We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

7

Modelling, Classification and Synthesis of Facial Expressions

Jane Reilly, John Ghent and John McDonald
Computer Vision and Imaging Laboratory
Department of Computer Science
National University of Ireland Maynooth,
Ireland

1. Introduction

The field of computer vision endeavours to develop automatic approaches to the interpretation of images from the real world. Over the past number of decades researchers within this field have created systems specifically for the automatic analysis of facial expression. The most successful of these approaches draw on the tools from behavioural science. In this chapter we examine facial expression analysis from both a behavioural science and a computer vision perspective. First we will provide details of the principal approach used in behavioural science to analyze facial expressions. This will include an overview of the evolution of facial expression analysis, where we introduce the field of facial expression analysis with Darwin's initial findings (Darwin, 1872). We then go on to show how his findings were confirmed nearly 100 years later by Ekman *et al.* (Ekman et al., 1969). Following on from this we provide details of recent works investigating the appearance and dynamics of facial expressions.

Given these foundations from behavioural science, we appraise facial expression analysis from a computer vision perspective. Here researchers attempt to create automated computational models for facial expression classification and synthesis. This chapter is divided into three sections, each of which deals with a different, but related problem within the field of facial expression analysis:

I. Classification of facial expressions:

Facial expressions play a major role in human communication. A system capable of interpreting and reacting to facial expressions would represent a major advancement in the field of human computer interaction. Although initial investigations into facial expressions focussed on classifying the six primary expressions (joy, sadness, fear, surprise, anger and disgust), in more recent years the focus has changed, due to a need for a consistent representation of facial expressions. Researchers began to concentrate on classifying expressions in terms of the individual movements that make up facial expressions. In this section, we appraise the current state of the art in facial expression classification, and provide details of our research to date in this area.

II. Modelling the dynamics of facial expressions:

Recent research has shown that it is not just the particular facial expression, but also the associated dynamics that are important when attempting to decipher its meaning. The

dynamics of facial expressions, such as the timing, duration and intensity of facial activity plays a critical role in their interpretation. Once we have classified which expression has been portrayed, we subsequently extract information regarding the dynamics of the expression.

III. Synthesis of facial expressions:

In the final section of this chapter we describe a technique which we have developed that allows for photo-realistic images of a person depicting a desired expression to be synthesised in real-time once a neutral image of the subject is present (Ghent, 2005a; Ghent, 2005b). This is achieved by applying machine learning techniques to the modelling of *universal facial expression mapping functions* (i.e. functions that map facial expressions independent of identity). We also demonstrate how the representation of expression used allows the intensity of the output expression to be varied. This ability to vary intensity means that the technique can also be used to generate image sequences of expression formation.

2. Behavioural and computational approaches for facial expression analysis

In this section we first review the state-of-the-art in expression analysis from a behavioural science perspective. Subsequent to this we detail computer vision based approaches for the classification and synthesis of facial expressions.

2.1 Facial expressions from a behavioural science perspective

Darwin first recognised the importance of facial expressions and the role which they play in human communication in 1872 (Darwin, 1872). During the subsequent years as behavioural scientists sought a means to objectively measure facial expressions, many different techniques and methodologies for describing facial expressions were developed (see (Fasel & Luettin, 2003) for a comprehensive review). Recent research has shown that it is not just the expression itself, but also its dynamics that are important when attempting to interpret its underlying meaning. In this section we provide details of the principal approach used in behavioural science to analyze, interpret and encode facial expressions.

2.1.1 Facial expression analysis

Within the field of facial expression analysis there has been a significant amount of research carried out investigating the six prototypical expressions (anger, fear, sadness, joy, surprise, and disgust). This focus is partially due to observations made by Darwin in 1872, when he observed that although people from different countries spoke different languages, from looking at their faces he could interpret their feelings and emotions. This idea coincided with his theory of evolution and the continuity of the species (Darwin, 1872).

In the subsequent years, the field of facial expression analysis remained an active research area within behavioural science. However, from a computer vision perspective, the research carried out by Ekman *et al.* has come to represent the seminal works on facial expression analysis. Approximately 100 years after Darwin's initial findings, Ekman *et al.* established that there are six universal facial expressions. They did so by performing a number of experiments to discover if American and Japanese students responded in a similar manner to a series of emotion evoking films. The reactions of the students to the films were captured and labelled as being one of the six primary expressions. While, they found that the students

had similar responses to the films, Ekman *et al.* were unable prove conclusively that the student's responses were not as a result of exposure to the American culture (Ekman & Friesen, 1969).

In the 1970's Ekman *et al.* came across footage of a remote tribe from New Guinea called the South Fore, which had little or no exposure to modern culture. Upon analysing the video footage they recognised the expressions portrayed by tribesmen as being variations of the six primary expressions. Ekman *et al.* subsequently travelled to New Guinea where they performed a number of studies. For example, they showed the tribesmen photographs of caucasians depicting the primary facial expressions. As anticipated, the South Fore tribe were able to recognise these expressions. Ekman *et al.* then asked the tribesmen to show what their face would be like if, for example, *'friends had come to visit'*, or *'their son was injured'*. Using the results of these experiments, they were able to prove that the South Fore people responded to and performed the facial expressions as anticipated (Ekman et al., 1971).

What distinguished these experiments from previous work was that as the South Fore people had not been exposed to modern culture, Ekman *et al.* were able to conclusively prove that these six primary facial expressions were not culturally determined. These results support Darwin's hypothesis that these primary facial expressions are biological in origin, and are expressed and perceived in a similar way across all cultures.

In everyday life, while these primary expressions do occur frequently, when analysing human interaction and conversation, researchers have found that displays of emotion are more often communicated by small subtle changes in the face's appearance (Ambadar et al., 2005). As a consequence, the focus of research into facial expressions has shifted from concentrating on identifying the six basic expressions to identifying the individual movements that make up an expression. The *Facial Action Coding System* (FACS) provides a means for coding expression in terms of these elementary facial actions. In Section 2.2.3 we look at the FACS in detail.

2.1.2 Facial expressions dynamics

Recent research has shown that it is not only the expression itself, but also its dynamics that are important when attempting to decipher its meaning (Cohn et al., 2005). The dynamics of facial expression can be defined as the intensity of the AUs coupled with the timing of their formation. Ekman *et al.* suggest that the dynamics of facial expression provides unique information about emotion that is not available in static images (Ekman & Friesen, 2002).

However, according to Ambadar *et al.*, only a few investigators have examined the impact of dynamics in deciphering faces. These studies were largely unsuccessful due to their reliance on extreme facial expressions. Ambadar *et al.* also highlighted the fact that facial expressions are frequently subtle. They found that subtle expressions that were not identifiable in individual images suddenly became apparent when viewed in a video sequence (Ambadar et al., 2005).

There is now a growing body of psychological research that argues that these dynamics are a critical factor for the interpretation of the observed behaviour. Zheng *et al.,* state that in many cases, an expression sequence can contain multiple expressions of different intensities sequentially, due to the evolution of the subject's emotion over time (Zheng, 2000). Despite the fact that facial expressions can be either subtle or pronounced in their appearance, and fleeting or sustained in their duration, most of the studies to date have focused on investigating static displays of extreme posed expressions rather than the more natural

spontaneous expressions. The following definitions explain the differences between *posed* and *spontaneous* facial expressions:

- **Posed facial expressions** are generally captured by asking subjects to perform specific facial actions or expressions. They are usually captured under artificial conditions, i.e. the subject is facing the camera under good lighting conditions, there is a limited degree of head movement, and the expressions are usually exaggerated.
- **Spontaneous facial expressions** are more representative of what happens in the real world, typically occurring under less controlled circumstances. With spontaneous expression data, subjects may not necessarily be facing the camera, the image size may be smaller, there will undoubtedly be a greater degree of head movement, and the facial expressions portrayed are in general less exaggerated.

The dynamics of posed expressions can not be taken as representative of what would happen during natural displays of emotions, similar to how individual words spoken on command would differ from the natural flow of conversation. Consequently, when analysing the dynamics of facial expressions, one must realise that while the final image in a posed sequence will be the requested facial expression, the entire sequence as a whole will not allow for the accurate modelling of the interplay between the different movements that make up the facial expression during its natural formation. This is because subjects often use different facial muscles when asked to pose an emotion such as fear as opposed to when they are actually experiencing fear.

2.1.3 Encoding facial expressions

While the importance of facial expressions was established in 1872, it wasn't until the 1970's that researchers began to analyse the individual movements that make up facial expressions. Many different techniques were developed which claimed to provide means for objectively measuring facial expressions. Although, the *Facial Action Coding System* (FACS) introduced by Ekman and Friesen in 1978 is arguably the most widely used of these techniques, as a result we have chosen to use it as a basis for describing expressions in our research.

