Modelling the Manifold of Facial Expression using Texture

Jane Reilly and John McDonald
Computer Vision and Imaging Laboritory, Department of Computer Science
National University of Ireland Maynooth
JReilly@cs.nuim.ie

Abstract

The speed and intensity of the appearance changes that occur during the formation of facial expressions provide important information about the underlying meaning of the expression itself. In the past we have demonstrated the effectiveness of using Locally Linear Embedding with facial shape information for estimating the dynamics of facial expression. This approach was only suitable for specific expressions, where the appearance change was principally due to a movement or distortion of the shape of facial features. However, for some facial expressions, the variation in the shape of the facial features is very subtle. These expressions are mainly characterised by the variation in the texture of the face. Hence such expressions are not amenable to the previous approach. In order to estimate the dynamics of these types of expressions it is necessary to develop non-linear appearance models that incorporate texture information. In this paper we use LLE to estimate the manifold of texture variation due to facial expression. We show that the resulting manifold effectively captures the underlying dynamics of facial expression and that it provides a suitable representation for differentiation between posed and spontaneous expressions.

1. Introduction

Although the importance of facial expressions and the role which they play in human communication was first established in 1872 [8], it wasn't until the 1970's that behavioural scientists began to develop techniques which objectively measured facial expressions. Many different techniques were developed, such as the *Maximally Discriminative Facial Movement Coding System (MAX)* introduced by Izard in 1979 [15], and the *Facial Action Coding System (FACS)* introduced by Ekman and Friesen in 1978 [9]. The FACS is the most successful of these techniques and is widely used in research, for more details see Section 3.1 or [9].

Using the foundations laid down by behavioural scientists, over the past number of decades computer vision researchers have created systems specifically for the automatic analysis of facial expressions. Within the field of facial expression analysis there has been a significant amount of research investigating the six prototypical expressions (anger, fear, sadness, joy, surprise, and disgust). However, in everyday life, while these primary expressions occur frequently, when analysing human interaction and conversation, researchers have found that displays of emotion or intention are more often communicated by small subtle changes in the face's appearance [1].

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 [7, 3, 5, 1, 14]. The dynamics of facial expression can be defined as the intensity of the facial movement 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 [10].

In this paper we provide details of our technique for manifold based analysis of the dynamics of facial expression. We compare the success of our technique when applied to shape and texture information separately. We demonstrate the success of our texture based technique for modelling the dynamics of facial expression formation. We propose that our texture based model provides a good representation of the manifold of facial expression formation, providing the basis for the classification of not only the expression itsself but also the estimation of the intensity of that expression.

The remainder of this paper is structured as follows, in Section 2 we discuss the theory behind the analysis of the dynamics of facial expressions. In Section 3 we provide some background information on the techniques and methodologies which we implement in our research. Following on from this we demonstrate the success of our technique for modelling the dynamics of facial expression, discussing some experiments and work to date in Section 4. The research presented in this paper builds on our previous works presented in [18, 19] and [20].

978-0-7695-3332-2/08 $25.00 © 2008 IEEE
DOI 10.1109/IMVIP.2008.9

2. Dynamics of Facial Expression

According to Ambadar *et al.*, 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 which were not identifiable in individual images suddenly became apparent when viewed in a video sequence [1].

There is a growing trend in psychological research which argues that the dynamics of facial expression play a critical role in the interpretation of the observed behaviour. Zheng *et al.*, state that an expression sequence often contains multiple expressions of different intensities sequentially, due to the evolution of the subject's emotion over time [24].

2.1. Posed vs Spontaneous Facial Expressions

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 main difference between posed and spontaneous facial expressions is that posed expressions are captured by asking the subject to perform specific facial actions, whereas spontaneous facial expressions are more representative of what happens in the real world. Posed expression are captured in a controlled environment, whereas with spontaneous expressions 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 often 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 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.

