

# Investigating the Dynamics of Facial Expression

Jane Reilly, John Ghent, and John McDonald

National University of Ireland, Maynooth

**Abstract.** This paper is concerned with capturing the dynamics of facial expression. The dynamics of facial expression can be described as the intensity and timing of a facial expression and its formation. To achieve this we developed a technique that can accurately classify and differentiate between subtle and similar expressions, involving the lower face. This is achieved by using *Local Linear Embedding* (LLE) to reduce the dimensionality of the dataset and applying *Support Vector Machines* (SVMs) to classify expressions. We then extended this technique to estimate the dynamics of facial expression formation in terms of intensity and timing.

## 1 Introduction

Darwin, in 1872, was the first to recognize the importance of facial expressions and the role that they play in human communication [1]. During the subsequent years as behavioral scientists sought a means to objectively measure facial expressions, many different techniques and methodologies for describing facial expressions were developed (see [2] for a comprehensive review). The *Facial Action Coding System* (FACS), created by Ekman and Friesen, in 1978, is the most comprehensive of these standards and is widely used in research. The FACS provides an unambiguous quantitative means of describing all movements of the face in terms of 46 *Action Units* (AUs) [3].

Recent research has shown that it is not just the expression itself, but also its dynamics that are important when attempting to decipher its meaning [4, 5, 6, 7, 8, 9, 10]. 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, namely the temporal information itself [11]. There is now a growing body of psychological research that argues that temporal dynamics of facial behavior - such as the timing, duration and intensity of facial activity - is a critical factor for the interpretation of the observed behavior [4]. This paper aims to not only classify expression but also capture the dynamics of facial expression formation.

In this paper we extend our previous work [12] to classify subtle changes in expression and capture the dynamics of facial expression. To achieve this we use *Locally Linear Embedding* (LLE). LLE was originally proposed in 2000 as a non-linear dimensionality reduction technique that computes low-dimensional neighborhood preserving embeddings of high-dimensional data by unfolding the underlying manifold [13]. We show that this non-linear dimensionality reduction

technique provides a means for classifying subtle changes in expression and allows for analysis of the dynamics of facial expression. In other words, if LLE can capture the underlying manifold of facial expression independent of identity, it would then provide information about the timing and intensity of that expression, hence capturing the dynamics of facial expression.

This paper is organized as follows. In Section 2, recent research on the importance of the dynamics of facial expressions is introduced from a behavioral science perspective. Following this we survey a number of computer vision based automated techniques for the analysis of both the appearance and dynamics of facial expression. Our proposed methodology is presented in Section 3. Experiments are presented in Section 4. Conclusions and a discussion on the results of these experiments is provided in Section 5.

## 2 Related Work

In this section we first review the state-of-the-art in expression analysis from a behavioral science point of view. Subsequent to this we detail computer vision based approaches for the automated analysis of facial expression.

### 2.1 Behavioral Science

Since the importance of facial expressions was first established in 1872 [1], many studies have been carried out attempting to interpret their meaning. However, according to Ambadar et al. [5], only a few investigators have examined the impact of dynamics in deciphering faces. These studies were largely unsuccessful due to their reliance on intense facial expressions. While the primary reasoning behind the research carried out by Ambadar et al. was to establish the importance of dynamics in deciphering facial expressions, it also highlighted the fact that facial expressions are frequently subtle. They found that subtle expressions that were not identifiable in static presentations suddenly became apparent in dynamic display.

Schmidt et al. studied the normative characteristics and individual differences that occur in spontaneous smiles. They found that spontaneous smiles, in contrast to posed smiles, have a relatively fast onset, averaging slightly over 0.5 seconds. Approximately one third of smiles included in the study consisted of AU 12 (Lip Corner Puller)<sup>1</sup> on its own, while the remaining smiles had one or more co-occurring AUs appearing within one second. [14]

According to Cohn, further development within the area of Automatic Facial Action Unit Recognition will eliminate the need for manual FACS coding [15]. This would prove to be a significant step forward as currently it takes approximately 100 hours of training to become even marginally proficient with FACS, and it takes 1 hour for an experienced FACS coder to encode one minute of video footage [3].

---

<sup>1</sup> for more information on AUs see the FACS manual [11].

The results of studies carried out by Cohn et al. confirmed human perception of facial expressions is heavily influenced by dynamics [5]. A system that could automatically FACS code the dynamics of facial expression, would be an important advance in the field of facial expression analysis in that it would provide key information on the timing and intensities of the various AUs.