The Facial Action Coding System (FACS)

The FACS was first introduced by Ekman & Friesen in 1978 (Ekman & Friesen, 1978). It is the most comprehensive standard for describing facial expressions and is widely used in research. It provides an unambiguous quantitative means of describing all movements of the face in terms of 47 *Action Units* (AUs). Unlike previous systems that use emotion to describe facial expressions, the advantage that FACS has over its competitors is the way in which the FACS explicitly distinguishes between AUs, and inferences about what they mean. However should one wish to make emotion based inferences from the FACS codes, resources such as the FACS interpretive database developed by Ekman, Rosenberg and Hager in 1998 are available (Ekman et al., 1998). For a complete list of AUs and descriptions of these AUs see (Ekman & Friesen, 1978).

As the field of facial expression analysis has evolved, so too has the FACS, with its latest amendment occurring in 2002 where intensity codes were included for all AUs (Ekman et al., 2002). There are five intensity ranges defined in total, ranging from ***A*** to ***E***, with ***A*** representing a subtle change in appearance, and ***E*** representing a maximum change in appearance. It should be noted that although there are five intensity levels, these intensities do not occur equally, for example intensity ***C*** occurs for a longer duration than intensity ***A***

during the formation of a given AU. An example of the affect that increasing intensity has on the appearance a facial expression is shown in Fig. 1.

Fig. 1: Expression intensity ranges, intensities displayed, from left to right, are: neutral, and intensities A, C, E

While the 2002 version of the FACS is a significant improvement on the previous version of the FACS, in that it included descriptions on how to grade the different intensities for all AUs, these FACS guidelines for intensity coding are somewhat subjective. Hence special effort is required to establish and maintain acceptable levels of reliability, especially in the mid-range. Sayette *et al.*, suggest that the reliability of intensity coding may be problematic and state that further work is needed (Sayette et al., 2001).

2.2. Computational approaches for facial expression analysis

Using the foundations laid down by behavioural science researchers, in this section we appraise computer vision solutions for the problems of classifying and synthesising facial expressions. Here researchers attempt to create automated computational models for facial expression classification and synthesis.

Although the FACS provides a good basis for AU coding of facial images by human observers, the way in which the AU codes have been defined does not easily translate into a computational test. The reason for this is that the FACS is an appearance based technique with the AUs being defined as a series of descriptions. As a consequence of this description based method, a certain element of subjectivity occurs with human coders.

Due to this subjectivity, the development of a system to automatically FACS code facial images is a difficult task. Although the development of such as system would be an important step in the advancement of studies on human emotion and non-verbal communication, potentially enhancing numerous applications in fields as diverse as security, medicine and education. However, this is as yet an unsolved problem. According to Cohn *et al.*, further development is required within the area of automatic facial AU recognition before the need for manual FACS coding of facial images is eliminated (Cohn et al., 2002).

The automated analysis of facial expressions is a challenging task because everyone's face is unique and interpersonal differences exist in how people perform facial expressions. Numerous methodologies have been proposed to solve this problem such as Expert Systems (Pantic & Patras, 2006), Hidden Markov Models (Cohen et al., 2003); Gabor filters (Bartlett et al., 2006a; Bartlett et al., 2006b; Bartlett et al., 2004; Bartlett et al., 2001,) and Optical Flow analysis (Goneid & Kalioby, 2002). For an overview of these techniques see (Tian et al., 2004).

While the computational analysis of facial expressions and their dynamics have received a lot of interest from various different research groups, in the past decade the techniques

developed by Bartlett *et al.* have come to represent the state of the art in facial expression analysis. Bartlett *et al.* proposed a technique which combines *Gabor wavelets* and *Support Vector Machines (SVMs)* to classify the six primary facial expressions and neutral in a seven-way forced decision, achieving an accuracy rate of 93.3% (Bartlett et al., 2001). In (Bartlett et al., 2004), Bartlett *et al*. extended on this technique to classify 18 AUs[1] with an agreement rate of 94.5% with human FACS coders.

Although this agreement rate is impressive, the most important contribution of this work lies in the design and training of the context independent classifiers. These classifiers are capable of identifying the presence of an AU whether it occurs singly or in combination with other AUs. The benefit of this approach is clear as there are an estimated 7000 possible AU combinations, if context independent classifiers were used, then potentially all possible AU combinations could be classified using only 47 classifiers.

In (Bartlett et al., 2006a), Bartlett *et al.* present a system that accurately performs automatic recognition of 20 AUs[2] from near frontal image sequences in real time, once again using Gabor wavelets and SVMs with a 91% agreement with human FACS coders. Following recent trends in using standardised approaches for characterising the performance of classification systems, Bartlett *et al.* have begun to use *Receiver Operating Characteristic* (ROC) curve analysis, whereby they report the success/failure of their classifiers in terms of the *area under the ROC curve* (AUC). In (Bartlett et al., 2006b), Bartlett *et al.*, present the results of their experiments using ROC analysis, reporting an AUC of 0.926. For details on ROC analysis see Section 4.1.2.

Within the field of facial expression synthesis, a number of approaches have been reported in the literature over the past 10 years (Raouzaiou et al., 2002; Zhang et al., 2006; Wang & Ahuja, 2003; Choe & Ko, 2001; Gralewski, 2004; Ghent, 2005a; Ghent 2005b). The approach presented in this chapter combines the FACS, statistical shape and texture models, and machine learning techniques to provide a novel solution to this problem.

In the past *Radial Basis Function Networks* (RBFN) have been applied to facial expression synthesis (King & Hou, 1996; Arad et al., 1994). However, in these approaches redundancy reduction techniques were not applied prior to calculating the mapping functions. This kept the dimensionality of the mapping functions high and meant that irrelevant information was used in calculating the mapping functions. In King's (King & Hou, 1996) approach, mapping functions were used to modify the locations of *Facial Characteristic Points* (FCP) which in turn were used to warp an image to depict an alternative expression. A weakness with this approach was, that in order to adequately model the appearance change due to expression, one must take account of the variation of both shape and texture. For example, to synthesise a smile the texture of the image must be modified to produce wrinkles. The technique described in this chapter overcomes this problem by manipulating both shape and texture of the input image.

More recently, Abboud (Abboud et al., 2004) applied PCA to the shape and texture of unseen images to lower the dimensionality of the problem in order to produce synthetic facial expressions. However, their approach used linear regression to perform facial expression synthesis. Our technique improves on this approach by using a RBFN to describe the non-linear nature of facial expressions. The results of the technique described in this chapter considerably outperform the results found in (Abboud et al., 2004).

[1] AUs {1, 2, 4, 5, 6, 7, 9, 10, 12, 15, 17, 20, 23, 24, 25, 26, 27, 44}

[2] AUs {1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15, 16, 17, 20, 23, 24, 25, 26, 27}

3. Building facial expression models

A common problem within the areas of machine learning and computer vision is the extraction of relevant features from high-dimensional datasets such as video sequences. Feature extraction refers to the process of transforming the input data set to a lower dimension where the resulting dimensions exhibit high information packing properties (i.e. they accurately represent the original data). In general dimensionality reduction techniques are possible due to the fact that datasets have a lower implicit dimensionality in that their input representation is redundant. For this reason dimensionality reduction techniques are often referred to as redundancy reduction techniques.
Images of faces exhibit these properties in that the dimensionality of the space of faces is far lower than that of the images themselves (Meytlis & Sirovich, 2007). Given this fact, prior to both classification and synthesis, we apply redundancy reduction techniques to reduce the dimensionality of the input images. This has the advantages of reducing the complexity of the machine learning task and as a consequence increasing their accuracy.
Dimensionality reduction techniques can be either linear, such as *Principal Component Analysis (PCA),* or non-linear, such as *Locally Linear Embedding (LLE)*. Linear dimensionality reduction techniques can in general be represented by $\boldsymbol{y=Ax}$, where the $\boldsymbol{A}$ represents a linear transformation which when applied to the input data $\boldsymbol{x}$ maps it to the lower dimensional output vector $\boldsymbol{y}$. Nonlinear dimensionality reduction techniques can be represented by $\boldsymbol{y=fx}$ where $\boldsymbol{f}$ is typically a particular family of nonlinear mappings and again, $\boldsymbol{x}$ and $\boldsymbol{y}$ are the high-dimensional input and low-dimensional output vectors, respectively.
In our research we have investigated the application of both PCA and LLE for facial expression analysis. Details of each of these techniques are provided in the Sections 3.2 and 3.3. For completeness, prior to explaining these techniques, we give details of the dataset used in our experiments.