Recently published research has shown that the dynamics of facial expression formation can be used to distinguish between posed and spontaneous expression of emotion. For example, Littlewort *et al.* developed a technique which differentiated between real and posed pain, achieving a 72% accuracy in a two-way forced choice [17]. Vural *et al.* used information relating to the timing and intensity of the appearance of the facial signals of tiredness, such as blink rate, eye closure and yawn to determine whether a driver was in a drowsy or alert state with 90% accuracy [23].

3. Proposed Methodology

In this section background information on the techniques which we implement for modelling the dynamics of facial expression formation are provided. We compare the effect of using textural data for modelling of the manifold of facial expressions as opposed to our previous works which use shape data alone. Our two experiments follow a similar structure were Locally Linear Embedding is applied to preprocessed datasets for both shape and texture. Details on how we preprocess our data is given in Section 3.2.

3.1. Facial Action Coding System - FACS

In this paper we use the *Facial Action Coding System* (FACS), which measures facial expressions according to the movement of muscles in the face. The FACS is based on an anatomical analysis of facial expressions [9], which allows us to subdivide our data into subsets where the variation in each expression is precisely characterized. The FACS provides an unambiguous, quantitative means of describing all movements of the face in terms of Action Units (AUs). An AU consists of the movement of one or more muscles in the face, causing an atomic change in the faces appearance. All expressions can be described using the AUs defined by the FACS, providing a measurable set of criteria that define whether or not a particular facial expression is present.

3.2. Preprocessing Techniques

In our research to date we have used the *Cohn-Kanade AU-Coded Facial Expression Database* (CK-database) [6]. This database contains approximately 2000 image sequences from over 200 subjects. The subjects came from a cross-cultural background and were aged approximately 18 to 30. The CK-database contains full AU coding and partial intensity coding of facial images and is the most comprehensive database currently available. Although the CK-database contains posed expressions which were captured under controlled conditions with the subject facing the camera, a certain amount of preprocessing was required prior to running our experiments.

For our shape experiment, we manually place 24 landmark points around the eyebrow region as shown in Figure 1. As we want to analyze the variance of the points that describe the shape of the eyebrows independent of identity,

prior to experimentation we align our data by performing *Generalized Procrustes Alignment* (GPA) [12]. GPA aligns two shapes with respect to position, rotation and scale by minimizing the weighted sum of the squared distances between the corresponding landmark points. More information on this technique can be found in [11]. Following on from this we perform *Shape Differencing*, whereby the neutral shapes for each sequence are subtracted from the sample set for that sequence.

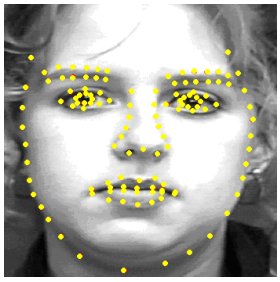

Figure 1. *Locations of the landmark points which we use to describe the shape of the face, in the experiments in this paper we are concerned with the 24 points which describe the shape of the eyebrow*

In contrast to our shape experiments, where our input data was a vector of xy coordinates corresponding to landmark points, with our texture experiments, our input data consisted of entire images of the subjects face. Due to the complex nature of this data, there are a number of issues which must be considered such as variations in illumination, scale and pose, along with interpersonal differences between subjects.

To counteract the effect that varying scale has on our experiments, we warped the subjects faces to a mean size. We did so by using the landmark points from our shape experiment to identify a bounding box around the eyebrow region. This box is used to crop out the required region of the face. These cropped boxes are then warped to a standard size, resulting in a matrix of images describing the texture of the eyebrow region.

As we want to uncover the underlying manifold of facial expression formation independent of identity, we implemented a technique called *Image Differencing*, which has been successfully used by Bartlett *et al.* as a preprocessing step in their experiments [2]. Image differencing works by subtracting the neutral image from each subsequent image in the sequence, and as a result the variances caused by varying identiy and light sources are minimised. For the purposes of clarity, the output images from this preprocessing when applied to the full face can be seen in Figure 2. In the experiments documented in this paper we apply this technique to the eyebrow region.