## 2.2 Computer Vision Approaches

There are many different techniques being currently applied to the problems of automatic facial expression recognition and automatic FACS coding such as: Expert Systems [16], Hidden Markov Models [7], Gabor filters [6, 17, 18, 19, 20, 21, 22, 23, 24] and Optical Flow Analysis [25]. For an overview of the current techniques see [26]. Facial expression dynamics, have also been used to assist the identification of individuals [4,9], where it was proposed that there are individual differences in how expressions are formed, it is suggested that this information could be used to augment facial recognition systems.

While 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 the study of facial expression. In their most recent publication [24] they present a system that accurately performs automatic recognition of 20 AUs from near frontal-view face image sequences using Gabor filters and Support Vector Machines.

Despite the fact that this system has the potential to provide information about expression dynamics, it has not as yet been fully explored. However, a system that could fully analyze the dynamics of facial expression in terms of intensity and timing would represent a major advancement in the field of facial expression analysis.

## 3 Proposed Methodology

In this paper we aim to classify expression and extract information regarding the dynamics of expression by first projecting shapes of individuals portraying specific expressions into a lower dimensional *Locally Linear Embedding* (LLE) space in order to capture an underlying manifold of that expression as it forms. And secondly, to classify that expression using *Support Vector Machines* (SVMs). Information about the dynamic of facial expression is then represented in the LLE space. In this section both the LLE and SVM algorithm are briefly discussed.

### 3.1 Locally Linear Embedding (LLE)

LLE is a non-linear dimensionality reduction technique that is capable of generating highly non-linear embeddings, and its main optimization involves a sparse eigenvalue problem [27]. The LLE algorithm is based on simple geometric intuitions. Essentially 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.

In 2000, the LLE algorithm was introduced by Saul and Roweis as an unsupervised learning algorithm that computes low dimensional, neighborhood preserving embeddings of high dimensional data [13, 27, 28]. Many extensions and adjustments to this core algorithm have been proposed, ranging from Robust-LLE [29, 30] to supervised [31, 32] and semi-supervised versions of LLE [33], however in this paper we are using the original algorithm as defined in [13].

### 3.2 Support Vector Machines (SVM)

SVMs are a type of learning algorithm based upon recent advances in statistical learning theory [34, 35, 36, 37] 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  $(x_1, y_1), \dots, (x_m, y_m) \in \mathbf{X} \times \{\pm 1\}$  where  $\mathbf{X}$  is some space from which the  $x_i$  have been sampled. We can construct a dual Lagrangian of the form

$$W(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (1)$$

which are subject to the constraints

$$\alpha_i \geq 0 \quad \forall i \quad \text{and} \quad \sum_{i=1}^m \alpha_i y_i = 0. \quad (2)$$

The solution to Equation 1 is a set of  $\alpha$  values which defines a hyperplane that is positioned in an optimal location between two classes. See [38], for more information.

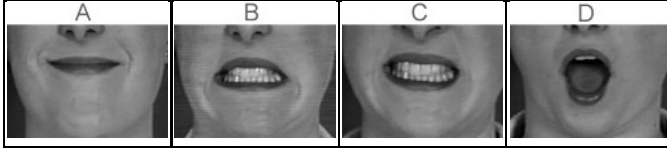
## 4 Experiments and Results

In this section we detail two experiments. The first is designed to classify which AU or combination of AUs a person is portraying, and the second is designed to extract information regarding the intensity and timing of a classified AU and hence provide a means for analysing the dynamics of facial expression. The data used for these experiments comes from the Cohn-Kanade [39] database.

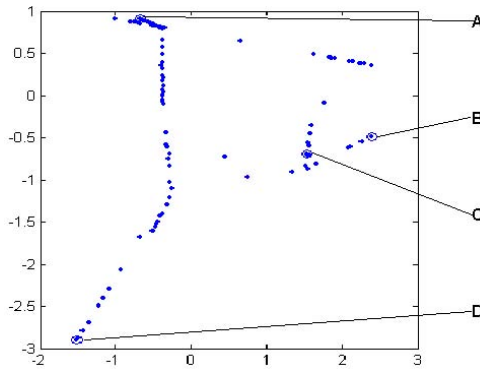
### 4.1 Classifying Action Units

In this experiment we classify four lower facial expressions; AU20+AU25, AU12, AU10+AU20+AU25 and AU25+AU27. The effect that these have on the mouth is illustrated in Table 1. Our training data consisted of 141 images of an individual performing 8 sequences from neutral to extreme expression. A single individual was used to develop the LLE space as this ensures that any variance