3.1 Facial expression dataset

In order to fully utilize the FACS AU and intensity coding we use a database containing videos of individuals performing a series of facial expressions which are fully FACS coded. In our research to date we use the Cohn-Kanade AU-Coded Facial Expression Database (Cohn & Kanade, 1999). This database contains approximately 2000 images sequences from over 200 subjects. The subjects come from a cross-cultural background and are aged between 18 - 30. This database contains full AU coding and partial intensity coding of facial images and is the most comprehensive database currently available.

Fig. 2**:** Example of landmark points used to describe the shape of a particular face

Prior to experimentation, we extract the expression features from our dataset by manually identifying a set of 122 landmark points on the face as shown in Fig. 2. To ensure that the variance in the data set is due to change in expression rather than subject we perform a global alignment step known as *Generalised Procrustes Alignment* (GPA) (Ghent, 2005a). Applying GPA minimises the sum of the squared distances between corresponding landmark points and in effect normalises the shapes with respect to scale, translation, and rotation. Given the set of globally aligned shapes there will be a high degree of correlation in the variation of landmark positions due to expression.

3.2 Principal Component Analysis – PCA

PCA is used to map high dimensional data to a low dimensional subspace. This method takes a set of data points and constructs a lower dimensional linear subspace that maximises the variance in the training set. PCA essentially performs an orthonormal transformation on the input data such that the variance of the input data is accurately captured using only a few of the resulting principal components.

These principal components are calculated in such a way that the squared reconstruction error between the input and output data are minimized. This is achieved by performing eigenvector decomposition on the covariance matrix of the input data. The resulting eigenvectors and eigenvalues represent the degrees of variability where the first eigenvalue is the most significant mode of variation. This process of maximising the variability of the input data and minimizing the reconstruction error can be achieved by rotating the axes. This axis rotation is the core idea behind PCA.

The higher the correlation between the input data the fewer the number of principal components needed to represent the majority of variance in the training set. However, if the input data points are co-linear then the data can be represented without any information loss using only one dimensional data along the principal axis. In correlated input data there also exists redundant information in the input space. PCA removes this by de-correlating the input data. The uncorrelated low dimensional features can be used to represent the correlated high dimensional input data making PCA a powerful data compression technique.

Given the set of globally aligned shapes there will be a high degree of correlation in the variation of landmark positions due to expression. In order to reduce this redundancy we perform PCA on the data set. To do this each shape is represented as a one-dimensional vector of the form:

$$\mathbf{P}_i = \left[\mathbf{p}_i^1 \quad \mathbf{p}_i^2 \quad \ldots \quad \mathbf{p}_i^n\right]^{\mathrm{T}} = \left[x_i^1 \quad y_i^1 \quad x_i^2 \quad y_i^2 \quad \ldots \quad x_i^n \quad y_i^n\right]^{\mathrm{T}} \tag{1}$$

For each $\mathbf{P}_i$, we define the difference vector as:

$$\boldsymbol{\delta}\mathbf{P}_i = \bar{\mathbf{P}} - \mathbf{P}_i \tag{2}$$

Where, $\bar{\mathbf{P}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{P}_i$ is the *mean shape*, using Equation (2) we may now define the covariance matrix of the dataset as:

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{\delta}\mathbf{P}_i\right)\left(\boldsymbol{\delta}\mathbf{P}_i\right)^{\mathrm{T}} \tag{3}$$

The m eigenvectors corresponding to the m largest eigenvalues of $\mathbf{S}$ are known as the *principal components* and are orientated in the directions of the m highest modes of variation of the original dataset. These eigenvectors form the basis of a low-dimensional subspace capable of representing the input shapes far more efficiently whilst capturing the majority of the variance in dataset. We refer to this space as the *Facial Expression Shape Model* (FESM). For example in our experiments the input space is 244-dimensional whilst the maximum number of dimensions that we use in an FESM is 20, which captures over 98% of the variance of the variance in the input dataset. Given an input point, $\mathbf{P}_i$, we can compute the corresponding FESM vector using:

$$\mathbf{b} = \mathbf{B}^{\mathrm{T}}\left(\overline{\mathbf{P}} - \mathbf{P}_i\right) \tag{4}$$

Where, $\mathbf{B}$ is an $n \times m$ matrix where the columns are the m eigenvectors described above. Conversely given an FESM vector we can project it into the input space by inverting Equation (4).

3.2 Locally Linear Embedding - LLE

According to Kayo *et al.*, as real world data is often inherently nonlinear, linear dimensionality reduction techniques such as PCA, do not accurately capture the structure of the underlying manifold, i.e. relationships which exist in the high dimensional space are not always accurately preserved in the low dimensional space (Kayo et al., 2006). This means that in order to capture the underlying manifold of real world data, a nonlinear dimensionality reduction technique is required. On one such nonlinear dimensionality reduction technique is LLE.

LLE was introduced as an unsupervised learning algorithm that computes low dimensional, neighbourhood preserving embeddings of high dimensional data (Saul & Roweis, 2003). The LLE algorithm is based on simple geometric intuitions, where it essentially computes a low dimensional representation of the input data in such a way that nearby points in the high dimensional space remain nearby and similarly co-located with respect to one another in the low dimensional space.

The LLE algorithm takes a dataset of N real valued vectors $\boldsymbol{X_i}$, each of dimensionality $\boldsymbol{D}$, sampled from some smooth underlying manifold as its input. Provided that the manifold is sufficiently sampled by the dataset, we can expect each point and its neighbours to lie on or close to a locally linear patch of the manifold. The LLE algorithm involves three main steps. Firstly, the manifold is sampled and the $\boldsymbol{K}$ nearest neighbours per data point are identified. Secondly each point $\boldsymbol{X_i}$ is approximated as a linear combination of its neighbours $\boldsymbol{X_j}$.

These linear combinations are then used to construct the sparse weight matrix $\boldsymbol{W_{ij}}$. Reconstructions errors are then measured by the cost function given in Equation 5, which sums the squared distances between each point and its reconstruction.

$$\varepsilon W = \sum_i \left| \overrightarrow{X_i} - \sum_j W_{ij} \overrightarrow{X_j} \right|^2 \tag{5}$$

In the final step of the LLE algorithm, each point $\boldsymbol{X_i}$ in the high dimensional space is mapped to a point $\boldsymbol{Y_i}$ in the low dimensional space which best preserves the structure and geometry of $\boldsymbol{X_i}$'s neighbourhood. The geometry and structure is represented by the weight

matrix W_{ij}. The mapping from X_i to Y_i is achieved by fixing the weights W_{ij}, and selecting the bottom d non zero coordinates of each output Y_i to minimise Equation 6. For more details on the LLE algorithm see (Saul & Roweis, 2003).

$$\Phi(Y) = \sum_i \left| \vec{Y}_i - \sum_j W_{ij} \vec{Y}_j \right|^2 \tag{6}$$

4. Classifying facial expressions

In this section we provide details of our approach towards the automatic classification of facial expressions, and the extraction and modelling of their dynamics. In our experiments we use *Support Vector Machine* (SVM) classifiers, and characterise the performance of our technique by performing *Receiver Operating Characteristic* (ROC) analysis on the results of our experiments. Details of both of these techniques are given in section 4.1. Following on from this we detail a number of experiments which demonstrate our approaches towards the automatic classification of facial expressions.

4.1 Facial expression classification and validation

4.1.1 Support Vector Machines - SVMs

SVMs are a type of learning algorithm based upon advances in statistical learning theory, and are based on a combination of techniques. One of the principal ideas behind SVMs is the kernel trick, where data is transformed into a high dimensional space making linear discriminant functions practical. SVMs also use the idea of large margin classifiers. Suppose we have a dataset $(x_1, y_1),\ldots,(x_m, y_m) \in X \times \{\pm 1\}$ where X is some space from which the x_i have been sampled. We can construct a dual Lagrangian of the form:

$$W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{7}$$

The solution to Equation 7, subject to the constraints $\alpha_i \geq 0 \forall i$ and $\sum_{i=1}^{m} \alpha_i y_i = 0$, is a set of $\boldsymbol{\alpha}$ values which define a hyperplane that is positioned in an optimal location between the classes.

A number of methods have been proposed for multi-class classification using SVMs, the *one-against-all* and the *Directed Acyclic Graph* (DAG) algorithms are the two main approaches (Ghent, 2005a). In our experiments we use the *one-against-all* approach.

4.1.2 Receiver Operating Characteristic (ROC) curve analysis

ROC analysis is a technique for visualizing, organising, and selecting classifiers based on their performance (Fawcett, 2003). Within the field of computer science, ROC analysis is becoming increasingly important in the area of cost sensitive classification, classification in the presence of unbalanced classes, robust comparison of classifier performance under imprecise class distribution and misclassification costs.