Figure 2. *Example of the results of performing image differencing on a pair of images, the neutral expression is shown on the left, and the extreme in the center. The resulting intensity difference image is shown on the right. For the purposes of clarity, the hue and saturation of the intensity image has been altered*

By performing these preprocessing steps, we are able to analyse the variances caused by facial activity, while minimizing the effect of scale, varying illumination and identity on our results. The application of shape and image differencing to our datsets 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.

3.3. Manifold of Facial Expressions

In this paper we are concerned with capturing the underlying manifold of facial expression formation, using the shape of the facial features, and also the texture of the appearance changes which occur. A manifold, in its simplest sense can be considered to be an abstract high dimensional space which is locally linear. Chang *et al.* proposed the concept of the manifold of facial expressions where each expression and its formation define a smooth underlying manifold in low dimensional space [4].

In order to capture the underlying manifold of an expression as it forms we apply dimensionality reduction techniques. According to Kayo et al., as real world data is often inherently nonlinear, linear dimensionality reduction techniques such as *Principal Component Analysis* (PCA), do not accurately capture the structure of this underlying manifold, i.e. relationships which exist in the high dimensional space are not always accurately preserved in the low dimensional space [16]. This means that in order to capture the underlying manifold of real world data, such as facial expressions, a nonlinear dimensionality reduction technique is required.

Over the past number of years a number of nonlinear dimensionality reduction techniques have been proposed, which are based on the concept of a smooth underlying manifold embedded in high dimensional data. See [22] for a comprehensive overview of these techniques. One of these nonlinear dimensionality reduction techniques is *Lo-*

cally Linear Embedding (LLE), which has been shown to provide a good visualisation of the underlying manifold of high dimensional data [16].

The LLE algorithm was introduced by Saul and Roweis in 2000 as an unsupervised learning algorithm that computes low dimensional, neighborhood preserving embeddings of high dimensional data [21]. The LLE algorithm is based on simple geometric intuitions where the algorithm attempts to compute a low dimensional embedding with the property 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 $\mathbf{X}_i$, each of dimensionality D, sampled from some smooth underlying manifold as its input. Provided there is sufficient data such that the manifold is well sampled, we can expect each data point and its neighbours to lie on or close to a locally linear patch of the manifold [21].

There are three main steps in the LLE algorithm. Firstly the manifold is sampled and for each sample, the *K* nearest neighbors are identified. Secondly each point $\mathbf{X}_i$ is approximated as a linear combination of its neighbors $\mathbf{X}_j$. These linear combinations are then used to construct the sparse weight matrix $\mathbf{W}_{ij}$. Reconstruction errors are then measured by the cost function Equation 1, which sums the squared distances between all the data points and their reconstructions.

$$\sum \mathbf{W} = \sum_i |\mathbf{X}_i - \sum_j \mathbf{W}_{ij}\mathbf{X}_i|^2 \qquad (1)$$

Finally each high dimensional observation $\mathbf{X}_i$ is mapped to a low dimensional $\mathbf{Y}_i$, which best preserves the geometry of $\mathbf{X}_i$'s neighborhood as represented by the weights $\mathbf{W}_{ij}$. Equation 2 is minimized by fixing the weights $\mathbf{W}_{ij}$ found by solving Equation 1, and finding the bottom d nonzero coordinates of each output $\mathbf{Y}_i$. For more details on the LLE algorithm see [21].

$$\phi\mathbf{Y} = \sum_i |\mathbf{Y}_i - \sum_j \mathbf{W}_{ij}\mathbf{Y}_i|^2. \qquad (2)$$

Effectively what this means is that the global coordinates Y_i, that best preserve the local relations between neighbouring points in the high dimensional space, are found over the lower dimensional manifold. As each individual coordinate is obtained only from local information within its neighborhood, the overlapping of each neighbourhood generates a global frame of reference.