**Table 1.** This table illustrates the effect of portraying four different expressions. The AUs portrayed, from left to right are; AU12, AU20+AU25, AU10+AU20+AU25, AU25+AU27.



in the LLE space is due to expression change rather than subject change. It should also be noted that two expressions are extremely similar and therefore it is hypothesised that a technique that can differentiate between these expressions can accurately classify subtle changes in appearance. During training 24 landmark points are manually placed at 24 specific locations on the mouth. We align the data to remove any unnecessary variation before applying the LLE algorithm. This is achieved by centering the data and then performing *Generalised Procrustes Alignment* (GPA) [40]. This technique aligns two shapes with respect to position, rotation and scale by minimising the weighted sum of the squared distances between the corresponding landmark points. More information on this technique can be found in [38]. We then calculate a shape model by performing LLE on the aligned training data. The LLE output using the aligned data as input is shown in Figure 1. We project new unseen data into the LLE space and



**Fig. 1.** Aligned LLE shape space. The Action Units portrayed, from top to bottom are; AU12, AU20+AU25, AU10+AU20+AU25, AU25+AU27.

use these outputs as inputs to the SVM classifier. As there exists four expressions to be separated, a one-against-all approach yields four separate SVM classifiers. This approach requires, at most, four evaluations to acquire a result. The results from the one-against-all SVM approach in conjunction with LLE are detailed in Table 2.

**Table 2.** LLE SVM results. In the table above A = AU20+AU25, B = AU25+AU27, C = AU10+AU20+AU25 and D = AU12.

AU	C	$\sigma$	Ts
A-v-all	6.1	0.1	84.6
B-v-all	6	0.3	83.8
C-v-all	7	0.8	98.0
D-v-all	8	0.5	98.7
Average			91.3

In Table 2,  $C$  is the soft margin variable,  $\sigma$  is the kernel parameter, and  $Ts$  is the percentage of correctly classified test data. The test data consists of 10 subjects (60 images) portraying AU20+AU25, 10 subjects (60 images) portraying AU25+AU27, 3 subjects (18 images) portraying AU10+AU20+AU25 and, 3 subjects (18 images) portraying AU12. There are low, medium and high intensity examples of each expression in the training set. The soft margin variable and the kernel parameter are calculated using a cross validation approach [38].

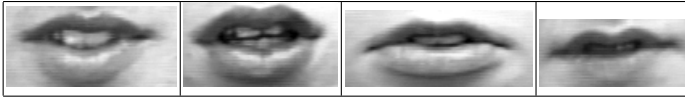
As can be seen from Table 2, this technique accurately classifies similar expressions with an average accuracy of 91.3%. This improves on our previous work as our PCA approach could only classify extreme appearance changes [38].

## 4.2 Examining the Intensity of Action Units

In this experiment information about the dynamics of facial expressions is extracted. The intensity over time of a composite AU lower facial expression and a single AU lower face expression are estimated according to the FACS intensity coding [11]. The FACS intensity coding ranges from A to E, with A representing a minor change in appearance, and E an extreme change in appearance. The input to these two experiments consisted of 2 datasets of mouth shapes taken from multiple subjects. These datasets consisted of multiple sequences each of a different subject. For each sequence we manually 6 frames from neutral to extreme. For AU20+AU25 a total of 10 subjects (60 frames) were landmarked, and 24 subjects (144 frames) were landmarked for the experiment on AU25.

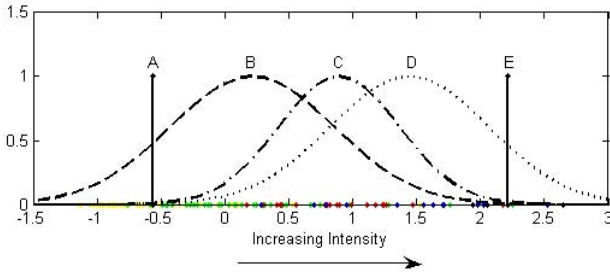
Similar to the previous experiment, the mouth shapes were first aligned using GPA. As LLE has a well documented tendency to cluster data into identity if the data contains more than one identity, and our datasets contain multiple identities, a further preprocessing step was required. For this we use a technique called *Shape Differencing*, whereby the neutral mouth shapes of each subject was subtracted from the sample set for that subject. As we are only interested in the difference between these two shapes and not the actual shapes themselves, by using shape differencing, the variances occurring due to the identity will be removed. This is a valid step as to analyse the dynamics of facial expression it is necessary to have a sequence of images containing a neutral image.