Given a classifier and an instance, there are four possible outcomes:

- True Positive (TP) - test correctly returns a positive result
- True Negative (TN) –test correctly returns a negative result

- False Positive (FP) –test incorrectly returns a positive result
- False Negative (FN) –test incorrectly returns a negative result

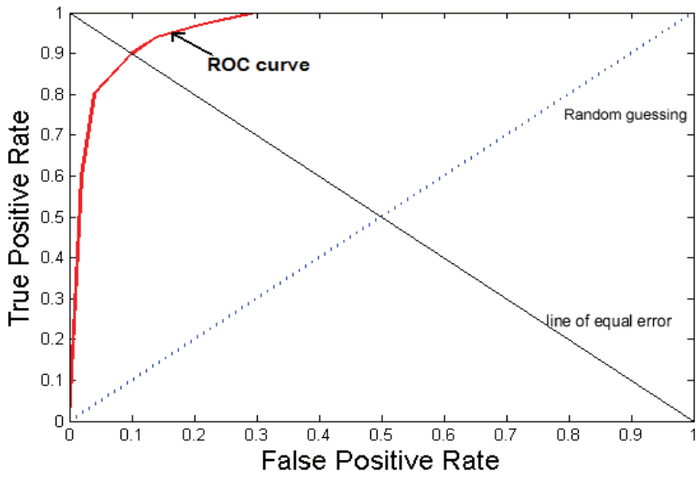

Fig. 3: Example ROC curve, with line of equal error & random guess line

These values are used to compile various ratios such as the *True Positive Rate* (TPR), which is *the number of true positives divided by the total number of positives*, and the *False Positive Rate* (FPR) is 1 – *the number of false positives divided by the total number of negatives*. ROC graphs are two dimensional graphs with the TPR plotted on the Y-axis, and the FPR plotted on the X-axis. A ROC graph depicts the relative trade offs between the benefits and costs of a particular classifier. An example of a ROC curve is shown in Fig. 3. The most frequently used performance metric in ROC analysis is the *area under the ROC Curve* (AUC). The AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. The AUC ranges from 0-1, and a random classifier has an AUC of 0.5.

4.2 Facial expression classification

The area of facial expression classification was originally concerned with classifying the six primary facial expressions (joy, sadness, fear, surprise, anger, and disgust). However, as the need for a consistent representation of facial expressions became apparent, this focus has changed, with researchers concentrating on classifying expressions in terms of the individual movements or AUs that make up the facial expressions. In this section we provide details of our work within these two distinct areas, concluding with details of our technique which models the dynamics of facial expression in terms of intensity and timing.

4.2.1 Classification of the six primary expressions

To date the research group at the *Computer Vision and Imaging Laboratory* (CVIL) at the National University of Ireland, Maynooth, have proposed a computational model for the classification of facial expressions (Ghent, 2005a). This model which is based on PCA and SVMs, can accurately classify the primary facial expressions at extreme levels of intensity. This model was created by firstly reducing the dimensionality of our data using PCA. This lower dimensional data was then used to train one-against-all SVMs, in this example we

have three expressions to classify (happiness, sadness and surprise), so we used three one-against-all SVMs (For more details on SVMs see Section 4.1.1.).

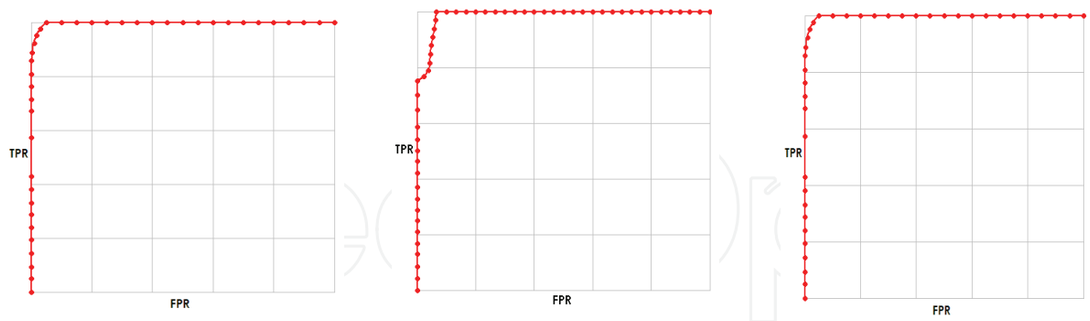

Fig. 4: ROC Curves for the three of the primary facial expressions, from left to right, Happiness, Sadness & Surprise

Following on from this, test data was projected into the training space, and presented as input to the SVMs. As a means of validating our results we performed ROC analysis. The resulting ROC curves are shown in Fig. 4, and the confusion matrices shown in Table 1. Our PCA-SVM technique has achieved a mean AUC of 0.91. Once we demonstrated the success of our technique at classifying a number of the primary facial expressions, such as happiness, sadness and surprise, we then extended on this work to analyse the individual movements that make up facial expressions.

4.2.2 Classification of individual action units (AUs)

Although FACS provides a good basis for the AU coding of facial images by trained individuals, the automatic recognition of AUs by computers remains a difficult challenge. As mentioned earlier, one issue with automating FACS coding is that 47 AUs have been defined and these can occur in a large number of combinations (estimated at over 7000 (Pantic & Rothkrantz, 2003)). One approach to overcome this problem is to subdivide the face into regions and classify expressions within these regions independently.

Emotion	AUC	#Neg	#Pos	FP	TP	FN	TN
Happiness	0.94	98	40	2	12	28	96
Sadness	0.804	121	17	0	2	15	121
Surprise	0.996	118	20	0	3	17	118

Table 1: ROC Results, AUC = Area under the ROC Curve, **#Neg** = number of negative samples, **#Pos** = number of positive samples, **FP** = False Positives, **TP** = True Positives, **FN** = False Negatives, **TN =** True Negatives

As the FACS separates facial expressions into upper and lower facial AUs, in this section we demonstrate our technique at classifying AUs in both of these regions of the face. In Experiment 1 we report our results for the classification of lower facial expressions involving the mouth, while in Experiment 2 we provide details of our experiments for the classification of upper facial expressions involving the eyebrow.

In these experiments we used LLE to reduce the dimensionality of our datasets. An LLE shape space is established by pre-processing training face shapes and using these as inputs

to into an LLE algorithm. SVMs are then trained on this data. In our experiments we use one-against-all SVM classifiers, so if for example there were four expressions in our training set, we would use 4 one-against-all SVMs. Once these classifiers have been trained, individual unseen shapes are pre-processed and projected into the LLE expression shape space. The results of this projection are then used as inputs to a previously trained SVM classifiers which output a FACS coding for this unseen shape. We perform ROC analysis on the results of the experiments.

Experiment 1 – Classification of lower face AUs

In our first experiment we classify four lower facial expressions; AU20+25, AU25+27, AU10+20+25 and AU12, the affect that these AUs have on the face is shown in Fig. 5. Our training data consisted of 73 images of an individual performing these four AU combinations, from neutral to extreme expression intensity. It should be noted that expression I & III are very similar and therefore we hypothesise that a technique that can accurately differentiate between these two expressions can accurately classify subtle changes in appearance. Our test set consisted of 522 images of multiple subjects from multi-cultural backgrounds performing the four lower facial expressions as shown in Fig. 5. In our test dataset we sampled the sequences at each intensity rating, including neutral (6 in total). In our training set we used the entire sequence and labelled each frame with an intensity score.

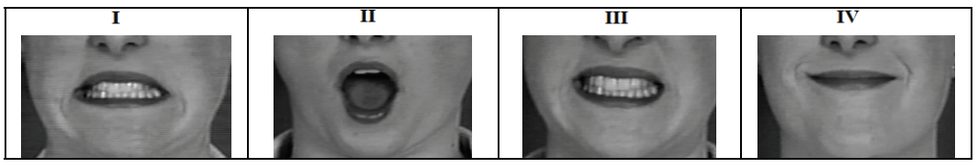

Fig. 5: This Figure illustrates the effect of portraying four different AU combinations on the mouth. From left to right the expressions portrayed are: I = AU20+25, II = AU25+27, III = AU10+20+25, IV = AU12

Results:

Using the outputs of our SVM classifiers we preformed ROC analysis, the optimal ROC curve for each of the four expressions are shown in Fig. 6. While from first glance it appears that our technique did not perform as well at the task of classifying the two similar expressions I - AU20+25 and III - AU10+20+25, it is important to note that the classifiers were tested across the entire intensity range. Possible reasons for the lower AUC's for these two expressions could be that as these two expressions are quite similar, in the more subtle intensities they may have been miss-classified.