While there are a number of issues with LLE and its poor generalisation to unseen data, extensions to the core algorithm which mitigate these issues have been proposed [16, 13].

4. Experimental Results

The aim of our experiments was to analyse the benefits of using textural information as opposed to shape data alone for capturing the underlying manifold of facial expressions involving the eyebrow. There are three AUs which effect the shape of the eyebrow namely; AU1 - *Inner Brow Raiser*, AU2 - *Outer Brow Raiser* and AU4 - *Brow Lowerer*. However, as there are no examples of AU2 in isolation within the CK-database, we were unable to include AU2 on its own in this experiment. The affect that the remaining AUs and their combinations have on the eyebrow region is shown in Figure 3.

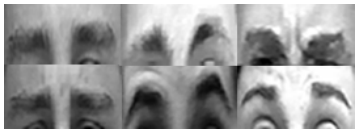

Figure 3. *Example of AU combinations that can occur in the eyebrow region which are contained within the CK-database, on the top row from left to right they are, the neutral expression AU0, AU1 and AU4, on the bottom row from left to right they are AU1+4, AU1+2 and AU1+2+4.*

In our first experiment, we were interested in capturing the underlying manifold of the variances caused by changes in the shape of the eyebrow using the expressions as shown in Figure 3. We did so by preprocessing our dataset as described in Section 3.2, applying the LLE algorithm to this preprocessed data. The resulting low dimensional space created by using the shape data can be seen in Figure 4.

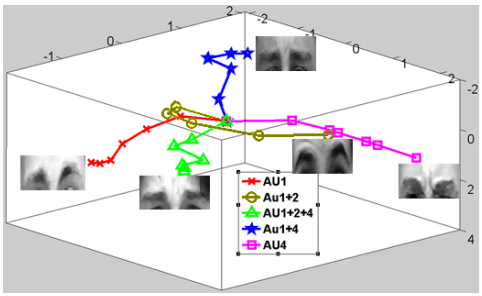

Figure 4. *Low dimensional representation of our shape data, as output from the LLE algorithm with K=13, and d=3.*

From Figure 4 we can see that the neutral expression is located in the centre of the space, and that the expression sequences for each of the five expressions radiate out from this neutral expression. Each of the points on the expression

lines represent an individual image from that sequence, going from the neutral expression in the centre to the extreme expression at the end of each line. The images displayed represent the final or extreme images in each of the image sequences, with the image at the top representing AU1+4, far left AU1, center left AU1+2+4, center right AU1+2 and far right AU4.

At first glance it appears that this low dimensional representation of the eyebrow data, has uncovered the underlying manifold of the expressions involving the eyebrow, in that for example we can see that the path of AU1+4 occurs mid way between the paths for AU1 and AU4. On closer inspection we can see that there are a number of inconsistencies in this expression space. For example the path for expression AU1+2 has been distorted as in the initial phases of the expression it appears to follow the direction of AU1, before doubling back to follow the path of AU4. Similarly, if we look at the expression paths for AU1+2 and AU1+2+4, we can see that the path for AU1+2+4 has been projected as being closer to AU1, than AU4, while the path for AU1+2 has been projected as being closer to AU4.

We hypothesise that the reasons behind these inconsistencies lie in the fact that in this experiment we are using shape data alone to analyse the manifold of the dynamics of the eyebrow region. Although alterations in the shape of facial features is one of the key indicators of the presence or absence of a particular AU, for certain AUs, these distortions do not always provide sufficient information. In order to identify the presence of some AUs, one must examine the other indicators, one such indicator is texture, such as the appearance and deepening of wrinkles on the face. An example of the textural changes that occur during the formation of AU4 are shown in Figure 5, where it can be observed that along with the distortion in the shape of the eyebrows, bulges and wrinkles also appear in the texture of the extreme expression.