**Table 3.** AU25 classified mouths

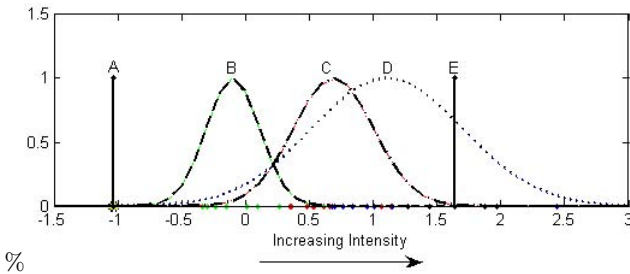


The images that were labeled as extreme in our datasets varied between each sequence, for example certain sequences only showed AU25 at an intensity of C. Table 3, shows a sample of images that were FACS coded AU25E. To deal with this variation, prior to running our experiments we manually FACS intensity coded the entire datasets.

LLE was then applied to the datasets, reducing the dimensionality to one dimension. We take the projections of the samples for intensities B, C, and D in the one dimensional LLE space and fit Gaussians to this data. Due to lack of samples for intensities A and E we just calculate the mean. The resulting Gaussians and means are shown in Figure 2 and Figure 3. As can be seen from these figures our technique accurately characterises the intensity of the facial expressions over time and hence provides a means for modelling their dynamics.



**Fig. 2.** Intensity over time for AU25 results



**Fig. 3.** Intensity over time for AU20 + AU25 results

## 5 Conclusion and Discussion

The accurate classification of facial expressions is a growing problem within several domains. The solution described in this paper takes a multidisciplinary approach drawing together psychological tools, statistical models and machine learning techniques. We first build a shape model that was based on an anatomical analysis of facial expression - FACS. The FACS provided us with a universal method of analyzing facial expression and allowed for the classification of facial expressions independent of identity.

In our first experiment, a shape model was calculated by using LLE to lower the dimensionality of the problem. One-against all SVMs were used to classify four lower face expressions (AU20+AU25, AU25+AU27, AU10+AU20+AU25 and AU12) resulting in an accuracy of 91.3% percent. While our previous work produced an accuracy of up to 99% percent with extreme expressions, it was unsuccessful at classifying subtle changes in expression [38] .

In our second experiment, multiple identities were used to build up intensity models for two lower face expressions. This intensity model demonstrates the changes in intensity over time, and can automatically FACS intensity code AUs.

The approach presented in this paper is an improvement on our previous work as it allows us to classify subtle changes in expression. We have shown that a low-dimensional non-linear representation of expression improves the accuracy of our classification results, enabling the classification of expression as it forms, and also providing an accurate means of classifying the intensity of AUs. This intensity information, in conjunction with timing information, provides the necessary basis for the automated analysis of facial expression dynamics.

This paper show that LLE provides an effective means of both classifying subtle changes in facial expression and estimating the intensity of facial. Future work will entail applying this technique to more comprehensive data sets incorporating a larger variety of AUs and individuals.

## Acknowledgements

This publication has emanated from research conducted with the financial support of Science Foundation Ireland.

## References

1. Darwin, C., Ekman", P.: 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.
2. Cohn, J., Kanade, T., Moriyama, T., Ambadar, Z., Xiao, J., Gao, J., Imamura, H.: A comparative study of alternative facs coding algorithms. Technical Report CMU-RI-TR-02-06, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (2001)
3. Ekman, P., Friesen, W., Hager, J.: Facial action coding system. Consulting Psychologists Press (1978)



