, T.; Tomaževič, D.; Likar, B.; Travnik, L. & Pernuš, F. (2004). Automated Construction of 3D statistical shape models. *Image Analysis and Stereology*, Vol.23, pp. 111-120.

Wallhoff, F. (2006). Facial expressions and emotion database. www.mmk.ei.tum.de/~waf/fgnet/feedtum.html, Technical University Munich.

Wang, J.; Yin, L.; Wei, X. & Sun, Y. (2006a). 3D facial expression recognition based on primitive surface feature distribution. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2006)*, pp. 17-22.

Wang, Y.; Pan, G.; Wu, Z. & Wang, Y. (2006b). Exploring facial expression effects in 3D face recognition using partial ICP. *Lecture Notes in Computer Sciences*, Vol.3851(2006), pp. 581-590.

Wang, J. & Yin, L. (2007). Static topographic modeling for facial expression recognition and analysis. *Computer Vision and Image Understanding*, Vol.108(1-2), pp. 19-34.

Whitehill, J.; Littlewort, G.; Fasel, I.; Bartlett, M. & Movellan, J. (2009). Towards practical smile detection. *IEEE Transactions on Pattern Analysis and machine Intelligence*, Vol.31(11), pp. 2106-2111.

Wimmer, M.; MacDonald, B. A.; Jayamuni, D. & Yadav, A. (2008). Facial expression recognition for human-robot interaction: a prototype. *2nd International Conference on Robot Vision*, pp. 139-152.

Yin, L. ; Loi, J. & Xiong, W. (2004). Facial expression representation and recognition based on texture augmentation and topographic masking. *12th Annual ACM International Conference on Multimedia*, pp. 236-239.

Yin, L.; Wei, X.; Sun, Y.; Wang, J. & Rosato, M. (2006). A 3D facial expression database for facial behavior research. *7th International Conference on Automatic Face and Gesture Recognition (FG2006)*, IEEE Computer Society TC PAMI, pp. 211-216.

Yin, L.; Chen, X.; Sun, Y.; Worm T. & Reale, M. (2008). A High-Resolution 3D Dynamic Facial Expression Database. *8th International Conference on Automatic Face and Gesture Recognition (FGR08)*, IEEE Computer Society TC PAMI. Amsterdam, The Netherlands.

Yousefi, S.; Nguyen, M. P.; Kehtarnavaz, N. & Cao, Y. (2010). Facial expression recognition based on diffeomorphic matching, *17th IEEE International Conference on Image Processing*, pp. 4549-4552.

Zhang, Z. (1994). Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, Vol.13(2), pp. 119-152.

Zhang, S. & Huang, P. (2006). High-resolution, real-time 3-D shape measurement. *Optical Engineering*, Vol.45(12), pp. 123601-1-8.