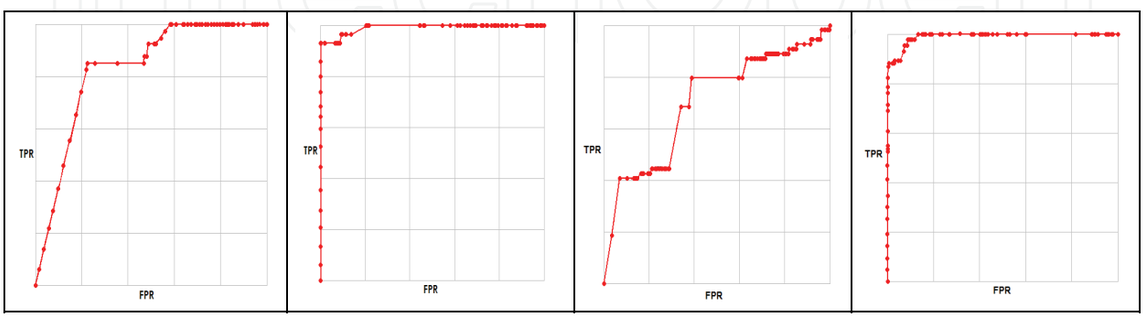

Fig. 6: ROC results from the application of LLE to lower facial expressions. From left to right: I=AU20+25, II=AU25+27, III=AU10+20+25, IV=AU12

Expression	AUC	#Neg	#Pos	FP	TP	FN	TN
AU20+25	0.803	422	20	93	15	93	329
AU25+27	0.951	382	28	1	22	6	381
AU10+20+25	0.687	432	18	31	8	10	401
AU12	0.989	417	105	0	62	43	417

Table 2: ROC Results, AUC=Area under the ROC Curve, **#Neg**=number of negative samples, **#Pos**=number of positive samples, **FP**=False Positives, **TP**=True Positives, **FN**=False Negatives, **TN=**True Negatives

From looking at the confusion matrices for these two classifiers, (see Table 2), for expressions I & III, we can see that there is a higher level of false positives than with the other expressions. However, this is still a positive result as the classification and differentiation between two such similar facial expressions across the entire intensity range is a non-trivial task. (For an in-depth analysis of these results across the entire intensity range see (Reilly, 2007)).

Experiment 2 – Classification of upper face AUs

The structure of our second experiment is similar to that of our previous experiment except that instead of attempting to classify each AU or AU group separately, we wanted to classify AU1 and AU4 independent of context. What we mean by this is that we wanted to classify AU1 regardless of whether it occurs on its own or in combination with other AUs within the eyebrow region such as AU2 and AU4. The motivation for the development of context independent classifiers is that the 47 AUs defined by the FACS can occur in over 7000 possible combinations. Due to this large number of possible combinations, it is not practical to design an independent classifier for each of these cases. Hence, in order to simplify this classification problem, if we design classifiers that will classify the presence of an AU regardless of whether it occurs in isolation or in combination with other AUs, we could potentially classify all of these possible AU combinations using 47 classifiers. This is a challenging problem as there is a significant overlap in how these AUs can alter the appearance of the face. This overlap is demonstrated in Fig. 7, where all the possible combinations of AU1 and AU4, within the eyebrow region, which are available within the Cohn-Kanade database are displayed.

Fig. 7: Example of AU combinations that can occur in the eyebrow region, from left to right they are, AU1, AU4, AU1+4, AU1+2, AU1+2+4

Our training data consisted of 42 images of an individual performing the five eyebrow movements as illustrated in Fig. 7, across the entire intensity range from neutral to extreme AU intensity. Our test set consisted of 84 images from multiple subjects from multi-cultural backgrounds performing the AUs as shown above. In our test dataset we sampled our data at each level of intensity from neutral to extreme, where 6 samples per sequence were taken. While there are three AUs associated with the eyebrow, there were not sufficient samples of AU2 within the Cohn-Kanade database to include this AU in this particular experiment. However, we hypothesise that a technique that can accurately classify the remaining AUs:

AU1 & AU4, regardless of whether they occur together or in isolation has the potential to perform similar classification in more complex regions of the face such as the mouth.

Results:

Using the outputs of our SVM classifiers we preformed ROC curve analysis, the optimal ROC curves for the context independent classification of AU1 and AU4 are shown in Fig. 8. From analysing the ROC curves we can see that our technique was successful at classifying AUI, where it achieved an AUC of 0.789. Our technique was less successful at classifying AU4, achieving an AUC of 0.62. This could possibly be because we had more samples containing AU1 in our training set. Although it could also be attributed to the fact that one of the indicators of the presence of AU4 is the appearance of wrinkles and bulges between the brows, and as our technique is shape based, this information is not captured.

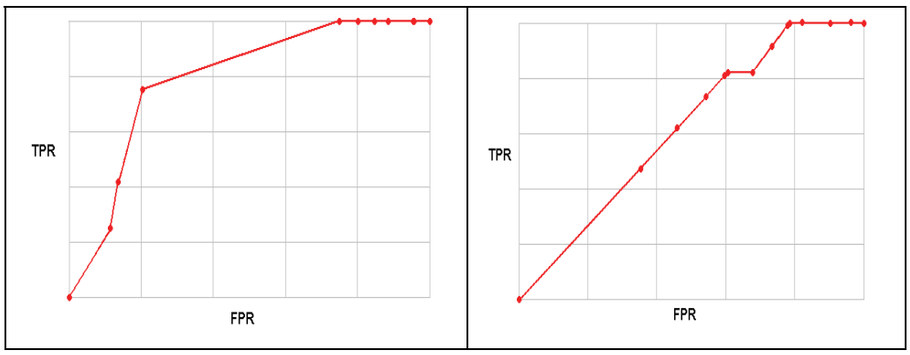

Fig. 8: ROC curves for the context independent classification of AU1 (Left), and AU4 (Right)

An example of the affect that AU4 has on the face is shown in detail in Fig. 9, where it can be seen that the presence of AU4 causes bulges to appear between the brows. Nonetheless, the development of context independent classifiers is a difficult problem. We envisage that our results will improve when we extend on our current models to also include texture information.

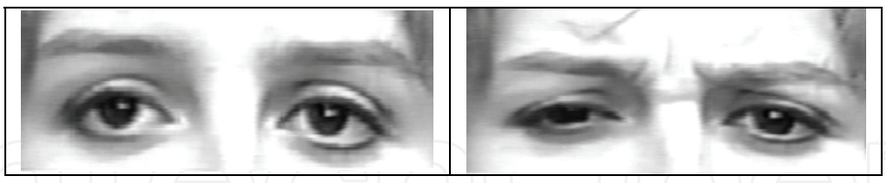

Fig. 9: Effect of AU4 on the eyebrow region, the neutral expression is shown on the left, and the extreme is shown in the right

Experiment 3 – capturing the dynamics of facial expression.

When investigating the dynamics of facial expression, we are interested in capturing the appearance changes that occur during facial expression formation in terms of the intensity and timing of those changes. When referring to facial expressions, intensity refers to the magnitude of the appearance change resulting from any facial activity, and the timing refers to the speed and duration of that activity. The dynamics of facial expression, such as the timing, duration and intensity of facial activity plays a critical role for the interpretation of the observed behaviour.

In the following experiment we extract information regarding the intensity and timing of a previously classified AU. The extraction of this information provides a means for analyzing

the dynamics of facial expression. In this experiment we estimate the intensity of AU25 – *which parts the lips*. The input to this experiment consisted of 24 subjects performing AU25, sampled at each intensity rating (i.e. per subject we sampled at intensity Neutral, A, B, C, D and E).

As we wish to capture the dynamics of AU formation, we perform an extra pre-processing step called *Shape Differencing*, whereby the neutral mouth shapes of each subject are subtracted from the sample set for that subject. The reason being is that we are only interested in the difference between the two shapes and not the actual shapes themselves. This is a valid step as in order to analyse the dynamics of facial expression it is necessary to have a sequence containing a neutral image.

Fig. 10: Examples of sequences of AU25 from the Cohn-Kanade database

As mentioned earlier in Section 2.2.3, the 2002 version of the FACS contains intensity ranges for each AU. The FACS intensity range goes from A to E, with A representing a minor change in appearance and E representing the maximum change in appearance. As the data for this experiment was taken from the Cohn-Kanade database, full intensity coding of the image sequences was not available. Therefore we manually selected the frames from the expression sequences which correspond to the FACS intensity coding descriptions, examples of our selections can be seen in Fig. 10.