4. Cohn, J.F., Schmidt, K., Gross, R., Ekman, P.: 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)
5. Ambadar, Z., Schooler, J., Cohn, J.: Deciphering the enigmatic face: The importance of facial dynamics to interpreting subtle facial expressions. *Psychological Science* (2005)
6. Bartlett, M.S., Movellan, J., Littlewort, G., Braathen, B., Frank, M.G., Sejnowski, T.J.: Towards automatic recognition of spontaneous facial actions. In Paul Ekman, editor, *what the face reveals* (2003) Oxford University Press.
7. I. Cohen, N. Sebe, L.C.A.G., Huang, T.S.: Facial expression recognition from video sequences: Temporal and static modeling. *Computer Vision and Image Understanding*, 91(1-2) (2003)
8. Ekman, P.: Facial expression of emotion: New findings, new questions. (1992)
9. Hadid, A., Pietikinen, M.: An experimental investigation about the integration of facial dynamics in video-based face recognition. *ELCVIA* 5 (2005) 1–13
10. Littlewort, G., Bartlett, M.S., Fasel, I., Susskind, J., Movellan, J.: Dynamics of facial expression extracted automatically from video. *Computer vision and Image understanding* (2006)
11. Ekman, P., Friesen, W., Hager, J.: *Facial Action Coding System Manual*. (2002)
12. Ghent, J., McDonald, J., Harper, J.: Statistical model for expression generation using the facial action coding system. Technical Report NUIM-CS-TR-2003-01, NUI Maynooth (2003)
13. Saul, L.K., Roweis, S.T.: An introduction to locally linear embedding. <http://www.cs.toronto.edu/~roweis/lle/publications.html> (2001)
14. Schmidt, K., Cohn, J.: Dynamics of facial expression: Normative characteristics and individual differences. In: *IEEE International Conference on Multimedia and Expo ( ICME2001)*. (2001) 728 – 731
15. Cohn, J.: 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 Facial Action Coding System (FACS) by P. Ekman and E. Rosenberg, ed., 2005.
16. 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
17. Zhang, Y., Ji, Q.: Facial expression understanding in image sequences using dynamic and active visual information fusion. *ICCV* (2003) 113 to 118
18. Donato, G., Bartlett, M., Hager, J., Ekman, P., Sejnowski, T.: Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(10) p. 974-989. (1999)
19. Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Recognising facial expression: Machine learning and application to spontaneous behaviour. *IEEE International conference on computer vision and pattern recognition* (2005)
20. 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
21. M. Stewart Bartlett, B. Braathen, G.L.T.J.S., Movellan, J.: Automatic analysis of spontaneous facial behavior: A final report. (2001) MPLAB-TR-2001-06, Institute for Neural Computation, University of California, San Diego.

22. M.S., Bartlett, G.L.I.F.J.M.: Real-time face detection and facial expression recognition: Development and applications to human computer interaction. (2003)
23. Bartlett, M.S., L.G.B.B.S.T..M.J.: A prototype for automatic recognition of spontaneous facial actions. (2003)
24. Bartlett, M., Littlewort, G., Lainscsek, C., Fasel, I., Frank, M., Movellan, J.: Fully automatic facial action recognition in spontaneous behavior bartlett. (2006)
25. Goneid, A., el Kaliouby, R.: Facial feature analysis of spontaneous facial expression. In Proceedings of the 10th International AI Applications Conference (2002)
26. Fasel, B., Luetttin, J.: Automatic facial expression analysis: A survey. *Pattern Recognition*, Vol. 36(1) (2003) 259–275
27. Saul, L.K., Roweis, S.T.: Think globally, fit locally: unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research* 4 (2003)
28. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by local linear embedding. *Science* 290 (2000) 2323–2326
29. Chang, Y., Yeung, D.Y.: Robust local linear embedding. Technical report, Department of Computer Science, Hong Kong University of Science and Technology, HKUST-CS05-12 (2005)
30. Hadid, A., PietikAanien, M.: Efficient locally linear embeddings of imperfect manifolds. Proceedings of the Third International Conference on Machine Learning and Data Mining in Pattern Recognition, Leipzig, Germany (2003) 188–201
31. Ridder, D.D., Kouropteva, O., Okun, O., PietikAanien, M., Duin, R.P.W.: Supervised locally linear embedding. ICANN 2003 (2003) 333–341
32. Kouropteva, O., Okun, O., Hadid, A., Soriano, M., Marcos, S., PietikAainen, M.: Beyond locally linear embedding algorithm. Technical report, Department of Electrical and Information Engerring, University of Oulu, Oulu, Finland, MVG-01-2002 (2002)
33. Ridder, D.D., Duin, R.P.W.: Locally linear embedding for classification. Technical report, Pattern Recognition Group, Department of Imaging Science and Technology, Delft University of Technology, Delft, The Netherlands, PH-2002-01 (2002)
34. Campbell, C.: kernel methods: A survey of current techniques. *Neurocomputing* 48 (2002) 63–84
35. Rogers, S., Williams, R.D., Campbell, C.: Class prediction with Microarray Datasets. In: *BioInformatics with computational intelligence paradigms*. Springer-Verlag (2004)
36. Rogers, S.: *Machine Learning Techniques for Microarray Analysis*. Faculty of engineering mathematics, University of Bristol (2004)
37. Scholkopf, B., Smola, A.J.: *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press (2002)
38. Ghent, J.: *A Computational Model of Facial Expression*. PhD thesis, National University of Ireland Maynooth, Co. kildare, Ireland (2005)
39. Cohn, J., Kanade: Cohn-kanade au-coded facial expression database. Technical report, Pittsburgh University (1999)
40. Gower, J.C.: Generalised procrustes analysis. *Psychometrika* 40 (1975) 33–50