Figure 5. *Effect that the presence of AU4 has on the appearance of the brow region. the neutral expression is shown on the left, and the extreme is shown on the right.*

Using this hypothesis, we performed this experiment again using textural information instead of shape data, to appraise the effectiveness of incorporating textural information for uncovering the manifold of the dynamics of eyebrow region. Firstly, we extracted textural data from the same subjects that were used in the shape experiment, and preprocessed this data as described in Section 3. The resulting low dimensional embedding can be seen in Figure 6. Again, the images portrayed in this figure are the final or extreme expressions for each expression sequence, clockwise from the top left the expressions portrayed are: AU1, AU1+4, AU4, AU1+2+4 and AU1+2.

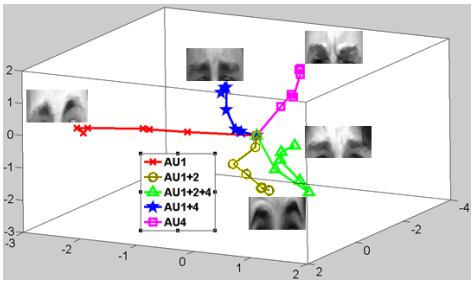

Figure 6. *Low dimensional representation of our texture data, as output from the LLE algorithm with K=13, and d=3.*

In Figure 6, it is clear that the use of textural information provides for a better representation of the underlying manifold of the expressions in the eyebrow region. We can see that similar expressions are located in the same regions of the space. Also unlike the shape space, this textural space enables us to analyse the dynamics of the expression formation. Where for example during the formation of AU1, the expression peaks one frame before the final frame in this sequence.

Another interesting aspect of this low dimensional space, is that during the formation of AU1+2+4, the subject did not progress directly from the neutral state to the expression AU1+2+4 as one would expect. Instead AU1+2 was formed first, raising the inner and outer eyebrows, before AU4 was applied to depress the brows. In Figure 7, we have zoomed in on this section of Figure 6, identifying key images in the sequence. This temporary appearance of AU1+2 during the formation of AU1+2+4 commonly happens when a subject is asked to pose the expression AU1+2+4. Whereas when a subject spontaneously performs AU1+2+4, the individual components that make up this movement occur simultaneously, in a coordinated manner.

The fact that this level of detail can be read from our low dimensional space directly, indicates that our technique provides the necessary foundations for the development of automatic methods for differentiating between posed and spontaneous expressions.

5. Conclusions and Future Directions

In this paper we have shown that using shape data in isolation provides an inconsistent representation of the mani-

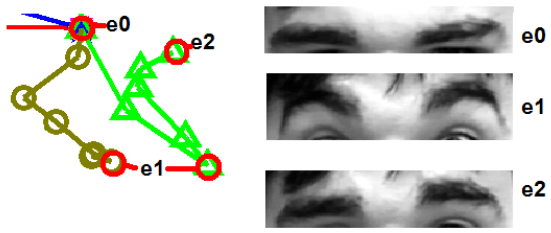

Figure 7. *Shows a zoomed in section of the Figure 6, where we can clearly see how during the formation of AU1+2+4, the subject first performed AU1+2, and then applied AU4. In this image, e0 is the neutral expression, e1 is AU1+2, and e2 is AU1+2+4*

fold of facial expressions. We have demonstrated the importance of textural information for extracting the manifold of the facial expressions involving.

We proposed that our texture based model provides a good representation of the manifold of facial expression formation and could be used as basis for the classification of not only the expression itsself but also the estimation of the intensity of that expression. The technique proposed in this paper provides the necessary foundations for the development of automatic methods for differentiating between posed and spontaneous expressions.

Future work will entail using these spaces as the basis for classifying not only the Action Unit present but also its intensity.

6. Acknowledgements

The research presented in this paper was conducted with the financial assistance of the Science Foundation Ireland.