In our initial experiments we developed the dynamical model as shown in Fig. 11. However, there is a significant overlap between intensities in the mid ranges, for example intensity D covers a large portion of the axis. We hypothesise that this is due to the fact that the FACS intensity codes are quite subjective and as a result the distinction between the different intensities across our dataset is a difficult task.

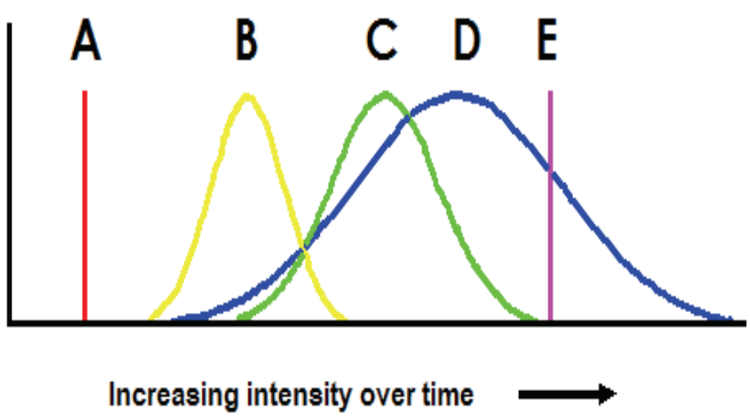

Fig. 11: Distributions for our FACS based dynamical model

To deal with this problem we relabelled our dataset based on three categories corresponding to low, medium, and high intensity displays of the AUs. As a result of this relabelling the outputs of the estimation process became more repeatable and representative of the underlying dynamics of expression formation. The results of the clustering of the dataset under the three-category labelling can be seen in Fig. 12. This is an extension of our previous works on modelling the dynamics of facial expression (Reilly, 2006).

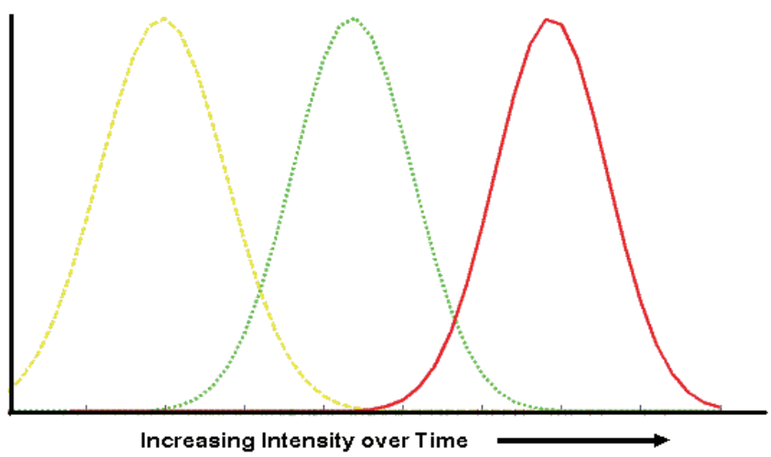

Fig. 12: Distributions for our new 3 stage simplified intensity model

Facial Expression Synthesis

In this section we describe a technique which we have developed that allows for photo-realistic images of a person depicting a desired expression to be synthesised in real-time once a neutral image of the subject is present (Ghent, 2005a; Ghent, 2005b). We also demonstrate how through simple linear interpolation in the expression space the intensity of the output expression can be varied and hence can be used to generate image sequences of expression formation. The approach combines facial appearance models, machine learning techniques, and the FACS to compute a *universal facial expression mapping function* (i.e. a function which maps facial expressions independent of identity).

5.1 Modelling the appearance of facial expressions

In our approach we use active appearance models to first derive low-dimensional spaces, known as *expression spaces*, which are capable of representing the variation in appearance of facial expressions. Active appearance models (Cootes et al., 2001) model the variation of appearance of a class of objects in terms of the variation of both the shape and texture of images of instances of the class. The shape of an object is characterised by the locations of a set of fiducial points on the object, known as landmarks, whereas the texture is defined as the set of pixel intensity values contained within the convex hull of the set of landmarks. By considering the complete collection of landmark sets, independent of the underlying texture data, we can derive a model of the variation in the image due to shape alone. By applying this technique to our dataset we derive separate shape and texture models which we call the *Facial Expression Shape Model* (*FESM*) and the *Facial Expression Texture Model* (FETM). Computation of the FESM is described in Section 3.2.

To compute the FETM we must first warp each input image to the mean shape. This provides us with a *shape independent* representation of the facial texture. Representing the resulting image as a one-dimensional vector (i.e. by concatenating each row of pixels).

$$\mathbf{P}_i = \begin{bmatrix} g_i^1 & g_i^2 & \dots & g_i^n \end{bmatrix}^{\mathrm{T}} \tag{8}$$

A similar procedure can be applied to the texture data as was applied to the shape data in deriving the FESM. That is, the covariance of the texture data and corresponding principal components may be computed. For example in our experiments each shape independent image contains approximately 90,000 dimensions (i.e. pixels) whilst the maximum number of dimensions that we use in an FETM is 20 in which case captured 92% of the variance in the input dataset.

These spaces allow us to constrain the expression synthesis mapping functions such that we can ensure that both the input and output of the functions represent facial images. This is achieved by using the projection of the input image in the expression space, as opposed to the input image itself, as input to the function. Furthermore since the output of the function is also a point in the expression space we ensure that the function only outputs facial images.

5.2 Learning universal expression mapping functions

As mentioned above, an expression mapping function maps a point corresponding to one expression type (e.g. neutral) for a particular individual to the point corresponding to a different expression (e.g. surprise) for that individual. The term "universal expression mapping function" is used to emphasize the fact that the function should perform this mapping for all individuals independent of identity.

To model this function we use *artificial neural networks* (ANN's) in conjunction with the FESM and FETM. Since we represent the shape and texture separately, it follows that for a given expression mapping we must construct the universal mapping as two separate functions. To do this we use a Linear Network (LN) and a Radial Basis Function Network (RBFN) to model the universal mapping function in the FESM and FETM, respectively (Ghent, 2005b). The reason for the different choice of network architecture is due to the fact that we have found the mapping in the FETM to be highly non-linear and hence cannot be modelled accurately using a LN.

Both networks are trained in a supervised manner using a subset of the database described in Section 3.1. For each image, the landmark set and the shape normalised texture are projected into the FESM and FETM, respectively, producing two vectors, $\mathbf{b}_s$ and $\mathbf{b}_t$. Training of the LN involves presenting the network with pairs of $\mathbf{b}_s$ vectors corresponding to the FESM representation of the neutral and expression shape for each individual. Training of the RBFN is performed in the same manner using the $\mathbf{b}_t$ vectors.

Once both networks are trained, synthesis is performed on an input image by identifying the landmark positions and computing the shape normalised texture. Again these are both projected into the FESM and FETM producing vectors $\mathbf{b}_s$ and $\mathbf{b}_t$. These vectors are then used as input to the respective ANN's resulting in two new vectors $\tilde{\mathbf{b}}_s$ and $\tilde{\mathbf{b}}_t$ corresponding to the predicted vectors for the individual portraying the output expression used in the training set. Reconstruction of the image is achieved by inverting Equation 2 for both the shape and texture.

5.3. Experiments

Many interactive media applications involving faces require the ability to animate a face in real-time. Such applications include personalised media creation (i.e. inserting an individual into a pre-captured sequence or movie), personalised online avatars, or giving a character in a computer game the identity of the user. In these situations the user typically presents themselves to the system in a cooperative manner and hence the system can request that the user portray a neutral expression.

To evaluate the performance of our approach in the context of this type of application we have applied it to the synthesis of non-neutral expressions where the input is a neutral expression. Universal mapping functions were developed for synthesising joy (AU6 + AU12 + AU25), surprise (AU1 + AU2 + AU5 + AU26), and sadness (AU15 + AU17).

To create a FESM and a FETM we use images from the Cohn–Kanade AU coded Facial Expression Database described in Section 3.1. Again, each face was manually labelled with 122 landmark points as shown in Fig. 2. Given both the images and the landmark sets the procedure described above was applied to produce the facial expression appearance model (i.e. the FESM and FETM). Since for each expression mapping function we assume that the input and output are specific expressions, only those two expressions are used in building the appearance model. Hence for each new expression mapping function we create an expression space tailored to that mapping. The neural networks are then trained on sample input-output pairs of the expression vectors under consideration in that space.

(a) (b) (c) (d) (e) (f)

Fig. 13: Columns (a)-(c) show the shape-free neutral, non-neutral, and synthesised images, respectively. Columns (d)-(f) show the same texture as columns (a)-(c) but here the correct shape is used (i.e. the original shape in (d) and (e), and the synthesised shape in (f))

(a) (b) (c)

Fig. 14: Examples of the (a) input, (b) desired output, and (c) actual output of the mapping functions for surprise, joy and sadness

Fig. 13 shows examples of the outputs of one of these mappings for five different subjects. The first two rows consist of images of subjects that were used during the training of the networks while the next three individuals (rows 3–6) were not used during the training of the networks. Column one consists of shape free original images of individuals depicting neutral expressions. Column two consists of shape free original images of individuals depicting AU6C+AU12C+AU25 as described by the FACS. Column three consists of synthetic images of individuals portraying these AU's as calculated by the RBFN with the neutral texture vector as input. Columns 4–6 are the same as the first three columns, respectively, except with shape taken into consideration. The shapes in column 6 are calculated using a LN in conjunction with the FESM.

Fig. 14 shows examples of original neutral and non-neutral images of unseen individuals in conjunction with images where the synthesised expression has been overlaid on the original neutral. As can be seen from the figure the rows 1-3 show the outputs for the surprise, joy, and sadness mapping functions, respectively. In order to quantitatively evaluate the performance of the technique we compute the correlation coefficient between the synthesised data and the real data.

Table 3 and Table 4 show the correlation coefficients between the estimated and real principal components for the FESM using a LN and FETM using an RBFN, respectively. Here, N_I is the number of individuals (i.e. image pairs) used in the experiments, *Pram* is the number of principal components used as input and output to and from the neural networks and, *Perc* is the percentage of variance that *Pram* can describe. N_S is the total number of individuals used for training (i.e. seen) and testing (i.e. unseen) the mapping functions, while *Avg*, *Max* and *Min* are the average, maximum and minimum correlation coefficients between the estimated shape parameters and the real shape parameters, respectively.

From the average of correlation coefficient of the unseen shape and texture taken from Tables 3 and 4 we can compute that the overall average for unseen data is 0.757. Using a

similar technique Yangzhou and Xueyin (Yangzhou & Xueyin, 2003) showed how a universal mapping function achieves results of *Avg*=0.51. Given that the input and output to LN and RBFN are vectors representing the neutral and extreme expression, respectively, a natural next step is to try to develop mapping functions for outputting expressions at intermediate intensities. By treating the neutral and extreme vectors as the end points of the trajectory traced out by the expression vector during the expression formation process, generating intermediate intensity expressions can be treated as equivalent to identifying intermediate points on this trajectory. We have found that by approximating this trajectory as linear and hence using linear interpolation to identify intermediate point yields excellent results.

AU	N_I	***Parm***	***Perc***		N_S	***Avg***	***Min***	***Max***
6,12,25	40	15	94.04	Seen	35	0.959	0.753	0.998
				Unseen	5	0.875	0.589	0.997
1,2,5,26	20	20	98.61	Seen	15	0.976	0.899	0.999
				Unseen	5	0.777	0.522	0.971
15,17	17	20	99.35	Seen	15	0.969	0.776	0.999
				Unseen	2	0.699	0.554	0.845
Total	**77**	**N/A**	**N/A**	**Seen**	**65**	**0.968**	**0.754**	**0.999**
				Unseen	**12**	**0.784**	**0.522**	**0.997**

Table 3: Correlation coefficients between real and synthesised shape vectors using a LN

AU	N_I	***Parm***	***Perc***		N_S	***Avg***	***Min***	***Max***
6,12,25	40	15	95.59	Seen	35	0.997	0.936	1.00
				Unseen	5	0.780	0.628	1.00
1,2,5,26	20	20	89.18	Seen	15	0.991	0.875	1.00
				Unseen	5	0.776	0.317	0.875
15,17	17	20	92.03	Seen	14	0.977	0.737	1.00
				Unseen	3	0.635	0.501	0.737
Total	**77**	**N/A**	**N/A**	**Seen**	**64**	**0.988**	**0.737**	**1.00**
				Unseen	**13**	**0.730**	**0.317**	**1.00**

Table 4: Correlation coefficients between real and synthesised texture vectors using a RBFN

Fig. 15 shows an example of applying this interpolation process to the FESM only. Here we have taken the input neutral shape and used the universal mapping function to estimate the shape of the extreme expression. We then linearly interpolate 3 equally spaced points between the two corresponding vectors in the FESM. Using the 4 new shapes as warp targets we warp the input (neutral) texture producing the images shown. Note here that due

to the fact that we are just using the LN and FESM that we do not need to convert the original image to greyscale and hence we can use this approach to generate colour image sequences. Also even though the image is not taken from the Cohn-Kanade database the results are still accurate. We have applied the technique to a number of images which are not part of this database and achieved similar results and so are confident that the technique generalises well.

Fig. 15: Examples of our synthesis results, by varying the interpolation in the FESM

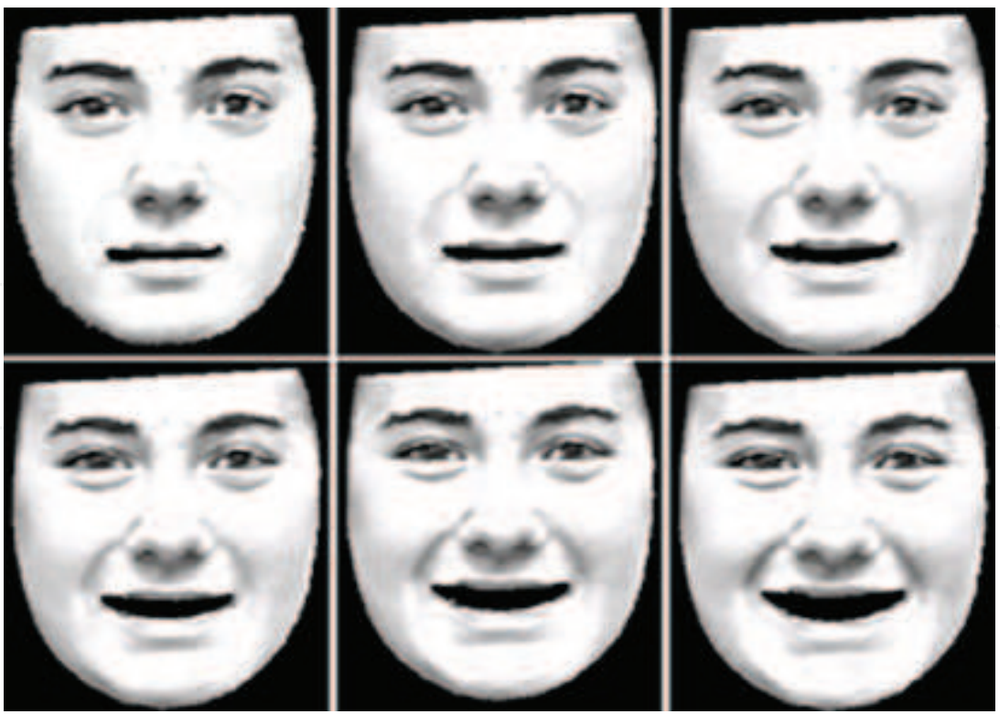

Fig. 16: Examples of our synthesis results, by varying intensity by interpolation in the FESM and FETM

Fig. 16 shows the result of applying the same procedure to both the shape and texture vectors of a given neutral image (i.e. interpolating both simultaneously). Since the FETM is computed using greyscale images the input and output of the process are also greyscale images. The advantage of interpolation in both spaces simultaneously can be seen in that here the texture is both warped and altered appropriately. This can best be seen by inspecting the cheek region (known as the *intraorbital triangle*) across the sequence. Specifically, in the neutral image there no creasing in this region, however in each subsequent image the creasing increases as would be expected in the expression associated with joy (due to the action of AU12). This alteration of the texture could not be achieved by simple warping alone and so requires a complete shape and texture model.

7. Conclusion

In this chapter we have discussed the field of facial expression analysis from both a Behavioural Science and a Computer Vision perspective. We introduced the field of facial expression analysis with Darwin's initial findings and then went on to provide details of the research conduced by Ekman *et al.* on facial expression analysis highlighting the importance of facial expression dynamics.

We then provided details of the current state-of-the-art in automated facial expression analysis, and presented our contribution to this field. In our facial expression classification section we demonstrated the success of our PCA based technique at classifying the primary facial expressions achieving and average AUC of 0.91. We developed separate LLE based shape models for the classification of upper and lower face AUs. Context independent classifiers were used to discriminate between two of the three AUs that occur within the eyebrow area.