References

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

[2] M. S. Bartlett. *Face Image Analysis by Unsupervised Learning*. Kluwer International Series on Engineering and Computer Science V. 612. Boston. Kluwer International Series on Engineering and Computer Science V. 612. Boston, 2001.

[3] M. S. Bartlett, J. Movellan, G. Littlewort, B. Braathen, M. G. Frank, and T. J. Sejnowski. Towards automatic recognition of spontaneous facial actions. *In Paul Ekman, editor, what the face reveals*, 2003. Oxford University Press.

[4] Y. Chang, C. Hu, and M. Turk. Manifold of facial expression. *IEEE International Workshop on Analsysi and Modelling of Faces and Gestures*, pages 25–35, 2003.

[5] I. Cohen, N. Sebe, and T. L.C.A.G.and Huang. Facial expression recognition from video sequences: Temporal and static modeling, (2003).

[6] J. Cohn and Kanade. Cohn-kanade au-coded facial expression database. Technical report, Pittsburgh University, 1999.

[7] J. F. Cohn, K. Schmidt, R. Gross, and P. Ekman. 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, 2001.*, 2002.

[8] C. Darwin and P. Ekman. *The expression of the emotions in man and animal*. Chicago: The University of Chicago Press, 1998. 1st edition in 1872, 2nd edition in 1889, 3rd edition with additional commentry by Paul Ekman in 1998.

[9] P. Ekman, W. Friesen, and J. Hager. Facial action coding system. *Consulting Psychologists Press*, 1978.

[10] P. Ekman, W. Friesen, and J. Hager. *Facial Action Coding System Manual*, 2002.

[11] J. Ghent. *A Computational Model of Facial Expression*. PhD thesis, National University of Ireland Maynooth, Co. kildare, Ireland, July 2005.

[12] J. C. Gower. Generalised procrustes analysis. *Psychometrika*, 40:33–50, 1975.

[13] A. Hadid and M. PietikAanien. Efficient locally linear embeddings of imperfect manifolds. *Proceedings of the Third International Conference on Machine Learning and Data Mining in Pattern Recognition, Leipzig, Germany*, pages 188–201, 2003.

[14] A. Hadid and M. Pietikdinen. An experimental investigation about the integration of facial dynamics in video-based face recognition. *ELCVIA*, 5(1):1–13, March 2005.

[15] C. E. Izard. The maximally discriminative facial movement coding system (max),. *Newark, Del.: University of Delaware, Instructional Resource Center.*, 1979.

[16] O. Kayo, nee Kouropteva. *Locally Linear Embedding Algorithm. Extensions and Applications.* PhD thesis, University of Oulu, Oulu, Finland, 2006.

[17] G. Littlewort, M. Bartlett, and K. Lee. Automated measurement of spontaneous facial expressions of genuine and posed pain. In *Proc. International Conference on Multimodal Interfaces, Nagoya, Japan.*, 2007.

[18] J. Reilly, J. Ghent, and J. McDonald. Investigating the dyanmics of facial expression. *Proceedings of the 2nd International Symposium on Visual Computing*, November 2006.

[19] J. Reilly, J. Ghent, and J. McDonald. Non-linear approaches for the classification of facial expressions at varying degrees of intensity. *Proceedings of the Irish Machine Vision and Image Processing Conference 2007*, September 2007.

[20] J. Reilly, J. Ghent, and J. McDonald. *Affective Computing, focus on Emotion Expression, Synthesis and Recognition*, chapter 1 Modelling, Classification and Synthesis of Facial Expressions, pages 1–26. I-Tech, 2008.

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

[22] L. J. P. van der Maaten, E. O. Postma, and H. J. van den Herik. Dimensionality reduction: A comparative review. Technical report, MICC, Maastrict University, 2008.

[23] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan. Drowsy driver detection through facial movement analysis. In *In Proc ICCV*, 2007.

[24] A. Zheng. Deconstructing motion. Technical report, EECS department, U. C. Berkley, 2000.