Given our approaches to classification of static expressions, we then extended on this work to create dynamical models which estimate the AU intensity. The performance of this approach was evaluated using both the full FACS intensity system and a simpler system of low, medium, and high intensities. Distributions of the resulting intensity estimations for a sample of the Cohn-Kanade database were presented.

In the final section of this chapter we described a technique which allows for photo-realistic expression synthesis (Ghent, 2005a; Ghent, 2005b). This was achieved by applying machine learning techniques to the modelling of *universal facial expression mapping functions*. Three mapping functions were developed for mapping from neutral to joy, surprise, and sadness. We also demonstrated how the representation of expression used allowed the intensity of the output expression to be varied. This ability to vary the intensity of output enabled us to generate image sequences of expression formation.

8. References

Abboud, B., Davoine, F. & Dang, M. (2004) , Facial expression recognition and synthesis based on an appearance model, *Signal Processing: Image Communication*, 19, 8, (September 2004), pp. 723-740.

Ambadar, Z., Schooler, J., Cohn, J. (2005) Deciphering the enigmatic face: The importance of facial dynamics to interpreting subtle facial expressions. *Psychological Science* (2005)

Arad, N., Dyn, N., Reisfeld, D. & Yeshurun, Y. (1994) Image warping by Radial Basis Functions:Application to Facial Expressions, *CVGIP: Graphical Models & Image Processing*,56,2, (March 1994), pp.161–172.

Bartlett, M.S. B. Braathen, G.L.T.J.S., Movellan, J.: Automatic analysis (2001)of spontaneous facial behavior(2001) MPLABTR-2001-06, Institute for Neural Computation, University of California, San Diego.

Bartlett, M.S., Littlewort, G., Lainscsek, C., Fasel, I., Movellan, J.: Machine learning methods for fully automatic recognition of facial expressions and facial actions. *IEEE International conference on systems, man and cybernetics (2004) 592–597*

Bartlet, M., Littlewort, G., Lainscsek, C., Fasel, I., Frank, M., Movellan, J. (2006a): Fully automatic facial action recognition in spontaneous behavior, *conference on Face and Gesture Recognition* (2006)

Bartlet, M., Littlewort, I., Frank,G., Lainscsek, C., Fasel, M., Movellan, J.: (2006b) Automatic Recognition of Facial Actions in Spontaneous Expresions, *Journal of Multimedia*, Vol 1. No. 6 September 2006

Choe B. & Ko H.S., (2001) Analysis & synthesis of facial expression with hand generated muscle actuation basis, *Proc 14th Conference on Computer Animation*, pp.12-19, Seoul, South Korea, November 2001

Cohen, I., Sebe, N., Huang, T.S. (2003) Facial expression recognition from video sequences: Temporal and static modeling. *Computer Vision and Image Understanding*, 91(1-2) (2003)

Cohn, J., Kanade: (1999) Cohn-Kanade au-coded facial expression database. *Technical report,* Pittsburgh University 1999

Cohn, J.F., Schmidt, K., Gross, R., Ekman, P. (2002) Individual differences in facial expression: Stability over time, relation to self-reported emotion, and ability to inform person identification. *Proceedings of Intel. Conf. On Multimedia and Expo,* 2002

Cohn, J.(2005) Automated analysis of the configuration and timing of facial expression. (2005) Afterword of What the face reveals (2nd edition): Basic and applied studies of spontaneous expression using the FacialAction Coding System (FACS) by P. Ekman and E. Rosenberg, ed., 2005.

Cootes, T.F.; Edwards, G.J.; Taylor, C.J. (2001) Active appearance models, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 23, 6, (June 2001), pp. 681 – 685

Darwin, C (1872), *The expression of the emotions in man and animal,* University of Chicago Press, ISBN-13: 978-0195112719, Chicago, USA

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969) Pan-cultural elements in facial displays of emotions. *Science*, 1969, 164(3875), pp. 86-88

Ekman, P., Friesen, W., Hager, J. (1978) Facial Action Coding System , consulting psychologists press, Palo Alto,CA 1978

Ekman, P. Friesen, W. (1979) Constants across cultures in the face and emotion, Journal of personality and social psychology 1971

Ekman, P. Friesen, W. (1999a) Facial expression handbook of cognition and emotion, new york, john Wiley and sons ltd. Chapter 16. 2002

Ekman, P, Rosenberg, E., Hager, J., (1999b) Facial action coding system aect interpretive database FACSAID http://nirc.com/expression/facsaid/facsaid.htm 1999

Ekman, P., Friesen, W., Hager, J. (2002) Facial Action Coding System Manual. (2002)

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

Fawcett, T. (2003) Roc graphs: Notes and practical considerations for data mining researchers. (2003)

Ghent, J. (2005a). *A Computational Model of Facial Expression*, PhD thesis,National University of Ireland, Maynooth, Co. Kildare, Ireland, July 2005.

Ghent, J. & McDonald, J. (2005b) Photo-realistic facial expression synthesis.*Image and Vision Computing*, 23, 12, (November 2005), pp. 305–328

Goneid, A., el Kaliouby, R.: Facial feature analysis of spontaneous facial expression. In Proceedings of the 10th International AI Applications Conference (2002)

Gralewski, L, Campbell, N., Thomas, B., Dalton, C. & Gibson, D., (2004) Statistical synthesis of facial expressions for the protraying of emotion, *2nd international conf. on Computer graphics and interactive techniques in Australasia and South East Asia*, pp. 190–198, Singapore, June 2004

Kayo, nee Kouropteva, O., (2006). Locally Linear Embedding Algorithm. Extensions and Applications. PhD thesis, University of Oulu, Oulu, Finland,2006.

King, I. & Hou, H.T. (1996) Radial Basis Network for Facial Expression Synthesis, *Proceedings of the International Conference on Neural Information Processing*, pp. 1127-1130, Hong Kong, 1996,

Meytlis, M. & Sirovich, L.; On the Dimensionality of Face Space, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29, 7, (July 2007), pp. 1262 – 1267

Pantic, M., Rothkrantz, L.J.M. (2000) Automatic analysis of facial expressions:the state of the art. *IEEE transactions on pattern analysis and machine learning* 22 (2000)

Pantic, M., Patras, I.: Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *SMC-B* 36 (2006) 433–449

Raouzaiou, A., Tsapatsoulis, N. & Karpouzis, K. (2002). ParameterizedFacial Expression Synthesis Based on MPEG-4, *EURASIP Journal on Applied Signal Proccessing*, vol. 10**,** (October 2002), pp. 1021–1038

Reilly, J., Ghent, J., McDonald, J., 2006 Investigating the Dynamics of facialexpressions, *International symposium on visual computing* 2006

Reilly, J., Ghent, J., McDonald, J., 2007 Nonlinear Approaches towards the classification of facial expressions *International Machine Vision and Image Processing conference*, 2007

Saul, L. K., and Roweis, S. T., (2003). Think globally, fit locally: unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research*, 4(119), 2003.

Sayette M., Cohn, J F, Parrott, D.J (2001) Psychometric Evaluation of the Facial Action Coding System for Assessing Spontaneous Expression. (Springer Netherlands) 2001

Wang, H. & Ahuja, N. (2003) Facial Expression Decomposition, *Proceedings of International Conference on Computer Vision*, pp. 958-965, Nice, France, October 2003

Yangzhou, D. & Xueyin, L. (2003) Emotional facial expression model building, *Pattern Recognition Letters*, 24, 16, (December 2003), pp. 2923-2934.

Zhang, Q., Liu, Z., Guo, B., Terzopoulos, D. & Shum, H. (2006). Geometry-Driven Photorealistic Facial Expression Synthesis, *IEEE Trans on Visualization and Computer Graphics*, vol.12, no.1, (Jan/Feb 2006), pp. 48-60

Zheng, A. (2000) Deconstructing motion. Technical report, EECS department, U. C. Berkley (2000)

Affective Computing

Edited by Jimmy Or

ISBN 978-3-902613-23-3

Hard cover, 284 pages

Publisher I-Tech Education and Publishing

Published online 01, May, 2008

Published in print edition May, 2008

This book provides an overview of state of the art research in Affective Computing. It presents new ideas, original results and practical experiences in this increasingly important research field. The book consists of 23 chapters categorized into four sections. Since one of the most important means of human communication is facial expression, the first section of this book (Chapters 1 to 7) presents a research on synthesis and recognition of facial expressions. Given that we not only use the face but also body movements to express ourselves, in the second section (Chapters 8 to 11) we present a research on perception and generation of emotional expressions by using full-body motions. The third section of the book (Chapters 12 to 16) presents computational models on emotion, as well as findings from neuroscience research. In the last section of the book (Chapters 17 to 22) we present applications related to affective computing.

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen