COMPUTATIONAL ANALYSIS OF FACIAL EXPRESSIONS

ARUNA SHENOY

A thesis submitted in partial fulfilment of the requirements of the University of Hertfordshire for the degree of Doctor of Philosophy

The programme of research was carried out in the School of Computer Science, University of Hertfordshire

September 2009

Abstract

This PhD work constitutes a series of inter-disciplinary studies that use biologically plausible computational techniques and experiments with human subjects in analyzing facial expressions.

The performance of the computational models and human subjects in terms of accuracy and response time are analyzed. The computational models process images in three stages. This includes: Pre-processing, dimensionality reduction and Classification. The pre-processing of face expression images includes feature extraction and dimensionality reduction. Gabor filters are used for feature extraction as they are closest biologically plausible computational method. Various dimensionality reduction methods: Principal Component Analysis (PCA), Curvilinear Component Analysis (CCA) and Fisher Linear Discriminant (FLD) are used followed by the classification by Support Vector Machines (SVM) and Linear Discriminant Analysis (LDA).

Six basic prototypical facial expressions that are universally accepted are used for the analysis. They are: angry, happy, fear, sad, surprise and disgust. The performance of the computational models in classifying each expression category is compared with that of the human subjects. The *Effect size* and *Encoding face* enable the discrimination of the areas of the face specific for a particular expression. The *Effect size* in particular emphasizes the areas of the face that are involved during the production of an expression. This concept of using *Effect size* on faces has not been reported previously in the literature and has shown very interesting results.

The detailed PCA analysis showed the significant PCA components specific for each of the six basic prototypical expressions. An important observation from this analysis was that with Gabor filtering followed by non linear CCA for dimensionality reduction, the dataset vector size may be reduced to a very small number, in most cases it was just 5 components. The hypothesis that the average response time (RT) for the human subjects in classifying the different expressions is analogous to the distance measure of the data points from the classification hyper-plane was verified. This means the harder a facial expression is to classify by human subjects, the closer to the classifying hyper-plane of the classifier it is. A bi-variate correlation analysis of the distance measure and the average RT suggested a significant anti-correlation. The signal detection theory (SDT) or the *d-prime* determined how well the model or the human subjects were in making the classification of an expressive face from a neutral one. On comparison, human subjects are better in classifying surprise, disgust, fear, and sad expressions. The RAW computational model is better able to distinguish angry and happy expressions.

To summarize, there seems to some similarities between the computational models and human subjects in the classification process.

Acknowledgements

Completing a PhD is truly a marathon event, and I would not have been able to complete this journey successfully without the help and support of countless people over the past three years. First and foremost I thank the University of Hertfordshire for providing me the studentship without which I would not have been able to undertake this PhD.

I would like to thank my supervisors: Neil Davey, Ray Frank, Tim Gale and Sue Anthony for their guidance and support during my PhD. I am especially thankful to my supervisor Neil Davey who not only supervised my work, but also hugely encouraged me to write papers. His cheerful nature, enthusiasm and everlasting energy are a source of inspiration to anyone. I thank Tim for helping me to understand the various aspects of psychology which was very new to me when I started my PhD. I thank Ray for helping with the issues in programming and his critical analysis of this thesis. I am also grateful to Sue Anthony for her wonderful support and guidance during the last year of my PhD especially with statistical analysis.

I thank my office colleagues Angela, Samar, Yi, Weiliang, Nicolos, Faizal, Johannes and Giseli for their great company and friendship. I have really enjoyed the company of Yi in particular, and am thankful to her for always being there as a very good friend.

I am very grateful to my parents, Ananthakrishna and Chithra for their unwavering support and belief in me without which I could not have been able to pursue my PhD. I am thankful to my mother for instilling in me from a very young age the desire for pursuing higher education and for the sacrifices she has made for me to be able to achieve them. I thank my elder sister Anu and my younger brother Aravind for their love and encouragement. I am thankful to my husband Sampath, whose love and unwavering support has enabled me to complete my Thesis. Completing this PhD in time has been quite a challenge as my son Suraj was just one year old at the commencement of my PhD. He has always been my main source of energy. I owe my achievement to my husband and my son.

Finally, I express my gratitude to the participants of the experiment, who extended their best cooperation in this endeavour.

Table of Contents

List of Tables

List of Figures

Publications and Presentations

Conference Proceedings

Shenoy, A., Gale, T.M., Frank, R. J. and Davey, N. 2007. Recognizing emotions by analyzing facial expressions. *UK Workshop on Computational Intelligence*. London. UK. [*Paper presented as a talk*].

Shenoy A., Gale T.M., Frank, R.J. and Davey, N. 2008. On the Recognition of emotions from facial expressions. Doctoral Symposium, *ACM Compute* 2008. Bangalore, India. [*Paper presented as a talk*].

Shenoy A., Gale T.M., Davey, N., Christansen, B., and Frank, R.J. 2008. Recognizing facial expressions: A comparison of Computational approaches. *International Conference on Artificial Neural Networks.* Prague, Czech Republic. This is also published in *Lecture notes in Computer Science - Artificial Neural Networks*, Vol. 5163, 2008. [*Paper presented as a talk*].

Shenoy, A., Anthony, S., Frank, R.J. and Davey, N. 2009. Discriminating Angry, Happy and Neutral facial Expression: A comparison of computational models. *International conference on Engineering Applications of Neural Networks*. London, UK. This is also published in *Communications in Computer and Information Science*. Vol 43. 2009 [*Paper presented as a talk*]. *This paper is now selected from the best papers to be extended and reviewed for inclusion in a special issue of the Springer journal, Neural Computing and Applications.*

Abstracts

Shenoy A., Gale T.M., Frank, R.J. and Davey, N. 2008. Representation and Classification of facial expression in a modular computational model. Neural Computation and Psychology Workshop. Oxford. UK. [*Abstract presented as a poster*].

Shenoy, A., Davey, N., Frank, R.J., Gale, T.M. 2008. A computational model of facial expressions: Classification and representation. *International conference on Cognitive and Neural Systems*. Boston, USA. [*Abstract presented as a poster*].

Shenoy, A., Anthony, S., Frank, R.J. and Davey, N. 2009. A comparison of the performance of humans and computational models in the classification of facial expression. *International conference on Cognitive Modelling*. Manchester, UK. [*Abstract presented as a poster*].

CHAPTER ONE

Introduction

1.1 Motivation

Facial expressions are an important part of social communication. They give an opportunity to both convey and understand emotions. The generation and recognition of facial expressions are two related, but distinct, aspects of this area of study. However, in normal day-to-day social circumstances they are equally important. This thesis concentrates only on analyzing facial expressions. The process of learning to understand the facial expressions of other people starts very early. The ability to recognize a facial expression as genuine or fake helps in making judgements and in responding accordingly. Emotions are conveyed through body language and voice; however, the main component of emotion display is by facial expression.

Darwin (1872) found that facial expression generation was universal and the same for all people across the globe. Later studies by Ekman and Friesen (1973) confirmed that there are six basic prototypical expressions namely, anger, happiness, fear, sadness, surprise and disgust. They also suggested that these expressions are universal across the various cultures in the world. A recent study that compared the expressions of blind and non-blind individuals suggests that the production of spontaneous facial expressions of emotions is innate (Matsumoto and Willingham, 2009). This indicates that some genetic wiring may be responsible for the generation of facial expressions of emotions. Studies on facial expression generation and recognition have been conducted with different types of experiments and tasks. Recent work in this involves designing artificial but biologically plausible facial expression recognition systems (Lyons *et al.*, 1998; Shen, 2005; Dailey, 2002; Liu and Wang, 2006).

With various facial expression recognition systems developed, a number of successful algorithms have been studied in the field of Computer Science. There has been an understanding that the theories, studies and results that have been obtained by psychologists may be successfully used to develop more efficient facial expression recognition systems (Pantic and Bartlett, 2007; Zheng *et al.*, 2009; Fasel and Luettin, 2003). In developing better biologically plausible computational systems, a further step may, in turn, be taken towards understanding and analyzing facial expression processing by humans.

The objective of this thesis is to study computational models for facial expression analysis using biologically plausible feature extraction techniques and dimensionality reduction methods.

Moreover, the results of this analysis are compared with those obtained from human subjects asked to perform a related task.

Generally, a typical facial expression recognition system has a cascade of three stages: pre-processing, dimensionality reduction and classification. Normally face images are of very high dimensions and may need efficient dimensionality reduction methods to provide good classification results. When the number of images increases, the need to use dimensionality reduction techniques also increases. In this thesis pre-processing techniques to extract features of the image and some dimensionality reduction methods have been discussed. The facial features such as: eyebrows, eyes, nose and chin play a prominent role in the recognition of facial expressions. Facial expressions are registered as changes in these features and their alignment (Ekman and Friesen, 1976; Ekman and Friesen, 1978; Hager, 2006). Chapter 3 discusses the computational techniques that pre-process images and extract the necessary features that enable efficient recognition. Once these features are extracted, they can be reduced in dimensionality and later categorized by a suitable classifier.

Chapter 3 discusses the necessary background of the pre-processing method for feature extraction that has been utilized in this thesis namely, Gabor filters. Earlier studies on simple cells in the visual cortex of the brain suggest their involvement with visual perception of static and moving images and also for pattern recognition (Hubel and Wiesel, 1995; Hubel and Wiesel, 1968). It has been argued that the best biologically plausible computational model to describe the receptive field of the simple cells is Gabor filters (Daugman, 1985).

A set of high dimensional face image can be projected to a lower dimension which may be its *true dimension* or the *intrinsic dimension*. This may enable the removal of redundancies and noise in the dataset. The *intrinsic dimension* is usually very low and defines the minimum dimensions that can be used to define the dataset without much information loss.

The pre-processing with Gabor filtering for feature extraction is followed with dimensionality reduction methods: Principal Component Analysis (PCA), Curvilinear Component Analysis (CCA) and Fisher Linear Discriminant Analysis (FLD). Classification methods such as Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) are also discussed in Chapter 3.

Different computational models that differ in the pre-processing techniques are investigated here. Chapter 4 and Chapter 5 discuss the experiments performed with two different datasets and critically evaluate the results. The hypothesis that non linear facial features (Jarudi and Sinha, 2003) may be better extracted by non linear Gabor filters (Shen and Bai, 2006) is inquired. Also, the view that non linear CCA could be more effective in reducing dimensionality than by the linear PCA technique is investigated.

A comparison of the performance in the classification of expressions by human subjects and computational models provide interesting similarities between the two, as described in Chapter 6.

1.2 Contribution

This thesis contains a comparison of the performance of computational models with human subjects in classification of basic prototypical facial expressions. A biologically plausible pre-processing system followed with dimensionality reduction techniques and classification constitutes the computational model. The major novel contributions are:

- Face images are of very high dimension and dimensionality reduction methods such as Principal Component Analysis (PCA), Curvilinear Component Analysis (CCA) and Linear Discriminant Analysis (LDA) are used to reduce the dimensions. For CCA, the actual dimension to which the dataset is reduced is the intrinsic dimension. The facial features: the eyes, nose and eye brows are aligned at different angles and orientations. The pre-processing is performed by Gabor filters in extracting these features. Gabor filtering can be used in combination with PCA reducing the dataset to a mere 22 components, whilst maintaining 95% of the variance in the original data. A non linear dimensionality reduction method namely, CCA in combination with Gabor filtering can reduce the dimension of the dataset to as low as 5 components. The classification accuracies obtained from SVM and LDA following these pre-processing and dimensionality reduction methods were compared. For some expressions the massively reduced dimensionality datum still gave good classification results. For example the expression surprise with Gabor pre-processing and a CCA projection gives 84.09%.

- A detailed PCA analysis of facial expressions was performed. The results show that differing eigenfaces discriminate different expression. However some faces discriminate more than one expression and this may be related to the confusion in recognizing some expression by human subjects, but this is highly speculative.

- Different regions of the face are associated with different expressions. Earlier research has also studied the facial muscles associated with an expression using the Facial Action Coding System (FACS). This describes the changing facial features in the event of an expression (Ekman and Friesen, 1976; Hager, 2006). Another approach to identifying areas of the face that are important in expression of emotion is to use an *'Effect size'* analysis. Surprisingly, I have not found any evidence that this has been done elsewhere. My results as described in Chapter 5 indicate the areas of the face that discriminate each of the six prototypical expressions from a neutral face. Some of the results were predictable and some were surprising.

- A comparison between the performance of human participants and of computational models, in facial expression classification was performed and the results are discussed in

Chapter 6. There seems to be some similarities in the average response time and the classification accuracy between the computational models and humans.

1.3 Terminology

In the field of computer vision, the words facial *expression* and *emotion* are used interchangeably; however, this is not the case in Psychology. This is because human emotions are not just expressed by changes in the facial features; emotions can also be displayed by changes in voice, body language, and gaze direction. In computer models facial expression recognition takes into consideration only the visual information.

Six different computational models have been tested. These models differ in the pre-processing techniques used. The terms used to describe them are:

RAW – This computational model uses face images without any pre-processing.

RAWPCA – This computational model uses face images without any features extracted but reduced in dimensionality by PCA. The number of principal components used is always precisioned to retain 95% of the variance of the original dataset.

RAWCCA – This computational model uses face images without any feature extraction but reduced in dimensionality by CCA. For the most part the number of dimensions was that which was indicated by an estimate of intrinsic dimensionality, and discussed in Section 3.3.3 of Chapter 3.

GAB – This computational model uses face images with features extracted by Gabor filters but do not use any dimensionality reduction methods.

GABPCA – This computational model uses face images with features extracted by Gabor filters and dimensionality reduction by PCA. As before 95% of the variance is maintained.

GABCCA – This computational model uses face images with feature extracted by Gabor filters and dimensionality reduction by CCA. As before normally intrinsic dimensionality is used as the indicator of the number of required dimensions.

1.4 Structure of the Thesis

This chapter has discussed the factors that motivated this PhD work. The main contributions made by this thesis in the field of facial expression recognition are also discussed. Chapter 2 presents background literature for the psychological, experimental work reported in the thesis. It also reviews computational models and databases existing to date.

The computational models that are used for pre-processing and dimensionality reduction are discussed in detail in Chapter 3. It presents the background for the use of Gabor filters for pre-processing, to enable feature extraction of a given face image. Face images are of very high dimensionality. A detailed discussion on dimensionality reduction methods namely, PCA, CCA and FLD follows. It also presents an evaluation of the classification by SVM and LDA. This chapter also discusses the *Effect size* and the *Encoding face*. This chapter also investigates the significance of PCA components for different expressions.

All the computational models that have been discussed in Chapter 3 are analysed in Chapter 4 with a small set of face images from the FERET dataset, with only two expressions, smiling and neutral. The six different computational models are tested and evaluated in their ability to classify the two facial expressions.

In Chapter 5, these experiments are extended to all six prototypical expressions and to a larger set of face images from the BINGHAMTON BU-3DFE dataset. In addition to trying the six models with classification by SVM, classification accuracy is compared to FLD. The '*Effect size*' for all expressions is implemented and it gives very interesting results that describe the areas of the face associated with different expressions. A detailed analysis of the PCA demonstrates how the significant components can be used to morph the expressions. The classification accuracies with different expressions and the models is discussed and critically analysed with similar computational models in the literature.

A comparison of the performance of human subjects in facial expression classification with computational models is made and statistically analyzed in Chapter 6. For the human subjects, the data recorded were response time and classification accuracy. These results are compared with computational models and critically evaluated with reference to relevant literature. Interesting comparisons between different expressions and between computational models and human performance are reported. The conclusions from individual chapters are presented in Chapter 7. The main contributions of this thesis are also presented. A section on future work suggests some possible extensions of this work based on the findings and observations made.

Some of the results and discussions of the methods are presented in Appendices at the end of this thesis. In Appendix A the steps to perform PCA is discussed along with the steps to reconstruct the original face images from the PCA components. The plots of the PCA components for each

expression are included in Appendix B. The LDA along with PCA is used as a Euclidean distance classifier and the cross validation results are in Appendix C. The cross validation results of the SVM classifier are also presented in Appendix C. Appendix D has the results of the Bivariate correlation analysis for the misclassifications by human subjects.

Some of the work from this thesis has been published as Conference papers, Poster abstracts and a copy of these are included in Appendix E.

CHAPTER TWO

Literature Review

2.1 Introduction

The human face is a portrait of various facial features with the potential to communicate nonverbally with others. Over the years, the ability to recognize and respond to facial expression has been the focus of research in social psychology. Much of that research has been conducted on various aspects of facial expression, such as establishing when infants learn to recognize facial expressions and investigating the role of the right hemisphere in facial expression recognition. These are just a few of the questions that have been addressed. Although over the last two decades interesting research has been undertaken in answering some of them, it has been argued that little progress has been made (Hager, 2006). This chapter discusses some of the work in the psychology of facial expression, including neuropsychology, and in computational modelling of facial expression as a background to the new empirical work reported in this thesis.

Since the focus of this thesis is on the recognition of facial expressions and not on face identity, psychological and computational models of face recognition will not be reviewed with the exception of the Bruce and Young (1986) face recognition model which does refer also to expression recognition. The review will include the universality of facial expressions and the importance of the facial expression recognition; the distinction between expression generation and recognition; the distinction between categorical and continuous perception of facial expressions and the debate between feature based and holistic based facial expression processing. The importance of facial expression recognition is exemplified with case studies of impairments and the relevant neuropsychological research is discussed. Selective impairments of some facial expression recognition due to brain injuries and disease are also considered. Feature based expression classifier such as the Facial Action Coding System (FACS) and emotion based classifiers are described. A section on databases reviews important aspects of an ideal database and methodologies used; and the dataset that has been used in the current work is also mentioned.

2.2 The psychology of facial expression

Bell (1844) seems to have published the first objective and scientific study of facial expression. Besides presenting valuable diagrams of the muscles of the face, Bell pointed out that in all the

positive emotions the eyebrows, the eyelids, the nostrils and the angles of the mouth are raised, while in the negative passions the reverse is true. Jenness (1932) reviewed previous work on the study of facial expressions. Various researchers were by then performing experiments with various types of expressions. Some of the work included questions of innateness and started to investigate whether any particular facial expressions are easier to recognize compared to others. Studies that were undertaken involved classification of facial expressions in images. Langfeld (1918) found laughter was easy to detect followed by amazement, bodily pain, hate, fear, disgust, doubt and the least easily detected was angry. This study was followed by Aluport (1924) repeating the same experiment but with a larger number of human subjects and found laughter the easiest to detect followed by bodily pain, fear, distrust, amazement, anger, doubt, and disgust. However, Jenness (1932) used the same data but with a very large number of subjects in comparison to others and found amazement to be detected most easily followed by laughter, bodily pain, anger, distrust, disgust, fear and doubt. In his review, Jenness mentions that due to inconsistencies in the experiments performed, it seemed difficult to arrive at a consensus. However, he predicted that it was the beginnings in the field of facial expression recognition and pointed to the necessity for new and better techniques of research and for more thorough consideration of the questions and difficulties involved.

Darwin (1872) argued that the emotional expressions are universal and the same for all people based on his theory of evolution. However, the theory that emotional expressions were universal was ignored and rejected by many at that time. The idea that facial expressions are not valid indicators of emotion was widely accepted even though the evidence was contradictory (Bruner and Tagiuri, 1954). In the mid fifties, Ekman started his study on facial expressions. He was to become a key figure in this field. He has researched extensively for over four decades in topics related or relevant to emotion and facial expressions. The theory proposed by Darwin about the emotional expressions being universal that was rejected by other researchers was once again addressed by Ekman and Friesen (1971) who suggested, based on evidence, that expressions are indeed universal. A very recent study by Matsumoto and Willingham (2009) compared the expressions of blind and non-blind individuals and their findings provide further evidence that the production of spontaneous facial expressions of emotion is not learned. They conclude that something genetically wired is responsible for the generation of facial expressions of emotions. Evidence by Ekman (1973) proving universality of facial expressions was given by their study spanning cultures across the globe that suggested constants across cultures in the emotional meanings of facial expressions. Ekman has since then proposed the existence of six basic prototypical facial expressions that are universal. Expressions found to be universal in nature are: anger, disgust, happiness, sadness, surprise and fear. Findings about the expression of contempt are less clear, although preliminary evidence support it as being universal (Ekman, 1986). Izard (1977) reported that ‘interest’ and ‘shame’ facial expressions are also universal. Since then there have been many other studies around the world that validate the universality of some of these facial expressions (Matsumoto, 2001). Also, the facial expression in response to the emotion felt are produced by all people all around the world and from all walks of life

(Matsumoto *et al.*, 2007) although some reviews report evidence that is suggestive of some Asian subjects having difficulty in displaying some expressions such a disgust and fear (Pantic and Rothkrantz, 2000).

Ekman and Oster (1979) learnt that in addition to the other expressions mentioned earlier, distress and disgust expressions are also present from birth. Social smiles may emerge in an infant, just 4 weeks old. The face of 3-week-old infants can show 'interest' (Oster, 1978) . Anger and contempt may be seen by 6-months (Izard, 1978). Meaningful surprise and fear configurations are seen in the second year of life (Ekman and Oster, 1979). The facial expressions are registered by changes in the forehead, eyebrows, eyelids, cheeks, nose, lips, and chin. Most often, in real life situations, there is a complex combination of facial expressions such as pleasant surprise (happy-surprise).

Though many facial expressions are universal in nature, the way these are displayed depends upon culture and the upbringing. People learn to manipulate expressions in a number of ways for example by amplifying (showing more than actually felt), reducing the intensity than actually felt, showing a combination of more than one expression, concealing the emotion, or show a neutral face or even simulating some expressions when nothing is felt (Matsumoto *et al.*, 2007; Matsumoto, 2007). There is also evidence that displaying expressions on the face can even affect the way you feel. This is called the facial feedback hypothesis. Strack, Martin and Stepper (1988) performed experiments to show that generating facial movement that shows a smile can positively affect the way we feel.

2.2.1 Facial expressions and its representation

Facial Expressions are a display of one or more emotions of an individual across the face. It may indicate the psychological state of the individual to the observers. Facial expressions can be thought of as mode of communicating the feeling or inner emotional state (Lisetti and Schiano, 2000). Humans can adopt a facial expression as a voluntary action. However, because expressions are closely reflective of emotion, they are more often involuntary in nature (Matsumoto *et al.*, 2007). Although we usually (not always) have control of our emotional expressions, when voluntarily expressing them, we may not be best at it. Among other things, the timing (onset and offset) and the coordination of the various regions of the face (brows, eyes, mouth) are usually conspicuously "off" in posed expressions (Ekman and Friesen, 1975). Similarly, we frequently have difficulty in voluntarily inhibiting genuine expressions. Facial expressions are not just emotional responses but a form of social communication. Fridlund (1994) strongly disagrees with Ekman in his writings, arguing that expressions carry no inherent meaning but the two basically agree that facial expressions tend to forecast people's future

actions. However, instead of describing expressions from the point of view of the expresser, as Ekman tends to do, Fridlund thinks more in terms of people who perceive the expressions.

2.2.2 Facial Identity and Expression

It is over 20 years since Bruce and Young (1986) presented the most influential model for face recognition. They proposed parallel pathways for recognizing facial identity and facial expressions and lip speech. A similar neuropsychological model is proposed by Haxby, Hoffman and Gobbini (2000). Figure 2.1 shows the functional model for face processing proposed by Bruce and Young. Haxby, Hoffman and Gobbini presented a neural model of face perception that has 'core' and 'extended' systems. The core system differentiates mechanisms for coding changeable facial properties and mechanisms coding invariant facial properties. The extended system includes neural regions that are involved in semantics, language, emotion and attention, which support the recognition of different facial characteristics. The Bruce and Young model is compatible with the neuropsychological model proposed by Haxby, Hoffman and Gobbini.

Most of the facial features such as the eyes and mouth in particular convey information about what the person is feeling and enables communication (Ellis, 1975). The relationship between the various facial features is referred to as configural information. This is an important factor for facial identity and facial expression. Young, Hellawell and Hay (1987) performed experiments with composite faces (creating a new face by using different upper and lower half of face images of popular celebrities). They demonstrated that the importance of configural information in perceiving of facial identity and those configurations are only properly perceived with upright faces. Calder and Young (2000) studied the configural information in the perception of facial expressions in similar way as Young, Hellawell and Hay studied facial identity by using composites of facial expressions. The facial expression in an aligned composite face took time in comparison to identifying the expression in misaligned face. This explains the composite effect of facial expressions and parallels the composite effect with facial identity by Young, Hellawell and Hay. In addition, Calder and Young also had evidence that composite effects of identity and expressions operate independently of one another. This supports the pathway explained by the Bruce and Young model.

The model by Bruce and Young that is compatible with the model by Haxby, Hoffman and Gobbini suggests that the facial identity and facial expression recognition pathways separate very early on, immediately after structural and visual analysis of faces. Some cases of prosopagnosia that have no impaired facial expression recognition but with difficulty in recognizing identity would support the independence of identity processing; however, these cannot necessarily be thought of to happen solely (or even at all) at the visuoperceptual level. Other causes such as cognitive impairments, amnesia etc cannot be ignored for such impairments. The Bruce and

Young model has been investigated recently by Calder and Young (2005) and they agree that there is some separation between the coding of facial identity and expression; however, the dominant view of distinct pathways is not strongly supported as they question at what stage the facial identity route actually bifurcates from the facial expression route. Although of interest, this question of the stage of separation does not, however, impact on this thesis since only facial expression recognition is under consideration here.

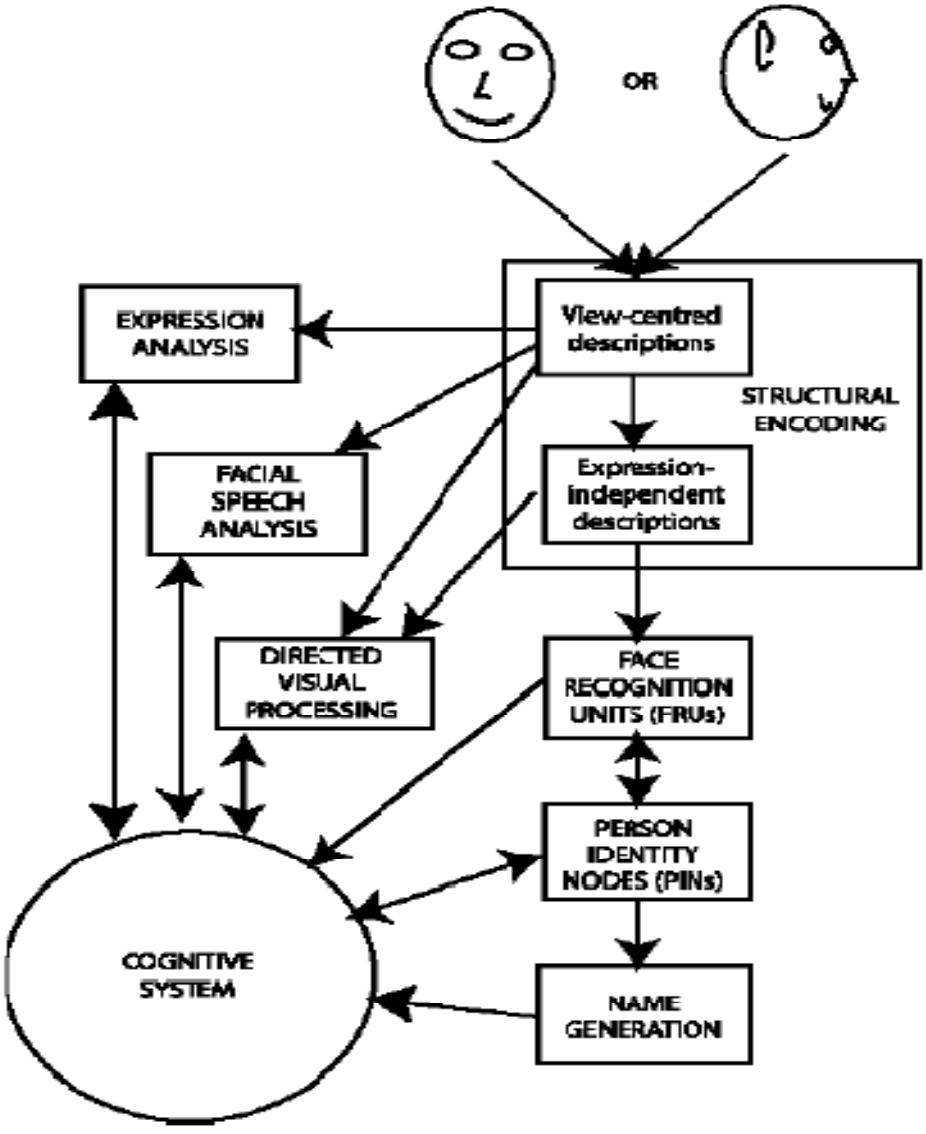


Figure 2.1: Bruce and Young's functional model for face processing

2.2.3 Facial Expression Recognition

Recognizing and understanding the facial expressions of other people is very important. Humans (and other primates) are biologically prepared for expression recognition, especially for the recognition of anger or threat (Ohman, 1993). The capability of a person to recognize facial expressions changes over time. Human infants as young as 5-6 months can discriminate between facial expressions of fear, anger, and sadness and that angry faces may be particularly 'attention-grabbing' for infants (Schwartz *et al.*, 1985; Serrano *et al.*, 1992). Some findings suggest that negative expressions (such as anger and fear) have greater impact on the perceiver that the positive ones (such as happy). For example, the so-called face-in-the crowd effect suggests that angry faces are detected faster than happy faces when they are presented alongside other faces (Fox *et al.*, 2000; Hansen and Hansen, 1988). Hansen and Hansen concluded that facial displays of threat (from angry faces) were detected automatically and faster and that the consequence of this would be to shift the attention of the person to it. This would presumably provide an evolutionary advantage. Happy faces were detected after a serial and linear search. However, other studies have shown that happy expression recognition is faster and easiest to be recognized and suggest that it could be attributed to the higher prevalence of this expression in social circumstances (Carvajal *et al.*, 2004; Kirita and Endo, 1995). Recent studies by Shimamura, Ross and Bennett (2006) suggest that memory for happy expression is longer than other expressions that were tested (surprise, angry and fear). This was also true when faces were turned upside down.

Facial expressions are an essential part of social cognition and convey information about the person's internal emotional state (Calder, 2005). The importance of facial expression recognition can be illustrated with individuals who have difficulty in perceiving it. In patients with brain damage or disease, the emotion recognition can be impaired. Adolphs, Travel, Damasio and Damasio (1994) found that bilateral amygdale damage results in harder fear expression recognition. Patients suffering from Alzheimer's disease, have impaired facial emotion processing and selective impairment in labelling facial expression of sadness (Hargrave *et al.*, 2002). Patients of Parkinson's disease have shown to have selectively impaired recognition of facial expressions of disgust (Suzuki *et al.*, 2006). Schizophrenia sufferers have been shown to exhibit difficulty in recognizing the emotion that corresponds to a given facial expression; specific deficits in recognizing happy faces have been documented as has the evidence that these patients were more inclined to attribute any facial emotion as fearful or sad (Tsoi *et al.*, 2008).

Two major theories explain how facial expressions are perceived and processed: the categorical view and the continuous view. The categorical or the discrete category view refers to specific emotions such as anger, happiness, surprise, fear, disgust, and sadness. Conversely, the dimensional theory or the continuous view suggests that the mental representation of emotional

space consists of continuous underlying dimensions in which similar emotions are clustered together while different ones are far apart (Chan, 2009).

Etcoff and Magee (1992) were the first to study the categorical perception of facial expressions. They experimented with the six basic prototypical expressions proposed by Ekman and Friesen (1976) but used computer generated drawings of these expressions. They found that faces within a category (such as two smiling faces from the happy category) were discriminated more poorly than faces in different categories (such as discriminating a happy face from fear face) that differed by an equal physical amount. They found that all expressions except surprise were categorically perceived. Thus they concluded that emotional face expressions are perceived categorically and posed a significant challenge to idea of continuous space of emotions. This means rather than being perceived as a linear progression, the continuum of expression is perceived as an abrupt discontinuity at the boundary between two categories, for example from happy to sad. They also suggested that people always seem to see faces exhibiting one or the other expression. This led to more research into this field and Calder, Young, Perrett, Etcoff and Rowland (1996) repeated these experiments with photographic-quality stimuli. The evidence from these experiments compliments the idea of categorical perception proposed by Etcoff and Magee. However, they do not agree fully with the idea of the mandatory assignment of an emotion category to the face. They also propose that categorical perception effects are evident when the population cells of the neural systems become more tuned to various expressions.

Other researchers however, do not agree with the theory of categorical perception for basic expressions and emotions. The idea of categorical perception for facial expression is challenged by results that show that similarity judgments of these expressions exhibit a graded, continuous structure (Dailey, 2002). Russell (1980) proposed the circumplex model for facial affect and later proposed that facial expression behave as fuzzy sets (Russell and Bullock, 1986). This research followed with other studies support that facial expression perception is a continuous, multidimensional and that some expression categories are more similar to each other than others (Dailey, 2002; Katsikitis, 1997; Russell *et al.*, 1989; Schiano *et al.*, 2004). When Young, Rowland, Calder, Etcoff, Seth and Perett (1997) experimented to find evidence supporting categorical or continuous facial expression perception, they found evidence supporting both. To date in spite of years of research on facial expression recognition by humans and automatic facial expression recognition systems, there has been no evidence that simultaneously explains all of these seemingly contradictory findings (Dailey, 2002).

2.2.4 Facial features and expressions

In the literature, facial features are described as either internal or external. It is normally assumed that the internal features such as eyes, nose, mouth and eyebrows and the configural relationship between them are important when compared to external features such as hair and

jaw line which are too variable to be useful for practical purposes (Sinha *et al.*, 2006). Featural processing involves using the individual features for processing and configural processing involves the relationship of various internal features. The holistic feature processing involves the interdependency between featural and configural information. Configural processing is already known to be important for face recognition, however, further experiments have now found that configural information is also necessary for facial expressions (Calder *et al.*, 2000). This does not mean that the individual features of facial expressions are not just encoded for identification but it implies that the configural relationship of the features plays an important role in the encoding of facial expression.

Different facial areas of the face are involved with different expressions. Bassili (1979) suggested that facial expressions are locally processed by brain unlike face recognition which is processed holistically. His investigation showed that the upper part of the face is important for some expressions and for other expressions, the lower part of the face is important. Zhang and Cottrell (2005) suggest that local features are good predictors in facial expression recognition and holistic processing is useful for facial identity recognition. An experiment by Kirkpatrick, Bell, Johnson, Perkins and Sullivan (1996) that had children detect facial expressions from the upper and lower half of the face suggested that the children concentrated on the features in the lower half of the face for expressions of happiness, sadness, surprise, and disgust. The features on the upper half of the face such as the eyebrows were used for the faces expressing anger and fear. The results of this study are consistent with the idea that certain groups of facial features are associated with specific emotions.

Expressions can be classified as macro expressions and micro expressions. Some expressions are so brief that they hardly last for a fraction of 2-3 seconds and they are called micro expressions. These micro expressions are usually revealing genuine emotions which the person tries to conceal and are not easily detected (Ekman, 2003). The macro expressions are the ones which last for a longer time than the micro expressions. However, even this does not last over 5-6 seconds. So, if this expression lasts on an individual's face, it indicates that the feeling was that intense which would also be displayed not just with the face but by the change in the voice tone or by words. Hence it is very hard to miss these emotions even if you are not looking at the persons face. The very long lasting facial expression however does indicate that they are not genuine and is faked or a mock expression (Ekman and Friesen, 1975).

To summarize, different features of the face are involved with various expressions and expression recognition involves featural and configural processing. Purely holistic based processing does not seem to be very useful for facial expression recognition (Schwaninger *et al.*, 2006).

2.3 Neuropsychology aspect of facial expression recognition

The recent neurological model for face perception that was proposed by Haxby, Hoffman and Gobbinni (2000) is compatible with the psychological model offered by Bruce and Young (1987). The core system of Haxby, Hoffman and Gobbinni's model contains two functionally and neurologically different pathways for the visual analysis of faces: one identifies those facial properties that change (such as expression, lip speech and eye gaze) and it involves the inferior occipital gyri and superior temporal sulcus (STS) of the brain, whereas the other identifies constant facial property (such as identity) and involves the inferior occipital gyri and lateral fusiform gyrus. The model proposed by Haxby, Hoffman and Gobbini and the model proposed by Bruce and Young agree that there are different pathways for the visual analysis of facial identity and expression. However, they differ in terms of how the processing takes place i.e. if there is a separate system to code the facial expressions perceived or if the processing takes place along with detection of other changing facial features.

Reviewing the broad subject of Neuroscience is beyond the scope of this thesis and it also requires in depth knowledge of various anatomical structures in the brain and its physiology. However, case studies of people with various anatomical lesions due to surgery and brain injury or damage and how it can effect on the ability to detect facial expression is discussed in the following section.

2.3.1 Human Brain injuries /Lesions and their effect

The left and right halves of the brain are specialized for different tasks. The right hemisphere of the brain controls the muscles of the left half of the body and vice versa. The left hemisphere of the brain performs tasks involving language and logic. The right half of the brain is involved with spatial abilities, face recognition, cognition, and visualization (Gisalason, 2007). Hence, any damage due to surgery or injury to the right hemisphere may result in impaired face and expression recognition. Though in the recent decades a lot of research has been done to study the cognition and behavioural impact of these injuries, less research has been done in the field of impairments of facial expression recognition.

Research with primates has shown that the temporal visual cortex is involved in processing facial expression. In addition, neuro-imaging studies of healthy normal people have shown areas of the brain involved in the processing of facial affect. Crocker and McDonald (2005) studied the effects of traumatic brain injury on facial expression recognition. They conducted experiments based on which they suggest that there is some impairment associated with recognizing facial expressions after brain injury and it was more with expressions pertaining to negative emotions.

Crocker and McDonald showed that the subjects of their study with traumatic brain injury were relatively normal on face recognition but abnormally poor when recognizing expression. This supports the notion that there are two distinct pathways for emotion and identity. In addition, these patients had an inability in naming an expression. This also suggests that to some extent there could be separate cognitive processes within the emotion recognition system. All these studies support the models proposed by Bruce and Young and Haxby et al, which have been discussed earlier. In a study by Buck and Duffy (1980), they learnt that people with right brain hemisphere damage showed more emotional deficit as compared to those with left brain hemisphere damage. Further studies by others have also shown that people with right hemisphere damage have difficulty in exhibiting emotion expression in comparison to a neutral one (Browndyke, 2002).

There is evidence from experiments by Ley and Bryden (1979) that when normal subjects were shown strong emotional expressions, the right hemisphere of the brain was highly active when compared to the left and also in comparison to neutral or weak expressions. Similarly, when a person displays a genuine expression, the intensity of the expression on the left side of the face is more than on the right (Browndyke, 2002). This very well gels with fact that the movements in the left half of the body are controlled by the right hemisphere of the brain.

The results of experiments by Adolphs, Damasio, Tranel and Damasio (1996) suggest that all patients with brain lesions or damage recognized happiness but there were significant impairments in recognizing negative emotional expressions when compared to control subjects. The patients with these impairments were significantly more likely to have damage to their right hemisphere of the brain, the visual and somatosensory cortical sectors in particular. Patients with brain injury on the left hemisphere showed normal recognition. The suggestion by earlier researchers that only the right hemisphere is involved in emotion recognition conflicts with study by Sprengelmeyer, Rausch, Eysel and Przuntek (1998) who suggested that the left hemisphere is important. He performed fMRI studies on people when they judged expression (anger, disgust and fear) and concluded that different neural structures were involved with each of these expressions. He also found that though the recognition of these expressions is based on different systems, they converge at the left frontal cortex which seems to conflict with earlier studies of right hemisphere involvement.

With these conflicting results in mind, Adolphs, Damasio, Tranel, Cooper and Damasio (2000) experimented with patients who had right or left brain hemisphere lesions and the task was emotion recognition. Though the results of these studies do not rule out the left hemisphere involvement in emotion recognition, it does show that there is very little association. It also shows that as Sprengelmeyer suggested, the frontal cortex is involved in emotion recognition; however, it may be making more meaning to the expression perceived (example- language) rather than in actually perceiving the expression on the face. A number of studies have covered this subject over the years, but no solution has yet been obtained that resolve this argument. Though the right hemisphere is still thought to be significantly involved in emotion recognition,

there is an ongoing debate on whether the right hemisphere is involved with all expressions/emotions and also, whether the right hemisphere is involved with negative emotions while left hemisphere is involved with positive emotions.

So far the discussions in this section on neuropsychology have dealt with expressions in general. Recent research has shown that various parts of the brain are involved with different expressions. Fox, Lester, Russo, Bowles, Pichler and Dutton (2000) studied threat detection, which is normally necessary in challenging social circumstances. The amygdala has believed to be engaged while processing specific expression such as fear; however, recently it has been found to have some role with perception of other negative emotions such as anger, sadness, disgust (Adolphs, 2002). Further evidence suggests the greater role of amygdala in recognizing signal of potential threat or danger.

2.3.2 Can the psychological and neuropsychological bases of facial expression recognition be used to develop computational models

Taking into account some of the studies in expression recognition, the next step would be in computational modelling of this system which will help us to understand the underlying mechanism involving expression recognition.

The primary visual cortex is located in the posterior part in the occipital area of the brain. It has been very widely studied with relevance for visual perception of static and moving images and also for pattern recognition. The primary visual cortex is the part of the brain that receives visual input from the retina. The primary visual cortex is divided into six functionally different layers labelled V1 to V6. The V1 part of the visual cortex is the first site where strong orientation and direction selectivity are observed (Hubel and Wiesel, 1968). Receptive fields of cells in the V1 layer of the visual cortex belong to one of the two categories: simple or complex. The Simple cells have smaller receptive fields that are elongated, with an excitatory central oval, and an inhibitory surrounding region. These cells are excited when the images for these receptive fields have a particular orientation and have low spontaneous activity. Some parts of the receptive fields of the simple cell respond to the onset of stimulus while other parts respond to the offset. The receptive fields of the complex cell are larger than that for simple cells and excite the cell as a response to movement in a particular direction. They exhibit greater spontaneous activity. The receptive fields of the complex cells respond to both onset and offset of the stimulus (Hubel and Wiesel, 1995; Leloglu, 1994).

The features of the face which are at various orientations and angles such as the eyes, eye brows etc can be extracted by computational models which mimic the simple cells of this visual cortex (Daugman, 1985). The receptive fields of simple cells can hence be well described by Gabor

filters (Marcelja, 1980; Daugman, 1980) which are limited by both space and frequency. There is evidence that simple cells found in pairs are tuned to same orientation and frequency with phase difference of approximately 90 degrees (Pollen and Ronner, 1981) and may represent the real and imaginary parts of a complex Gabor filter. Hence, the nearest biologically plausible feature extraction method mimicking simple cells would be Gabor filters and is explained in detail in Section 3.2 of Chapter 3. My work uses Gabor filters for feature extraction for the computational models of facial expression recognition.

2.4 Computational models of facial expression recognition

This section discusses research over the last decade in developing computational models for facial expression recognition. Whilst there has been a considerable amount of research done on facial identity recognition, they have concentrated on issues dealing with the identification of face by name, categorization of face by gender, race and age (Buchala *et al.*, 2004c; O'Toole *et al.*, 1994; Calder *et al.*, 2001). Some approaches in studying facial expressions such that it can aid in recognizing the facial identity, gender, age and race as in real life situations the facial expression are unpredictable, multiple and always present (Lisetti and Schiano, 2000). A considerable number of systems have been developed which deal with the issue of facial expression analysis. Padgett et al (Dailey, 2002; Dailey *et al.*, 2000; Padgett *et al.*, 1996; Padgett and Cottrell, 1998) were the first to develop computational models for facial expressions. Every model is different in the technical approaches used (Lisetti and Schiano, 2000). Lyons et al (1998) used Gabor filters for facial feature extraction in experiments with facial expression. He suggested that Gabor representation shows a significant degree of psychological plausibility.

Another computational model for expression and recognition was proposed by Calder et al; it used the idea of encoding the positions of various features of the face with respect to the average face (Calder *et al.*, 2001). Before performing PCA, the faces are said to be *warped*. On other words, the facial features in a given set of faces are morphed to the average face of the set to obtain the same standard positions for the features of all faces. PCA is performed to obtain a low dimensional representation of the face shape and texture. Their experiments have shown that PCA can code facial expressions and that PCA can code facial expressions in psychologically plausible form.

Most facial expression processing systems use part based or feature based processing for expression recognition. The best example is the use of FACS.

The Facial Action Coding System (or FACS in short) is a widely used method for describing the various internal facial feature behaviours. FACS allows psychologists to code expressions from static facial images. Ekman and Friesen (1978) developed the FACS by studying which muscles

on the face undergo changes in a particular expression. The unit of measurement in FACS is Action Unit or AU. An example of what an AU constitutes can be seen in Table 2.1. The contractions and relaxations of the muscles result in changes in the appearance of the face whilst displaying facial expression. The purpose of designing this system was to best discriminate one expression from another. This has been used by skilled human coders to determine the category into which the facial display fits into.

Encoding a facial expression in FACS produces a list of AU's. Normally, every AU records changes with more than one muscle. An expression can be coded as a combination of more than one AU. A total of 44 facial action units have been defined. Experienced human coders use FACS to manually code any facial expression and decompose it into its specific AU's. This has been one of the highly used efficient methods for expression recognition. A few examples of AU are shown in Table 2.1.

Table 2.1: Examples of Action Units (AU). The first column is the AU number, followed by the description for changes in the muscle, the third column describes the muscle involved and the final column shows an example for that AU.

Action Unit (AU)	**Description**	**Facial muscle**	**Example image**
1	Inner brow raiser	Frontalis, pars medialis	
15	Lip corner depressor	Depressor anguli oris (or Triangularis)	
26	Jaw drop	Masseter, relaxed Temporalis and internal Pterygoid	

FACS coding is performed by highly trained human coders and lately, some automatic computational modelling has been investigated by Cohn and Kanade (2000), Bartlett (2005) and Pantic (2006).

A majority of studies done so far have been based on the categorization of Ekman's prototypical expressions and the problems associated are: firstly, the six basic prototypical expressions are not defined with FACS or with any facial codes to be identified universally and are quite confusing. Secondly, two different expressions can have two or more features involved in a very similar manner such as smiling mouth and raised eyebrows for a pleasant surprise and happy expression (Pantic and Rothkrantz, 2000; Pantic and Bartlett, 2007). One important thing about FACS is that it is not a model for facial expression processing and does not claim to define which of the combinations of AU's represents any expression (Schwaninger *et al.*, 2006).

2.4.1 Facial expression recognition systems

Over the last few years a number of computational models have been developed that perform facial expression classification. Ideally, any facial expression recognition system designed should be capable of tasks comparable with the human visual system. The human visual system is believed to perceive the face as a whole and not as a collection of facial features (Pantic and Rothkrantz, 2000) and is capable of filling information in order to aid identification, if any part of the face image is occluded or covered. This is a very difficult task for any computational system to do.

In general any facial expression classification system would have the three basic units: Face detection, feature extraction and facial expression recognition.

2.4.1.1 Face detection

Determining the exact location of a face within a large background is a very tricky job for a computational system. An ideal face detection system should be capable of detecting faces within a noisy background and in complex scenes. Most often, there are variations in pose and lighting conditions, diverse range of sizes of the face, colour, texture and also movements across the face due to facial expressions and head movements (Fasel and Luettin, 2003).

The facial components such as the eyes, nose, eyebrows etc are the prominent features of the face. The face may be represented as a whole (*holistic face representation*) or as a set of these facial features (*analytic face representation*). It can also be represented as a combination of these and is called hybrid representation.

There has been much research in the field of face recognition over the last two decades (Susskind *et al.*, 2007; Essa and Pentland, 1997; Bartlett *et al.*, 1999; Fasel and Luettin, 2003) and the most commonly used face detector in automatic facial expression analysis is the one that is proposed by Viola and Jones (2004) which makes use of a cascade of filters, which are trained by Ada Boost.

2.4.1.2 Feature extraction

Once the face is detected, the next step is to extract the features that may be relevant for facial expression analysis. If the face is represented as a holistic face model then the *template based* feature extraction method may be used. If the face is represented as an analytical face model, then *feature based* extraction methods may be adopted. The most efficient of all are the hybrid methods which uses the analytic and holistic method for face representations (Pantic and Rothkrantz, 2000). The template based methods are also referred as *appearance based* feature extraction methods and feature based methods are also referred to as *geometric* feature extraction.

A review on these methods by Pantic lists a number of methods that have been used with feature extraction. Some of the holistic or template or appearance based methods used are: active appearance models (AAM) which makes use of PCA, labelled graph to fit on a face image by using elastic bunch graph technique and applying Gabor jets at these points and also, gradient optical flow method which estimates motion of specific points on the face. Some of the feature based methods include: multiple feature detectors applied on specific features of the face, extracting brightness distribution data on the face and optical flow method for specific areas such as the facial features on the face.

Gunduz et al describes feature extraction methods can be broadly classified into 4 categories (Gunduz and Krim, 2003):

- Geometric feature based – These methods extract the shape and locations of facial features such as the mouth, eyes, brows, and nose. They are presented as a feature vector that represents the face geometry.

- Template based – These methods match the facial components using an appropriate energy function. A simple example for template matching is that a test image represented as a two-dimensional array of intensity values is compared using a suitable metric such as the Euclidean distance with a single template representing the whole face.

Edwards et al (1998) used Active Appearance Model (AAM) for representing the shape and gray level property of the image. The images were hand-labelled at points that represent the key positions of facial features. PCA is applied to shape and gray level data separately (Turk and Pentland, 1991). PCA is applied again to this vector of concatenated shape and gray level parameters resulting in components describing 'appearance'. In order to perform face recognition, the appearance parameter minimizes the error between the new face image and the synthesized AAM image. Hence, these methods are also called as appearance based models. The other methods include Independent component analysis (ICA) and Gabor filters are used to extract wavelet feature vector for the facial components (Hong *et al.*, 1998). These are holistic and rely on the statistical technique and an unsupervised learning method. Linear discriminant analysis (LDA) (Belhumeur *et al.*, 1997) is another type of appearance based technique except that it is a supervised learning method.

- Colour segmentation based – Here, the skin colour is used to detect the face features. Any non skin colour on the face is viewed as a feature such as eyes, mouth, nostrils etc (Vezhnevets *et al.*, 2004).

Fasel (2003) suggests that there can be other approaches to feature extraction. They are deformation extraction and motion extraction, both can be implemented holistically or locally. Deformation based methods can be applied to both static images and captured frame of an image sequence. They rely on neutral face images to extract facial features associated with an expression efficiently so that permanent wrinkles and creases are not picked up as changes in facial features. In contrast, motion based methods directly focuses on the facial changes that occur due to facial expression.

Deformation methods are: *Image* based or *model* based. Motion extraction methods that focus on facial features relevant to facial actions are: *dense optical flow*, *feature point tracking* and *difference images*. The following are some of the methods that have been discussed in literature so far.

A. DEFORMATION METHODS:

- **Image based deformation methods :**

Holistic -

- Neural network based such as Multi layer perceptron, feed forward network (Dailey, 2002) and back propagation algorithm (Lisetti and Schiano, 2000).

- Gabor Wavelets (Fellenz *et al.*, 1999; Dailey, 2002).

Local –

- With windows placed across areas of interest such as the facial features, PCA and neural networks are used (Padgett and Cottrell, 1996).

- Local transient facial features such as wrinkles and creases which occur during an expression are measured by image density profiles or by determining the density of high gradient components over the areas of interest (Lien, 1998).

- **Model based deformation methods:**

Holistic –

- Active appearance models (Lanitis *et al.*, 1997; Edwards *et al.*, 1998).

- Labelled graphs use sparse distributed fiducial feature points with Gabor jets. Each Gabor jet is a filter response of a Gabor filter at that point on the face image (Hong *et al.*, 1998; Lyons, 1999). These points are placed at specific areas of the face image in order to perform better feature extraction.

Local -

- Geometric face method uses the relationship between the features such as mouth, eyes and nose (Kobayashi and Hara, 1997). The entire face is represented by 30 facial characteristic points (FCP) and in combination with neural networks, the measurements are made.

- A two view point based method adopted by Pantic and Rothkrantz (2000) represent frontal and side view of face as facial points at the facial features. Multiple feature detectors are applied to study the contours of the salient features such as eyebrows, eyes and mouth.

B. MODEL METHODS

- **Dense optical flow methods :**

  Holistic –

  - The methods used here analyze whole face motions with wavelets and multi resolution optical flow. Optical flow can define the relative changes in the brightness pattern of an image. The use of optical flow to track motion is much useful with facial expressions because facial features and skin naturally have a great deal of texture. Using PCA, a low-dimensional representation of the high dimensional dense flows for each frame can be used (Lien, 1998).

  Local –

  - The same techniques such as that used in holistic processing is adopted except that the areas of interest are restricted to specific regions of the face representing facial actions (Mase, 1991).

- **Motion models:**

  Holistic –

  - Changes in facial features in particular lips are tracked by creating force field around these areas by making use of the gradients found in images(Terzopoulos and Waters, 1993). Sophisticated 3D motion and muscle models for facial expression recognition have been used to track the changes (Essa and Pentland, 1997).

  Local –

  - These models allow local regions in space and time to accurately record non rigid facial motions and also motion associated with the edges of the mouth, nose, eyebrows and eyelids by a very small number of parameters (Black and Yacoob, 1997; Yacoob and Davis, 1994).

- **Feature point tracking :**

  Local –

  - Facial feature point is based on facial features in regions of brows, eyes, nose, and mouth. However, the forehead, cheek and chin regions also have important

expression information. Feature points are placed on the face in areas of high contrast especially at locations of intransient facial features which are always present on the face but may be deformed due to facial expressions. Motion analyses are performed by measuring these displacements and are tracked. Other studies have used different component models for facial features such as lips, cheeks, eyes and eyebrows. They use feature point tracking to study the deformation of these facial features (Lien, 1998; Tian *et al.*, 2005). Similarly, a rectangular area enclosing the feature can also be tracked with the help of feature points (Rosenblum *et al.*, 1996).

- **Difference images**

  Holistic –

  - Differences of image intensities can be obtained by subtracting a given face image from a previously stored neutral face image of the same subject. The results depend on the alignment of the faces in consideration (Fellenz *et al.*, 1999; Donato *et al.*, 1999).

  Local –

  - Region based difference image models belong to local methods.

However, most often extraction methods are one of the two categories: Holistic (appearance based) or feature based (geometric based).

The feature based methods extract the information from the facial deformation of the features during the display of an expression. They emphasize on the contours of the eyebrows, lips, corners of the mouth, eyes or the geometrical relationship between the features represented as a set of fiducial points on the face (Buenaposada *et al.*, 2008).

In comparison to reducing the image to a set of facial features which removes a lot of information as in feature based methods, holistic appearance based methods make use of the entire face as a whole. Over the years, though both methods have been used and the reviews do not support one over the other with mixed findings, the combination of both appearance based (holistic) and motion based (feature) may seem to be more powerful as some evidence support this (Bartlett *et al.*, 1999). The use of appearance based model for feature extraction is found to be good with expression recognition (Littlewort *et al.*, 2006). The recent trend is the use of hybrid systems which use both holistic and feature based (Schwaninger *et al.*, 2006). In a hybrid method, instead of using eigenfaces, PCA is applied only to specific facial areas that have facial features to obtain 'eigenfeatures'. These systems are capable of performing efficiently even in situations where there is severe changes in the appearance of a face due to occlusions (Swets and Weng, 1996). Similar other methods use SVM's which are trained to recognize

facial features. The combined configuration of these features can then be used by some high level classifier (Schwaninger *et al.*, 2006).

2.4.1.3 Facial expression classification

The final step in facial expression analysis is classification which classifies or identifies the expression. The classification task always ends up as one of the basic emotions or a facial action. In other words, the classifiers of facial expression are message based or sign judgement based. In the message based systems determines the underlying affect, the outputs of which will be judged as an emotion such as 'angry' , 'happy' etc. The sign judgement systems are based on detection of facial action units. For example, a brow furrow could be judged as 'angry' in a message based and as a movement of facial muscles with the sign judgement system. A higher level decision making process needs to be followed in the sign judgement systems to interpret these muscle movements.

Irrespective of the classification category used, the classifiers can either follow a template based or a neural network based or a rule based classification method (Pantic and Rothkrantz, 2000). Template based methods include discriminant functions such as LDA, PCA and spatio-temporal energy templates. Rule based methods make use of expert system rules. Back propagation learning methods are the most often used neural network based classification method.

Another way of classifying these methods as reviewed by Fasel (2003) suggests that classification can be achieved by one of the two approaches: spatio-temporal approach or spatial approaches.

Spatio-temporal approach: This approach emphasizes space and time. The image template refers to space and a sequence or few templates refer to time. The spatio-temporal approach includes Hidden Markov Models to model the dynamics of facial actions (Lien, 1998). A number of classifiers have been developed that use this approach. Another class includes 2D motion field, where instead of a sequence just two templates are used whose Euclidean distance will provide the estimate for the expression (Essa and Pentland, 1997).

Spatial approach: This involves the use of neural networks (Lisetti and Schiano, 2000; Padgett and Cottrell, 1996; Kobayashi and Hara, 1997). Neural networks can be applied to face images with or without undergoing feature extraction and representation by methods such as PCA, ICA and Gabor filters (Fellenz *et al.*, 1999; Dailey and Cottrell, 1999). Use of dimensionality reduction methods such as PCA, ICA, and CCA can also be performed which reduces the complexity of the classification task in terms of time for classification and also computational complexity. These methods can be used either holistic or locally. A number of classifiers in the past have used this approach.

Another way of classification by Lisetti and Schiano (2000): Image motion, Anatomical models, neural networks and hybrid systems. The Image motion approach analyzes the motion and extract dynamic muscle action between successive images or in a sequence and is called as Optical flow (Mase, 1991). An array of arrows is used to indicate the direction and the magnitude at each image location. Other methods use anatomical models of the face in order to interpret the expression (Essa and Pentland, 1997; Terzopoulos and Waters, 1993). The problem with this technique is that of producing the anatomical model, which is difficult considering the vast range of feature differences on the face across individuals. Neural network methods could be supervised or unsupervised learning networks. With face expression, they work with 2D images and receive pixel intensity of the image as the inputs. Support Vector Machines (SVM) can also be used for classification. This is a learning algorithm that separates two classes of data such that there is maximum separation between them. A number of studies with facial expression use SVM for classification (Vert, 2002; Cortes and Vapnik, 1995; Zheng *et al.*, 2004b; Liejun *et al.*, 2009).

The ideal facial expression system should be capable of identifying an expression irrespective of age, gender, ethnicity, and also with varying degrees of intensity of the expression. Also, the recent advances especially with recognition of facial expression in moving sequences suggest that the timing of these facial expressions is also a very important factor. Designing an ideal robust facial expression system that is capable of detecting all expressions in various lighting conditions, pose, gaze, even in the presence of facial hair, glasses, different hair style, and also capable to fill in the gaps in the areas of the faces that are obstructed or occluded that will match a good human expression expert is a very difficult task.

This thesis does not discuss face recognition and concentrates solely on methodologies involving facial expressions. This thesis uses a holistic appearance based method namely; Gabor filters for feature extraction followed by dimensionality reduction methods such as PCA, CCA, LDA and classifiers namely SVM and FLD.

2.5 Databases

There are number of databases which are frequently used with experiments on facial expression classification. Some earlier studies are intended to judge human performance and have used sketches (Etcoff and Magee, 1992; Jenness, 1932). All other works use either static images and more recently, moving image sequences of posing individuals. Each of these methods has their own benefits and drawbacks. It is impossible to make a one to one comparison with the results of different computational models of facial expression classification. The primary reason being, none of them use the same database. There are number of issues that have been raised by

researchers earlier in reference to databases. A number of researchers have discussed the factors that affect quality of the database that make comparisons from these experiment results difficult (Zheng *et al.*, 2009; Lisetti and Schiano, 2000; Pantic and Bartlett, 2007; Fasel and Luettin, 2003; Buenaposada *et al.*, 2008; Pantic and Rothkrantz, 2000).

They include:

- The intensity of the expression on the face of the subject.
- Are the images from spontaneous expression or posed for the camera by subjects?
- Presence of noise - is the recording performed in a laboratory or in real life situations.
- Is the expression on the face significant or is it the internal feeling- both need not be the same.
- Is the subject aware of being recorded?
- With image sequences, the timing of the facial expression is important.
- Age of the person – preferably with not many permanent wrinkles which can contribute to variation in feature shape.
- Presence of facial hair or glasses.
- Ethnicity, Gender.
- Does the database have all six basic prototypical expressions?
- In real life situations, it cannot be guaranteed that the subject will not move.

Currently, a number of databases exist. I have used two types of datasets – FERET (Philips *et al.*, 1998) and BINGHAMTON BU-3DFE (Yin *et al.*, 2006).

2.6 Discussion

This chapter has discussed facial expression with respect to three different domains: psychological, neuropsychological and computational. The process of generating expressions is innate and evidence suggests universality of expressions across the globe. The ease with which facial expression are recognized by humans, the processes involved in the human brain, the importance of the ability to recognize, brain lesions and impairments associated with them have also been discussed. The process by which human beings perceive facial expressions and recognize them is complicated. Very early contribution in the field of recognition of facial expressions by humans has been discussed. The classification accuracy of facial expressions by humans is much higher in comparison to any computational models that have ever been developed so far. Most of the computational models work with static images which do not represent normal ecological environment and though recently work is being done on moving

video images, they are posed expression rather than spontaneous; none of these depict natural social circumstances that we normally deal within real life situations.

In addition, humans do not make judgements with six basic expressions in mind; it is much more than that. Micro expressions, macro expressions, deception, are some other factors that are involved in addition to the complex expressions which are combinations of more than one expression. Judgements in social circumstances in real life situations take into consideration other factors such as body language, voice, tone and also the environment around us. Facial expression is only one component of emotional display. Hence, an ideal computational system would be the one that takes into consideration each one of these small factors that have been mentioned.

Having discussed about these factors, with information from existing literature, it is very difficult to compare the results of various computational models of facial expression recognition. In addition, comparing the results of human performance in experiments of facial expression classification and results from computational models of facial expression recognition is also a daunting task. The major factors for these difficulties are differences in stimuli and methodologies used in these experiments. Few other factors include differences in the ability of brain lesion patients or people with various brain diseases, and also, gender and age of the participants in neuropsychology based experiments. Also, possible effects of other disorders such as autism, anxiety and depression should not be ignored as they can also affect the ability to perceive and judge emotions or expression and also, in exhibiting them. This chapter discussed the simple cells of the visual cortex of the brain and the next chapter discusses the biologically plausible Gabor filters that mimic the simple cells.

Facial expression recognition is a very interesting field of research and has brought together psychologists, psychiatrists, neurophysiologists and computer scientists. A better understanding between these fields would result in developing better, biologically plausible facial expression systems that are able to match the human classification performance.

CHAPTER THREE

Computational Techniques

3.1 Introduction

Human beings appear to detect and process faces and face expression with minimal effort. However, to develop a computational model capable of doing this is a non trivial task. The processes involved in developing such a computational model, and how best it may be developed to mimic a human like performance, will be explained in this chapter.

Computational techniques that are used with images include pre-processing techniques for feature extraction, dimensionality reduction and classification algorithms. This chapter explains the feature extraction method used here, namely, Gabor filters. This is followed by a discussion of dimensionality reduction methods. Face images are high dimensional in nature and though not many of face images are used in experiments, it presents a challenge in terms of mathematical complexity and the memory space required in storing them (Donoho, 2000). However, high dimensional data could have many variables which are redundant and therefore not necessary. There are a wide variety of dimensionality reduction methods which enable this problem to be circumvented.

In the literature, various dimensionality reduction methods have been proposed such as: Principal Component Analysis (Smith, 2002; Jolliffe, 2002), Fisher Linear Discriminant Analysis and Curvilinear Component Analysis (Demartines and Hérault, 1997b), Independent Component Analysis (Comon, 1994), Self Organising Maps (Kohonen, 2001) are also widely used. The discussion of all these methods is beyond the scope of this thesis; however, some of these techniques are used here and are discussed in this chapter. This chapter also discusses the classifiers used : Support Vector Machines (Chang and Lin, 2001) and the Fisher Linear Discriminant (Fisher, 1936).

3.2 Feature Extraction

Feature extraction is a method of capturing relevant information from the image in order to perform the desired task, using the reduced representation, instead of the full sized image. From the neurophysiology point of view, human sensory processing involves data reduction (Barlow,

1989) as well as feature extraction in the perceived image (Daugman, 1985). The cues on the face help humans to recognize the person and also the expression on their face. In order to develop a model capable of detecting these facial expressions, the face in an image has to be detected, followed by the expression. For this features on the face may be extracted. The facial features are the prominent components on the face, such as, the eyebrows, eyes, nose, mouth, and chin (Pantic and Rothkrantz, 2000). For any given face, these attributes have typically been placed into two groups: 'internal' attributes, comprising the eyes, nose and mouth, and 'external' attributes comprising the hair and jaw-line or chin (Jarudi and Sinha, 2003). The facial expressions are registered by changes in these features and these facial features may be aligned at various angles or orientations. Face expression are analysed as either holistic, analytic or hybrid. Holistic is representation as a whole. In analytic, the face is represented as a set of the above features and in hybrid, a combination of holistic and analytic techniques are used. Once these features are extracted, they have to be reduced in dimensionality and categorized by a classifier.

Computer-based recognition of facial expressions using features has a long history (Cao *et al.*, 2005), and various methods have been proposed. All the methods can be classified into two broad-based categories: (i) feature based approaches and (ii) holistic or probabilistic approaches. Most often, the feature-based methods utilize the Facial Action Coding System (FACS) designed by Ekman and Friesan (1978). Combinations of various muscle movements over the face represent an action unit (AU). In FACS, the emotions of the face are represented by values of 44 action units (AUs), and their combinations may describe any facial expression. Each expression is generated by the combination of several of these action units. More than 7,000 combinations of AUs have been observed. However, FACS itself is purely descriptive, uses no emotion and simply provides the necessary parameters to describe facial expressions and not the expression itself. The probabilistic-based method does not give preference to facial features such as the eyes and mouth. Instead, the feature vector can be the random distribution of image intensities (pixel values) and these vectors may differ for each emotion. The vectors are calculated per emotion and classification algorithms such as Neural Network (NN) and Hidden Markov Models or hybrid models (HMM or NN) are applied (Teo et al., 2004).

3.2.1 Gabor Filters

Mathematically, Gabor filters are a function obtained by modulating a sinusoidal function with a Gaussian function. A Gabor filter can be one or two dimensional (2D). A 2D Gabor filter is expressed as a Gaussian modulated sinusoid in the spatial domain and as a shifted Gaussian in the frequency domain. The key parameters of a Gabor filter are orientation and frequency. It is used to enhance certain features that share orientation and/or frequency and thereby enables useful pre-processing required for facial expression, recognition and analysis. By using a suitable

Gabor filter at the required orientation, certain features can be given high importance and other features less importance.

The Gabor filter is a Gaussian (with variances $S_x$and S_y along the x and y - axes respectively) modulated by a complex sinusoid plane (along x and y - axes respectively) and is described by Equation 3.1. The sinusoidal signal frequency is described as cycles/unit length and is described by Equation 3.2. The equation is complex in nature and has a real and imaginary part (Derpanis, 2007; Drakos and Moore, 1999). The Gaussian function is described by Equation 3.3.

$$g(x,y) = s(x,y)h(x,y) \tag{3.1}$$

The complex sinusoid is given by Equation 3.2.

$$s(x,y) = e^{-2\pi j(Ux+Vy)} \tag{3.2}$$

where U and V are the centre frequencies in the x and y directions.

The Gaussian envelope is given by Equation 3.3.

$$h(x,y) = \frac{1}{2\pi S_x S_y} \mathrm{e}^{-\frac{1}{2}\left\{\left(\frac{x}{S_x}\right)^2+\left(\frac{y}{S_y}\right)^2\right\}} \tag{3.3}$$

Hence, the full Gabor filter is given by the Equation 3.4.

$$g(x,y) = \frac{1}{2\pi S_x S_y} \mathrm{e}^{-\frac{1}{2}\left\{\left(\frac{x}{S_x}\right)^2+\left(\frac{y}{S_y}\right)^2\right\}-2\pi j(Ux+Vy)} \tag{3.4}$$

Figure 3.1 shows the real and imaginary part of the 2D Gabor filter.

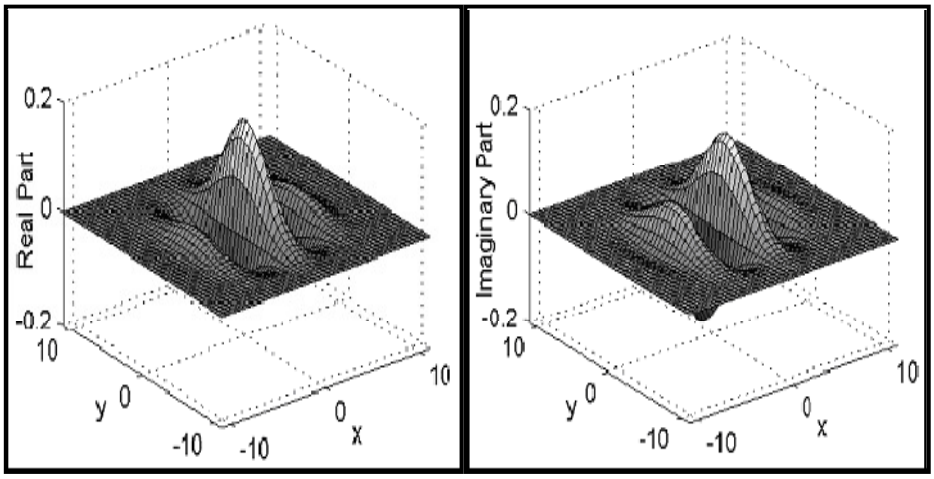


Figure 3.1: Plot of Real and Imaginary part of 2D Gabor filter. The main difference between the two images here is that they are out of phase.

A Gabor filter is therefore described by the following parameters:

1. The S_x and S_y of the Gaussian explain the shape of the base (circle or ellipse).
2. The frequency (f) of the sinusoid plane.
3. The orientation (θ) of the applied sinusoid.

As the Gabor filter is complex in nature, the image when filtered produces two parts: the imaginary part and the real part. The magnitude image can be obtained from the imaginary and the real parts. Here, only the magnitude of the filter is used for feature extraction. Figure 3.2 show examples of various Gabor filters (magnitude) at different frequencies and orientations.

There are two ways of performing Gabor filtering on face images:

- Analytical methods: These make use of specific points on the face, which are important feature points (fiducial points). There are two methods for identifying or locating these feature points: The Elastic Graph based method and Non graph based methods.

  - With Elastic graph based analytic methods, the face is represented by a rectangular graph with local features extracted at deformable nodes using Gabor wavelets, referred to as Gabor jets. Wiskott extended this method to Elastic Bunch Graph Matching (EBGM), where graph nodes are located at a number of facial landmarks (Wiskott *et al.*, 1999).

- Computationally, the Elastic graph method is complex, hence other simple methods of manually locating the feature points, or using colour, or edge information from the images, have been proposed and these are called Non-graph based methods (Shen, 2005). Escobar proposes to use Log-Polar images for Gabor feature extraction. The face image is Log-Polar transformed before it is convolved with Gabor wavelets. This technique is supposed to be more robust against the variance of scale and rotation. In this system, facial feature points are located manually and the coordinates are Log-Polar transformed as well (Escobar and Ruiz-del-Solar, 2002). The colour and edge information can also be used to extract facial features (Wu *et al.*, 2002).

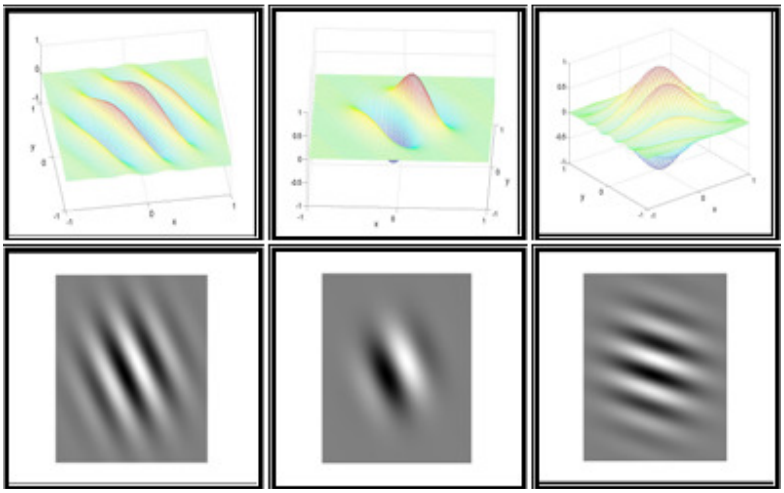
Once the location process is completed, recognition can then be performed using Gabor jets extracted from those feature points (Shen and Bai, 2006).

- Holistic methods: These methods normally extract features from the whole face image. An augmented Gabor feature vector is thus created, which produces a very large representation for the image. Once the feature vector is available, various methods have been proposed in the literature for using the feature vector and these will be further explained in the following section.

A well designed Gabor filter bank can capture the features of an image. These include repeating patterns in the image, the details of a pattern and its edge. Experimental results in texture analysis and character analysis demonstrate these features in the capture of local information with the different frequencies and orientations in the image (Zheng *et al.*, 2004a).

According to Daugman, aspects of the visual cortex can be mimicked by Gabor filters. The various 2D receptive-field profiles encountered in populations of simple cells in the visual cortex are well described by an optimal family of 2D Gabor filters (Grigorescu *et al.*, 2002; Jones and Palmer, 1987; Daugman, 1985; Kulikowski *et al.*, 1982; Escobar and Ruiz-del-Solar, 2002)

Jones and Palmer (1987) showed that the real part of the Gabor function fits very well with the receptive field weight functions for the simple cells in the cat striate cortex. Feature extraction using Gabor filters is considered effective for facial image representation (Jain and Farrokhnia, 1991; Movellan, 2002). Studies by Pollen and Ronner (1981) have shown that pairs of adjacent cells in the visual cortex of the cat are in quadrature (separated in phase by 90^{o}). The two adjacent cells can be regarded as the real and imaginary parts of a complex function and treat it as a complex receptive field (Movellan, 2002).



(a) (b) (c)

Figure 3.2: (a), (b), (c) are examples of Gabor filter with different frequencies and orientations. The top row shows their 3D plots and the bottom row, the intensity plots of their amplitude along the image plane. Normally filters at five different frequency scales and eight orientations are used over the image.

Since the local frequency and orientation of the features of the face are unknown, in most face recognition applications 40 Gabor filters are used (Shen and Bai, 2006). Five scales and eight orientations account for the forty filters normally used. Figure 3.3 shows all the 40 filters in 5 rows and 8 columns. The filter in the top row is of the highest frequency scale and is of decreasing scale in the rows below. Each column has filters for a particular angle.

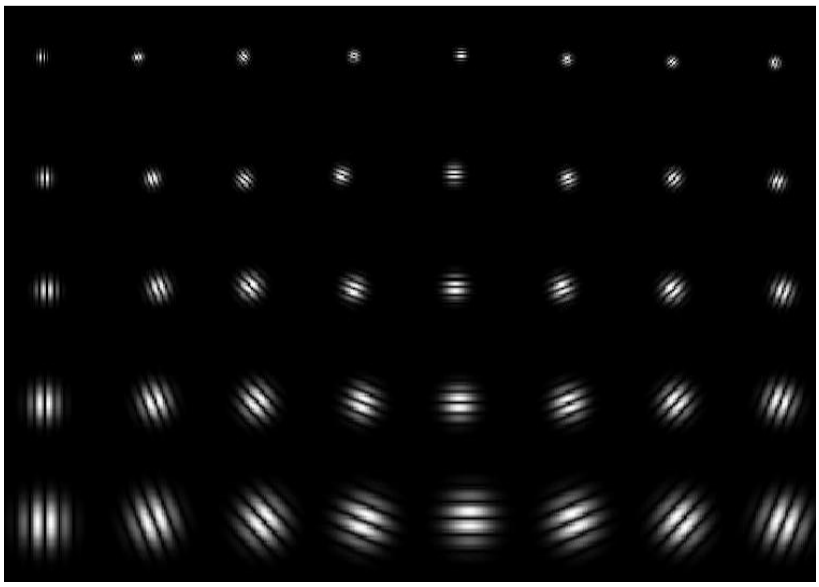
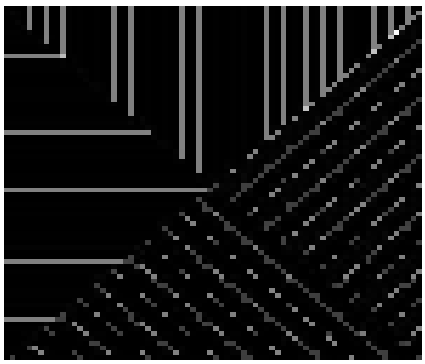
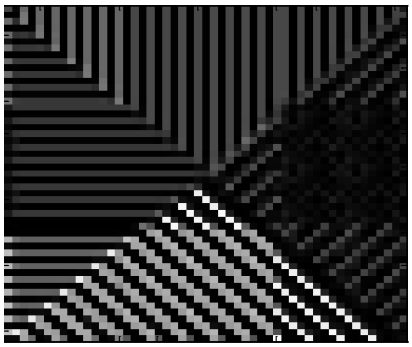


Figure 3.3: Gabor filters at five scales and eight orientations

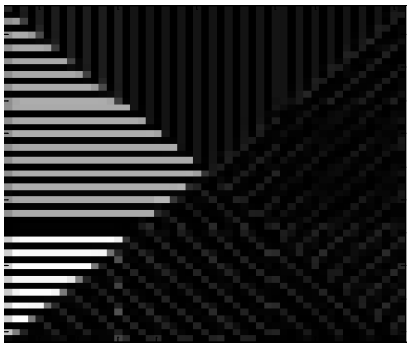
The effect of applying the filters can be best seen in the image which has lines at various angles and orientations. Figure 3.4(a) shows an image with lines at various angles. Figure 3.4(b) and 3.4(c) show the effect of applying a particular Gabor filter on Figure 3.4 (a). The highlighted lines in Figure 3.4 (b) and Figure 3.4 (c) shows the way the Gabor filter exaggerates lines at particular orientation similar to the results obtained earlier by others (Asirvatham, 2002) .



(a)



(b)



(c)

Figure 3.4: Gabor filtered images at various angles and orientations (a) Image with lines at various angles (b) Frequency, $\boldsymbol{f}$ = 12.5 and orientation, $\boldsymbol{\theta}$ = 135 degrees (c) Frequency, $\boldsymbol{f}$ = 25 and orientation, $\boldsymbol{\theta}$ = 0 degrees

An image such as a face has features at various angles and orientations and various frequencies. A Gabor filter bank with filters at 5 different frequency scales and 8 different angular orientations is capable up of capturing all the features of the face. Figure 3.5 is a sample image and the filters shown in Figure 3.3 are applied on the sample image. The resultant output from the filter bank is shown in Figure 3.6.

Figure 3.5: Sample Image of size $\mathbf{64 \times 64}$

In all the experiments performed here, the magnitude image is used with 5 frequency scales and 8 angular orientations.

By using the holistic method, features from the whole face image can be extracted. An augmented Gabor feature vector, which is the resultant image from the filter bank, is far greater in size than the original data for the image. This is because, as 40 filters are used; every pixel is represented by a vector of size 40. So a 64×64 image is transformed to size $64 \times 64 \times 40$.

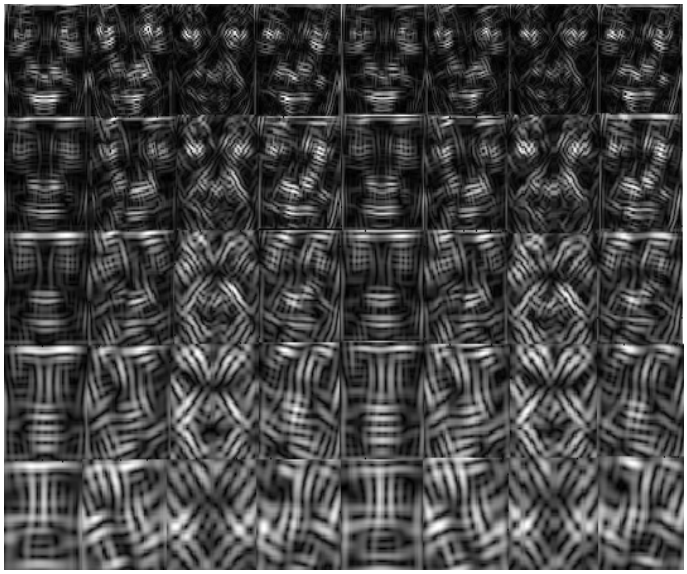


Figure 3.6: Magnitude part of the convolution output of a sample image shown in Figure 3.5 and the Gabor kernels (shown in Fig. 3.3).

Once the feature vector is obtained, it can be handled in various ways such as:

- The final image can be of the sum of the magnitudes of the Gabor filter coefficients at each location in the filter bank output.

- The pixel value in the final image would be the *L2 max norm* value of the feature vector obtained from the Gabor filter bank. This is simply the largest value from the Gabor filter bank output for every pixel of the original image (Grigorescu *et al.*, 2002; Kruizinga and Petkov, 1999)

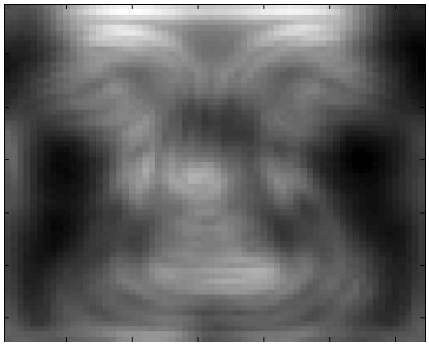
- Some methods use the feature vector as a concatenated vector and then perform dimensionality reduction such as PCA or even ICA (Liu and Wechsler, 2003).

- For the individual images (40 images) the energy content is obtained from the grey scale value. The mean and the variance can be obtained for every image. Thus the mean and variance is obtained for the entire filter bank (40filters). The final vector is represented by 80 bytes: 2 for each (mean and variance) Gabor filter output for every input image (Shen and Bai, 2004).

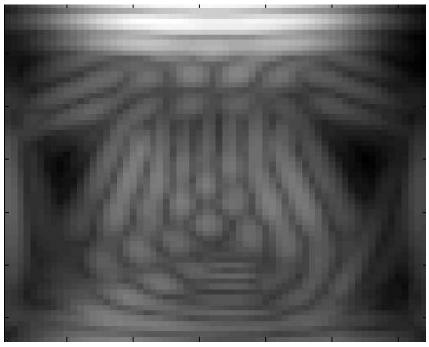
- The final image from the filter bank can also be the average of the corresponding pixels of the individual Gabor filter bank outputs.

- The final image from the filter bank could be the threshold output where the pixel value after performing the L2 max norm is compared with the threshold value and assigned magnitude 1 if greater than the threshold or 0 if less than the threshold (Kruizinga and Petkov, 1999).

(a)

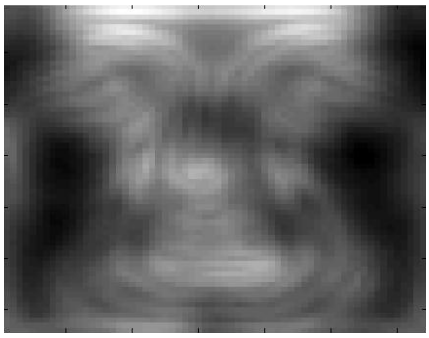


(b)



(c)

(d)



(e)

Figure 3.7:(a) Original Image (b) Sum Image (c) Superposition output (*L2 max norm*) (d) Threshold Output (e) Average Output

Figure 3.7 (c) shows the *L2 max norm* superposition output for the original image of Figure 3.7 (a). Similarly the outputs of the 40 filter banks can also be averaged or summed to give an output as in Figure 3.7 (b). All images displayed here are from the magnitude part of the Gabor filter outputs. The computational model used in the experiments here makes use of the *L2 max norm* superposition output. The technique adopted to find the *L2 max norm* superposition output can be explained with Figure 3.8. Each of the 40 filters produce an output of size 64 $\times$ 64, the

final output of the entire filter bank at a pixel (x, y) is obtained by comparing the pixel value at the same co-ordinates for all 40 filters. The pixel value at (x, y) is the largest at that point in all the 40 filter outputs. This is done for all the pixels of the entire image to get the *L2 max norm* superposition output for the filter bank.

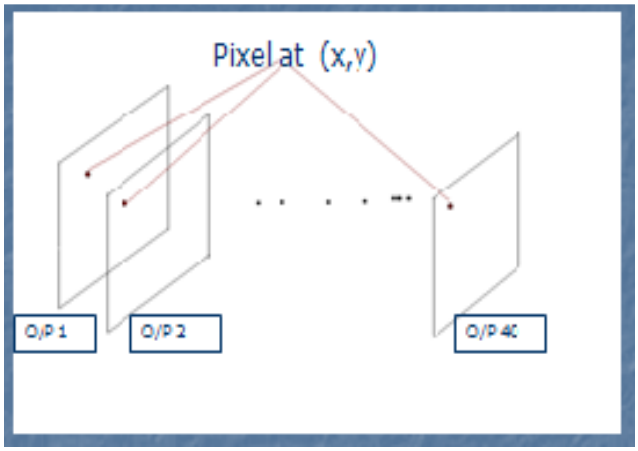


Figure 3.8: All 40 filter outputs used to find the *L2 max norm* superposition

3.3 Dimensionality Reduction

There are many techniques for dimensionality reduction such as, Principal Component Analysis or SVD decomposition (Smith, 2002; Jolliffe, 2002), Independent Component Analysis (Hyvärinen and Oja, 2000), Curvilinear Component Analysis (Demartines and Hérault, 1997b), Linear Discriminant Analysis (LDA), Fisher Linear Discriminant (Fisher, 2001), Multidimensional scaling, Projection pursuit, Discrete Fourier transform, Discrete Cosine transform (Jain, 1988), Wavelets, Partitioning in the time domain, Random Projections, Multidimensional scaling, Fast Map and its variants (Fodor, 2002; Gunopulos, 2001). The following methods are used here and are described in detail in this chapter: Principal Component Analysis (PCA), Curvilinear Component Analysis (CCA) and Linear Discriminant Analysis (LDA). Also, Fisher Linear Discriminant (FLD) which is an extension of LDA is described in its use for classification.

3.3.1 Principal Component Analysis

Principal Component Analysis (PCA) transforms higher dimensional datasets into lower dimensional uncorrelated outputs by capturing linear correlations among the data, and preserving as much information as possible in the data. PCA transforms data from the original coordinate system to the principal axes coordinate system such that the first principal axis passes through the maximum possible variance in the data. The second principal axis passes through the next largest possible variance and this is orthogonal to the first axis. This is repeated for the next largest possible variances and so on. All these axes are orthogonal to each other. On performing the PCA on the high dimensional data, Eigenvectors or principal components are obtained (Smith, 2002; Shlens, 2005). The required reduced dimensionality is obtained by retaining only the first few principal components.

PCA projects a $D-$dimensional dataset X into a $d-$ dimensional dataset Y, where $d \leq D$. Projecting the data from their original $D-$dimensional space onto the $d-$dimensional subspace spanned by these vectors then performs a dimensionality reduction that often retains most of the intrinsic information in the data (Smith, 2002; Jolliffe, 2002). The first principal component is taken to be along the direction with the maximum variance. The second principal component is constrained to lie in the subspace perpendicular to the first. Within that subspace, it points in the direction of the maximum variance. Then, the third principal component is taken in the maximum variance direction in the subspace perpendicular to the first two, and so on.

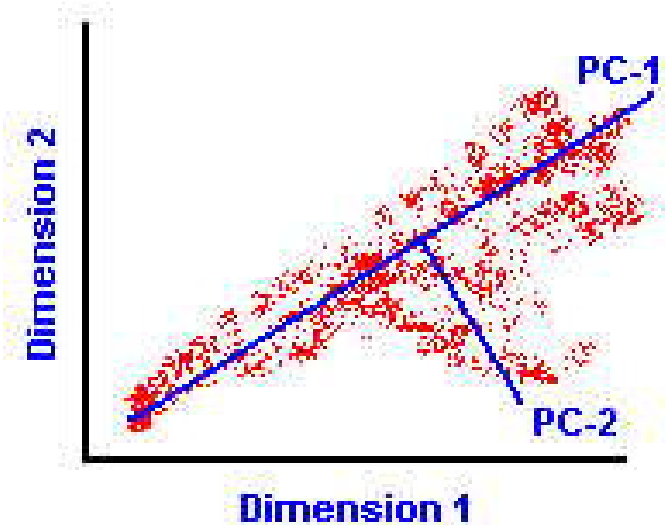


Figure 3.9: The blue lines represent 2 consecutive principal components. Note that they are orthogonal (at right angles) to each other.

Figure 3.9 shows the first two principal components. The steps involved in obtaining Principal components are detailed in Appendix A.

If face images are used in the PCA, then the principal vector or Eigenvectors are called Eigenfaces. The Eigenfaces are face like and capture variations of the faces in the dataset (Turk and Pentland, 1991). Figure 3.11 show the Eigenfaces for a dataset of 80 images which has 40 neutral expression and 40 smiling faces of equal number of male and female subjects from the FERET dataset (Philips *et al.*, 1998). Figure 3.10 shows examples of the images from the FERET dataset.

Figure 3.10: Example faces from the FERET dataset. The top row shows neutral faces and bottom row shows smiling faces

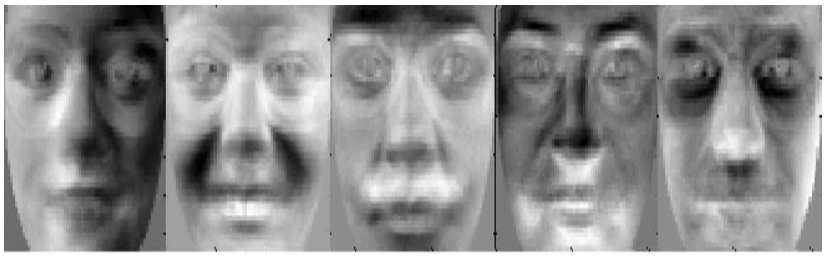


Figure 3.11: The first five Eigen faces for a set of FERET faces

Each image is of size 64×64 (4096 dimensions) and on performing PCA; it produces 79 Eigenfaces and components. Figure 3.11 shows the first 5 Eigenfaces in the order of importance. The total number of components to be retained for dimensionality reduction is based on the proportion of the variance of the first few components and the total variance of the complete dataset. In this work on performing PCA, the number of components to be retained is selected so as to preserve at least 95% of the variance of the data set. For this dataset of 80 face images (neutral and smiling), the first 66 components retain 95% of the total variance of the dataset. Hence, the PCA projection reduces the original 4096 dimensions to 66 components. This is still a large number and could be suggestive that the redundancy is not captured by a linear technique such as PCA and requires a non-linear technique such as CCA which is explained in the next section of this chapter.

3.3.2 Curvilinear Component Analysis

Curvilinear Component Analysis (CCA) is a non-linear projection method that attempts to preserve distance relationships in both input and output spaces. It is very similar to multidimensional scaling. CCA is a useful method for redundant and non linear data structure representation and can be used in dimensionality reduction. CCA is useful with highly non-linear data, where PCA or any other linear method fails to give suitable information (Demartines and Hérault, 1997a).

The $D-$ dimensional input X should be mapped onto the output $d-$ dimensional space Y. The $d-$ dimensional output vectors $\{y_i\}$ should reflect the topology of the inputs $\{x_i\}$. In order to do this, Euclidean distances between the x_i's are considered. Corresponding distances in the output space y_i's is calculated such that the distance relationship between the data points is maintained. CCA puts more emphasis on maintaining the short distances than the longer ones. Formally, this reasoning leads to the following error function:

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(d_{i,j}^{X} - d_{i,j}^{Y}\right)^2 F_\lambda\left(d_{i,j}^{Y}\right) \qquad \forall\, j \neq i \qquad (3.5)$$

where $d_{i,j}^{X}$ and $d_{i,j}^{Y}$ are the Euclidean distances between the points i and j in the input space X and the projected output space Y respectively and N is the number of data points. $F_{i,j}^{Y}$ is the neighbourhood function, a monotonically decreasing function of distance. In order to check that the relationship is maintained a plot of the distances in the input space and the output space $(dy - dx)$ plot is produced. For a well maintained topology, dy should be proportional to the value of dx at least for small values of dx's.

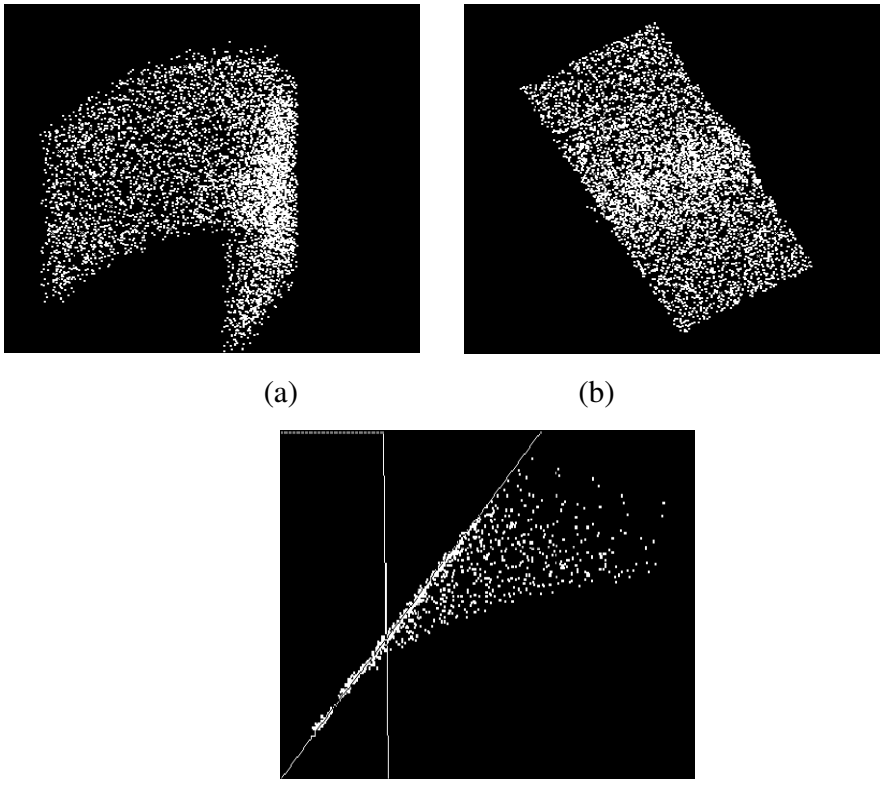


Figure 3.12: (a) 3D horse shoe dataset (b) The 2D CCA projection of the horse shoe dataset (c) $(\boldsymbol{dy - dx})$ plot of the projection showing that small distances are maintained, although it is not possible to maintain the larger distances.

Figure 3.12 shows CCA projections for the 3D data taken initially. The $(dy - dx)$ plot shown is good in the sense that the smaller distances are very well matched (Demartines and Hérault, 1997a).

For the dataset mentioned earlier as in PCA, with the CCA only 14 components are retained, the $(dy - dx)$ plot of this is shown in Figure 3.13.

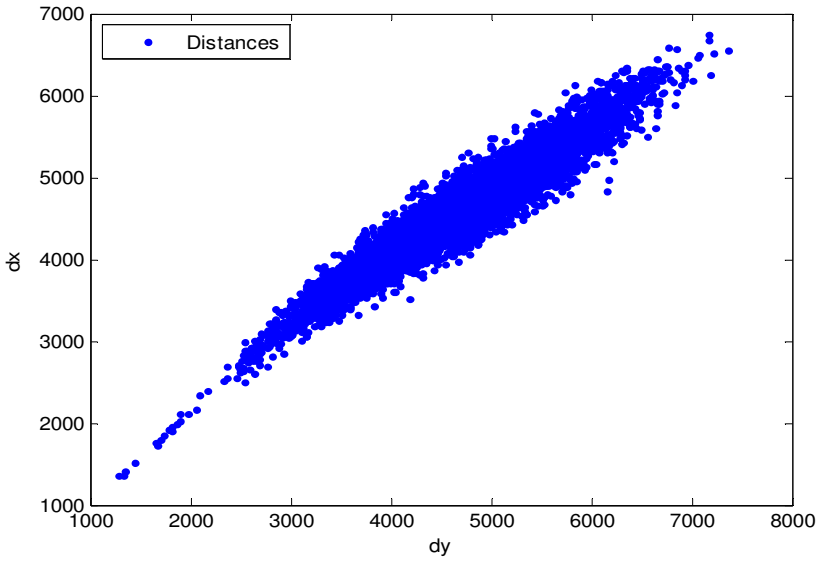


Figure 3.13: The $(\boldsymbol{dy - dx})$ plot for the dataset with 80 images of equal number of smiling/neutral, male/female faces and where 14 components were retained.

3.3.3 Intrinsic Dimension

One problem with CCA is deciding how many dimensions the projected space should occupy and one way of estimating this is to use *Intrinsic Dimension* of the data manifold. The Intrinsic Dimension (ID) can be defined as the minimum number of free variables required to define the data without any significant information loss. Due to the possibility of correlations among the data, both linear and nonlinear, a $D-$dimensional dataset may actually lie on a $d-$dimensional manifold $(d \leq D)$. The ID of such data is then said to be d. There are various methods of estimating the ID. These are based on the *fractal dimension* (Camastra and Vinciarelli, 2001) and there are three popular methods in estimating this. These are the Box Counting, Information Dimension and Correlation Dimension methods. The box counting method and the information dimension method are suitable when the dimensions are small but are not practical for use with large or high dimensional dataset with faces. With face images, the best intrinsic method to use is the Correlation Dimension.

The Correlation Dimension method was developed by Grassberger and Procaccia (1983). This method finds the closeness between the points at different scales, and then the dimension is calculated by measuring how the closeness of the neighbouring point is affected by the scales used. A measure of this closeness is called the Correlation Integral $C(l)$.

It can be calculated as follows:

$$C(l) = \frac{2}{(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} K \tag{3.6}$$

$$where \begin{cases} K = 1\,, if\ d_{i,j} \leq l \\ K = 0\,, if\ d_{i,j} > l \end{cases}$$

N is the number of data points, l is the length variable and $d_{i,j}$ is the Euclidean distance between the i^{th} and j^{th} data points of the dataset. The total number of pair wise points closer to each other than length l is proportional to l^d (Grassberger and Proccacia, 1983).

Assume

$$C(l) = k\, l^d \tag{3.7}$$

where d is the dimension of the data and k is a constant.

$$\log C(l) = \log k + d\, log l \tag{3.8}$$

So,

$$\frac{\log C(l)}{\log l} = \frac{\log k}{\log l} + d \tag{3.9}$$

$$\text{Take } l \to 0 \text{ then } \log(l) \to \infty$$

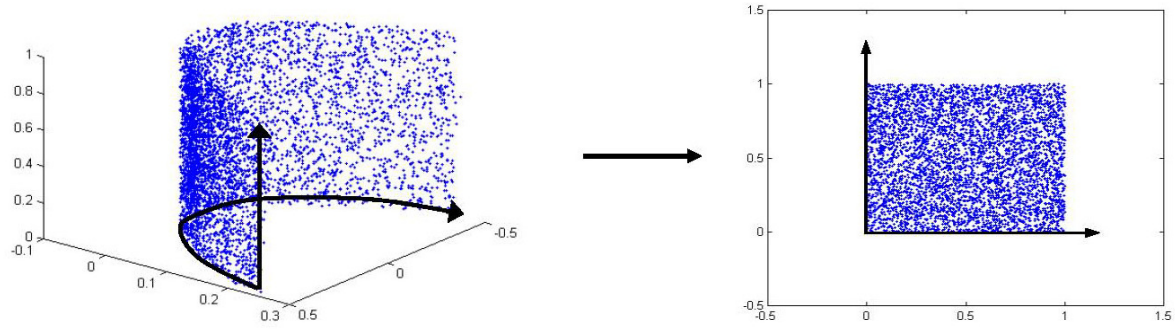
So, the dimension d can be calculated as:

$$d = \log C(l) / \log l \tag{3.10}$$

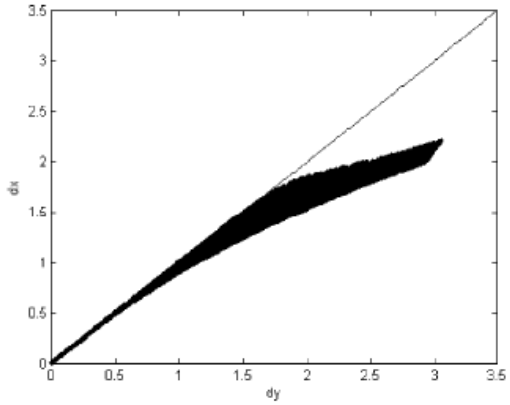
The Correlation Dimension d_c can be calculated by measuring the closeness property at all scales as follows:

$$d_c = \lim_{l \to 0} \frac{\log C(l)}{\log l} \tag{3.11}$$

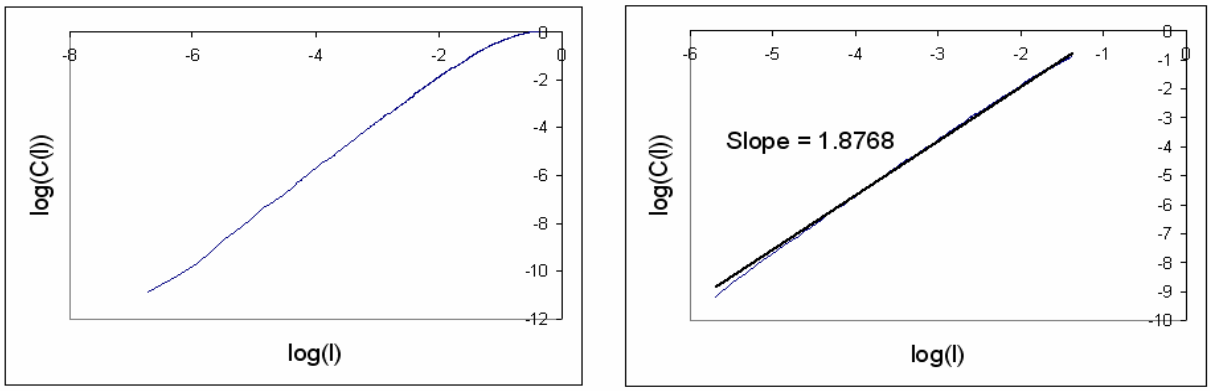
Figure 3.14(a) shows the well known horse shoe data set and the plot of $log\ C(l)$ versus $C(l)$ which is the Correlation Dimension for the horse shoe data is shown in Figure 3.14(c) and Figure 3.14(d) shows how the Correlation Dimension is estimated by considering the most linear part of the curve and measuring its slope. Though the 2D non linear projection of the of the 3D horse shoe distribution looks perfect as shown in Figure 3.14(a), the $(dy - dx)$ plot of the projection will have smaller distances maintained and larger distances are not so well maintained and is shown in Figure 3.14(b) (Buchala *et al.*, 2005). Different intervals on the curve shown in Figure 3.14 (c) must be selected and the slope from the linear portions of this curve gives the correlation dimension.



(a)



(b)



(c) (d)

Figure 3.14: (a) A 2-dimensional nonlinear projection of 3-dimensional horseshoe distribution (b) The $(\boldsymbol{dy - dx})$ plot of the projection showing that small distances are maintained, although it is not possible to maintain the larger distances. (c) Correlation Dimension plot of the horse shoe data. (d) The Correlation Dimension is calculated as the slope of the most linear part of the curve.

3.3.4 Fisher Linear Discriminant Analysis

Fisher Linear Discriminant Analysis (FLDA) has been successfully applied to face recognition, which is based on a linear projection from the image space to a low dimensional space by maximizing the between-class scatter and minimizing the within-class scatter. It is most often used for classification (Welling, 2005; Fisher, 2001). The main idea of the FLD is that it finds projection to a line so that samples from different classes are well separated (Veksler, 2006).

3.3.4.1 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a special case of FLD in which both classes have the same variance. It makes use of the class label for dimensionality rather than just the features of the data points. Belhuemer was the first to use the LDA on faces and used it for dimensionality reduction (Belhumeur *et al.*, 1997) and it can be used as a classifier.

In other words, LDA moves images of the same face closer together, while moving images of different faces further apart. For a two class problem it is commonly known as Fisher Linear discriminant analysis after Fisher who used it in his taxonomy based experiments (Fisher, 1936). Eigenfaces attempt to maximise the scatter of the training images in face space, while Fisherfaces which are obtained by performing the linear discriminant analysis (LDA) attempt to maximise the between class scatter, while minimising the within class scatter.

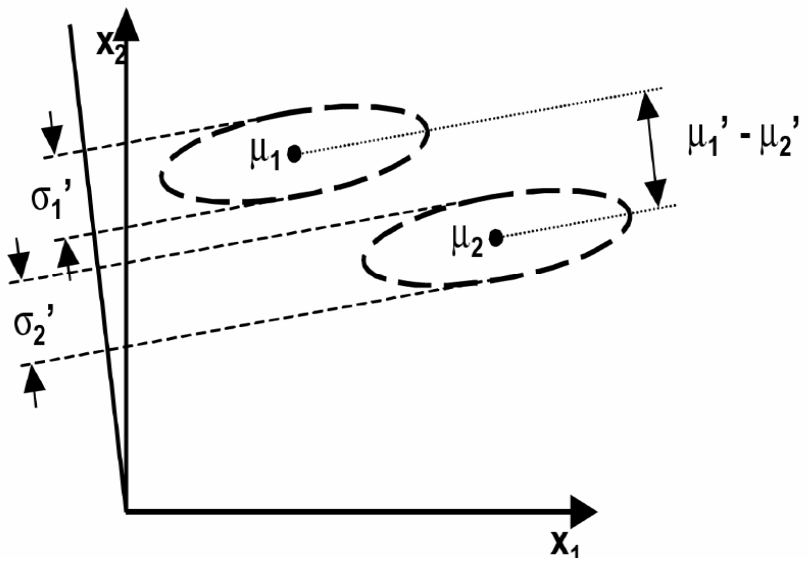


Figure 3.15: Figure shows the classes which are overlapping along the direction of X1. However, they can be projected on to direction X2 where there will be no overlap at all.

In a dataset with two classes, the dimension most important for classification would be the one with maximum difference in the means of the two classes. In the example shown in Figure 3.15, the difference between the classes is higher in the direction X1, but with considerable amount of overlap. So, the best direction is X2 due to the lesser within class variance. The better class separability can be obtained by the within class variance.

The between class scatter covariance matrix is given by:

$$S_B = (\mu_2 - \mu_1)(\mu_2 - \mu_1)^T \tag{3.12}$$

The within class covariance matrix is given by:

$$S_W = \sum_{i=1}^{C_i} \sum_{N \in C_k} (X^n - \mu_i)(X^n - \mu_i)^T \tag{3.13}$$

where μ_1 and μ_2 are the means of the datasets of the class 1 and class 2 respectively. C is the number of classes and C_k is the k^{th} class. The eigenvector solution of $S_W^{-1}S_B$ gives the fisher face.

3.3.4.2 Expression encoding power

When PCA is performed, the first few components encode the maximum variance. However, as face data has multiple properties though the first few components encode maximum variance they may not be of interest and if the property of interest of the data is encoded by the last few components then this method would be disadvantageous. Hence, the selection of the components should be such that they are based on the importance of the property rather than the total variance. The LDA seems to be a perfect answer to this as an analysis can be performed on the separation matrix (Etemad and Chellappa, 1997) to obtain the discriminant power of the components in a similar way as we find the eigenvalues on the covariance matrix (Turk and Pentland, 1991).

The discriminating power is defined as the ratio of projection of the between-class variance to the projection of the within-class variance. The discriminant power of the dataset can be explained in terms of eigenvalues. This is obtained by first summing up all the eigenvalues which are obtained for the separation matrices to get a measure of the total discriminating power. This result is divided into each individual eigenvalues to get its proportion of the total power. The larger the eigenvalue, the greater is the discriminating power. The eigenvalues can be expressed as relative percentages. If μ_1 is the mean of neutral face image dataset and μ_2 is the mean of the smiling face image dataset and with λ_i being the eigenvalue of the i^{th} component of

the property, the discriminating power (e.g., expression, age, gender and so on) or the encoding expression power P_i is given by Equation 3.14.

$$P_i = \frac{\lambda_i}{\sum_j^n \lambda_j} \tag{3.14}$$

where P_i is a measure of the encoding power of the i^{th} component of the property (e.g., Expression, Gender, Age and so on) and n is the number of non-zero eigenvalues.

The LDA can be used to estimate the encoding power of the various face properties such as expression, gender, age, identity and race. Using the two classes namely, neutral and smiling, LDA successfully transforms it into a space which has very large between class variance and very small within class variance.

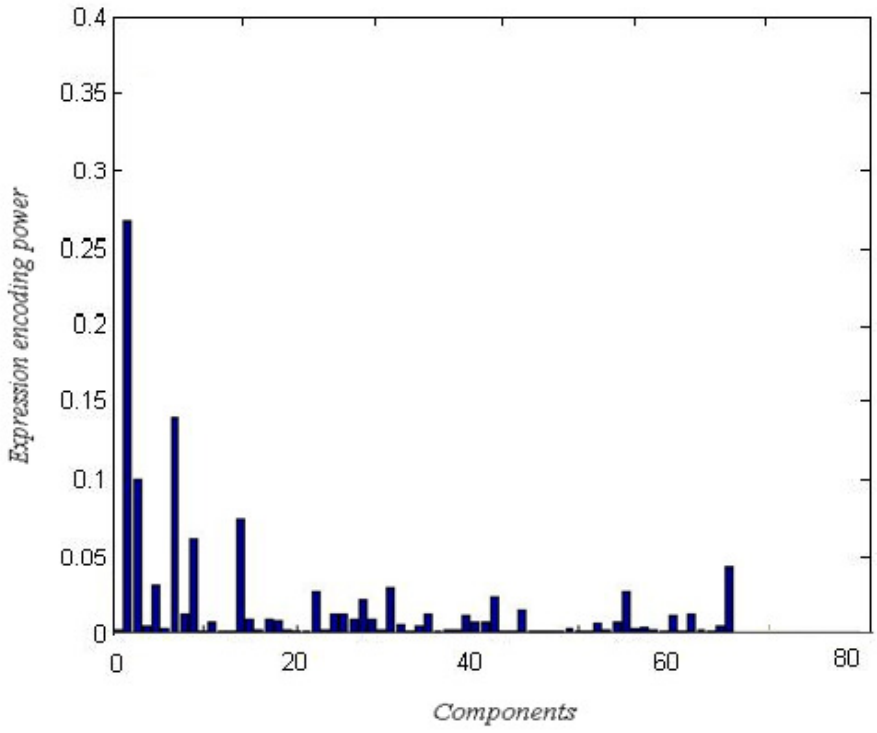


Figure 3.16: Expression encoding power for the first 66 components of the FERET dataset as mentioned earlier with PCA. The second component has the highest expression encoding power.

By using the within class variance and the between class variance, the encoding power for the expression property can be obtained by using Equation (3.14) and can be viewed as in Figure 3.16. Figure 3.16 shows the encoding power for the expression property of the face and it

suggests that some of these initial components are not significant for expression and some of them are significant (the larger the value the more significant).

With high dimensional data, it is often not possible to perform LDA as there can be a problem with singular matrices and therefore PCA is normally used to pre-process the data and reduce its dimensionality. Also, if the number of dimensions is more than the number of data points the computational complexity with LDA is overcome by using PCA first (Belhumeur *et al.*, 1997). With face images, this is often true and hence, PCA is used to reduce the data to 66 components from the original 4096 dimensions (64×64 image) and the LDA (with two classes for smiling and neutral) helps in finding the encoding power of the expression property.

The steps involved in finding the Fisher face are as follows:

1) For N samples $\{x_1, \ldots, x_N\}$, C classes $\{X_1, \ldots, X_c\}$, the average μ_i for each class i is calculated along with the total average μ.

2) The Scatter for each class i is calculated as:

$$S_i = \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T$$

3) The within class scatter is calculated as:

$$S_W = \sum_{i=1}^{c} S_i$$

4) The between class scatter is calculated as:

$$S_B = \sum_{i=1}^{c} | X_i | (\mu_i - \mu)(\mu_i - \mu)^T$$

5) The linear transformation or LDA is given by a matrix V whose columns are the eigenvectors of $S_W^{-1} S_B$ (called *Fisherfaces*).

6) The Eigenvectors are solutions of the generalized Eigenvector problem:

   $S_B V = D S_W V$ where V will have the Eigenvector which in this case is called the Fisherface and D will have the

eigenvalue and in this case with 2 classes will have only one non zero value.

7) If S_W is non-singular, then we can obtain a conventional eigenvector problem by writing:

$$S_W^{-1} S_B V = DV$$

8) In practice, S_W is often singular since the data are image vectors with large dimensionality while the size of the data set is much smaller. Hence we project original data to the PCA space $S_{BB} = P^T \times S_B \times P$ and $S_{WW} = P^T \times S_W \times P$ where P is the matrix of Eigenfaces obtained from the PCA and used for fisher face.

9) Hence, the eigenvaues are obtained by solving: $S_{WW}^{-1} S_{BB} V = DV$

A LDA projection of the dataset that was used with PCA and CCA gives the Fisher face shown in Figure 3.17.

Figure 3.17: The LDA reduced the dimensionality from 66 to one and the corresponding Fisher face is shown here.

3.3.5 Effect Size

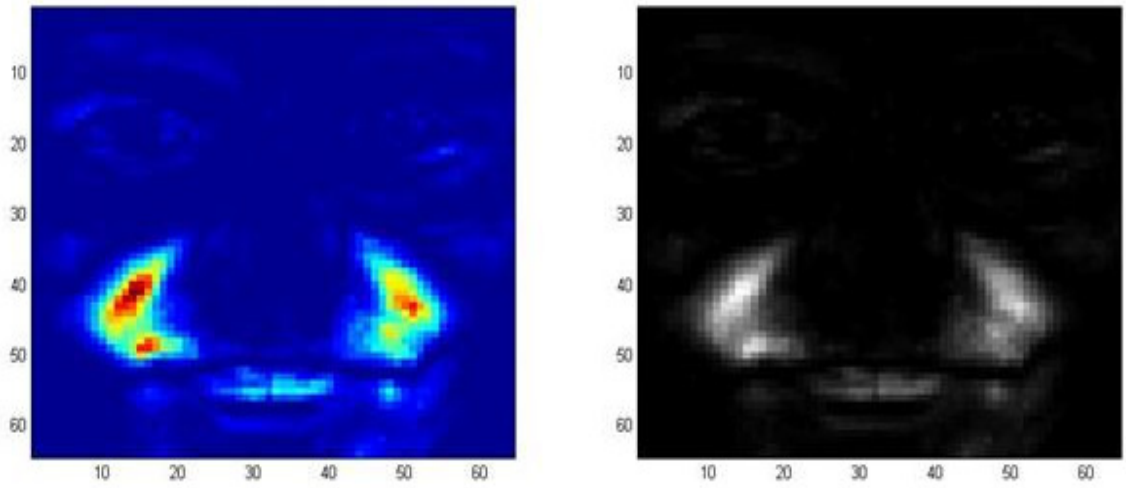
Effect size is a way of expressing the difference between two groups. Here two groups: Smiling and Neutral are used. Cohen (1988) defined d as the difference between the means, $\mu_1 - \mu_2$, divided by standard deviation, σ of either group.

$$d = \frac{\mu_1 - \mu_2}{\sigma} \tag{3.15}$$

μ_1 and μ_2 are the means of two groups and σ is the standard deviation of the whole population is calculated by Equation (3.16).

$$\sigma = \sqrt{\frac{(\sigma_1^2 + \sigma_2^2)}{2}} \tag{3.16}$$

σ_1 and σ_2 are the standard deviation of the two classes, Smiling and Neutral respectively and N is the total number of samples. The 'Encoding face' is obtained by finding the Effect size of each pixel in an image. In other words which pixels discriminate most between smiling and neutral faces can be seen and the result of this analysis is shown in Figure 3.18.



(a) (b)

Figure 3.18: (a) Colour image of the encoding face (b) The gray scale image of the encoding face. The features picked up are clearly seen in colour image than in the gray scale image.

3.4 Classification

A number of classifiers can be used in the final stage for classification; however, Support Vector Machines have been used for all the classification of expressions.

3.4.1 Support Vector Machines

The Support Vector Machine (SVM) classifier is becoming very popular these days although the subject can be said to have started in the late seventies (Vapnik, 1979) and it has been used in pattern classification and regression (Cortes and Vapnik, 1995). They belong to a family of generalized linear classifiers.

The basic idea of an SVM is to find the optimal separating hyper-plane, that has the maximal margin of separation between the classes, while having a minimum number of classification errors. This means the SVM classifier tries to find the plane which separates the two different classes such that it is equidistant from the members of either class which are nearest to the plane.

SVM's are used extensively for a lot of classification tasks such as: handwritten digit recognition (Cortes and Vapnik, 1995) or Object Recognition (Blanz *et al.*, 1996). SVM's can be slow in test phase, although they have a good generalization performance. In total the SVM theory says that the best generalization performance can be achieved with the right balance between the accuracy attained on the training data and the ability to learn any training set without errors, for the given amount of training data. The SVM shows better classification accuracy than Neural Networks (NNs) if the data set is small. Also, the time taken for training and predicting the test data is much smaller for a SVM system than for a NN (Zheng *et al.*, 2004b).

Consider an input training set, $S = \{(x_1, y_1), (x_2, y_2), \ldots . (x_N, y_N)\}$ of objects $x_i \in X$ and their known classes $y_i \in \{-1, +1\}$. The Output of the classifier is $f : X \longrightarrow \{-1, +1\}$ which predicts the class $f(x)$ for any (new) object $x \in X$. This can be explained by the Figure 3.19. The two classes are separated by an optimum hyper-plane, illustrated in Figure 3.19, minimizing the distance between the closest +1 and -1 points, which are known as *Support Vectors*. *Support Vectors* are the data points that the margin is closest to. The right hand side of the separating hyper-plane represents the +1 class and the left hand side represents the -1 class (McCulloch, 2005).

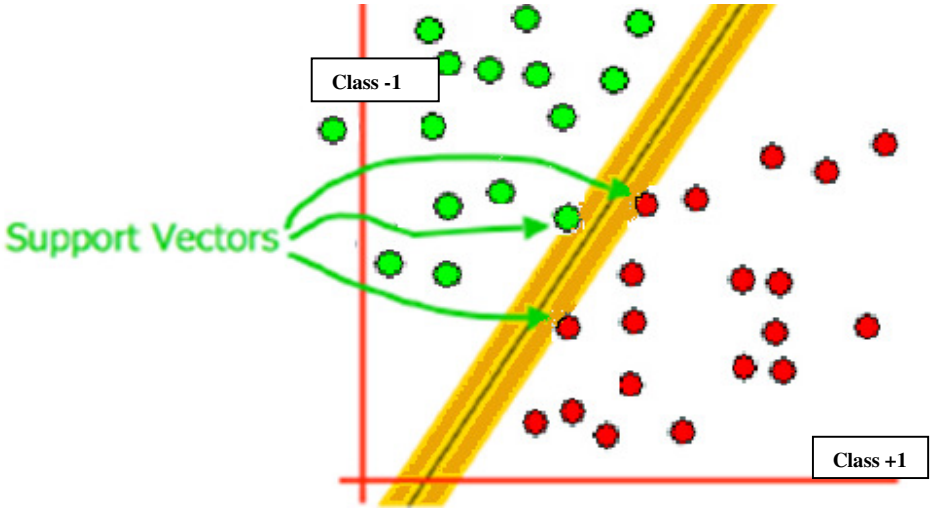


Figure 3.19: SVM Classifier with optimal hyper-plane

Maximizing the Margin (γ) between the two classes would be the optimal hyper-plane. With a data point x_2 of class -1 and another data point x_1 of class +1, the hyper-plane between the two classes can be defined by the equation:

$$\mathbf{w}.\mathbf{x} + b = 0 \tag{3.17}$$

The decision function for the classifier is given by:

$$f(x) = sign(\mathbf{w}.\mathbf{x} + b) \tag{3.18}$$

If the two classes are linearly separable, then the following equation is always true:

$$y_i(\mathbf{w}.\mathbf{x} + b) \geq 1 \quad \forall i \tag{3.19}$$

For data point x_1 on the margin of class +1:

$$\mathbf{w}.x_1 + b = +1 \tag{3.20}$$

And for the data point x_2 on the margin of class -1:

$$\mathbf{w}.x_2 + b = -1 \tag{3.21}$$

Hence,

$$(\mathbf{w}.x_1 + b) - (\mathbf{w}.x_2 + b) = \mathbf{w}(x_1 - x_2) = 2 \tag{3.22}$$

For the separating hyper-plane, the normal vector is given by:

$$\widehat{\mathbf{w}} = \frac{\mathbf{w}}{\|\mathbf{w}\|} \tag{3.23}$$

The margin (γ) is half the projection of $(x_1 - x_2)$ on to the normal vector and is given by:

$$2\gamma = \frac{\mathbf{w}(x_1 - x_2)}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \tag{3.24}$$

This implies $\gamma = \frac{1}{\|\mathbf{w}\|}$ and to maximize this term the following term has to be minimized

$$\min\left(\frac{1}{2}\right)\|\mathbf{w}\|^2 \tag{3.25}$$

subject to

$$y_i(\mathbf{w}.\mathbf{x_i} + b) \geq 1 \qquad \forall i \tag{3.26}$$

The SVM is trained to find the value of α that maximizes the following equation, so by applying the Lagrange multiplier to the Equation (3.25) and (3.26), we get:

$$L(\vec{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j \; y_i y_j \left(x_i \cdot x_j\right) \tag{3.27}$$

under the constraints $0 \leq \alpha_i \leq C$ for $i = 1 \ldots\ldots N$ and $\sum_i^N \alpha_i y_i = 0$. C is the cost parameter and α is the optimizing parameter for the training process.

In the example shown in Figure 3.20, the objects belong either to class GREEN or RED. The separating line defines a boundary on the right side of which all objects are GREEN and to the left of which all objects are RED. Any new object falling to the right is labelled, i.e., classified, as GREEN or classified as RED if it falls to the left of the separating line.

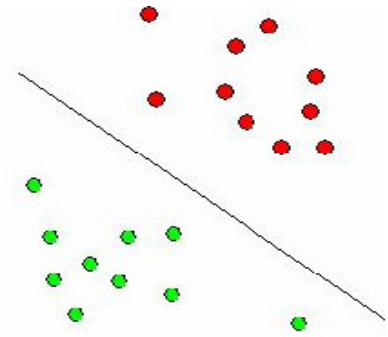


Figure 3.20: A Linear Classifier

Most classifications are not this simple, and a more complicated example is shown in Figure 3.21. In this example, it needs a non-linear separator rather than a straight line to separate the two classes.

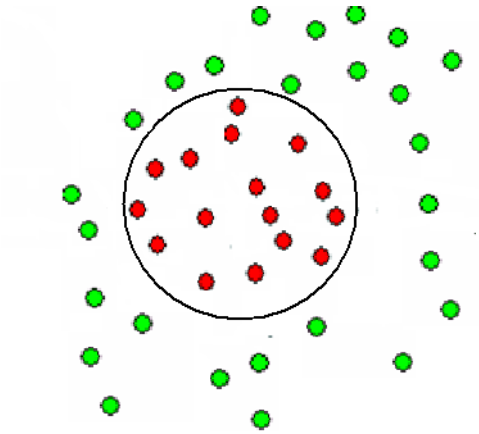


Figure 3.21: A non Linear Classifier

A SVM rearranges the original objects (data points) according to a mathematical function (kernels) and transforms it into a feature space which allows the classification to be accomplished more easily, and is illustrated in Figure 3.22. Mapping the input data points into a different co-ordinate space is called projecting into the feature space.

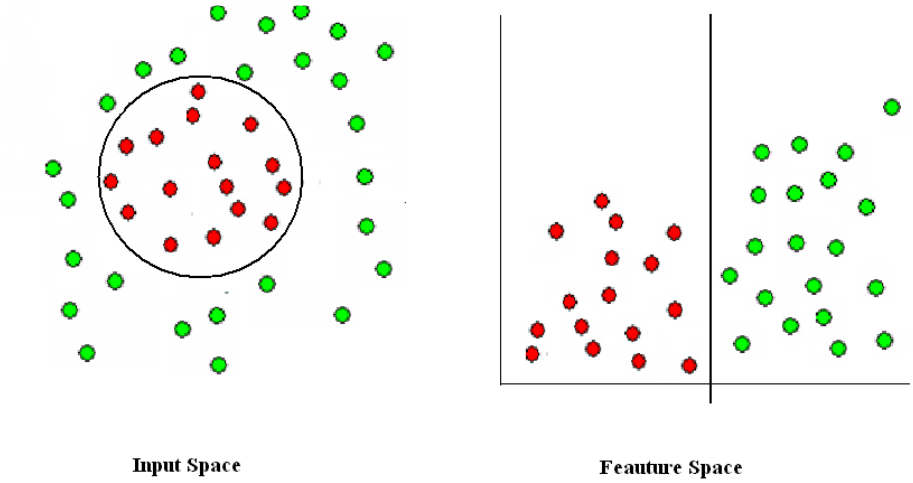


Figure 3.22: Transformation from input space to Feature space by the Support Vector Machine. The data points cannot be separated in the Input space by a linear separator. Hence on projecting onto a polar coordinate system (Feature space); the data points can be separated by the linear separator.

Figure 3.22 shows the Feature space and the Input space. When the input data points are projected into the polar coordinate system, they can be easily separated by a straight line (linear separator) which is a circle (non-linear separator) in the original two dimensional input space.

In general, kernels are used to map the datasets to a higher dimensional feature space which is normally linear in nature and normally there is no need to explore the actual feature space. By using a Kernel all the computations can be done on the original data in the input space. In Equation 3.22 the $K(x_i, x_j)$ can be replaced for the dot product $(x_i \cdot x_j)$ and it is called the kernel function and most often; for classification purpose a Radial Basis function (RBF) is used. By using a RBF kernel, the input space is projected into a very high dimensional space and can linearly separate any data in such a large feature space. There are two parameters when using RBF kernels: C and γ. Here, C is the cost parameter and γ is the kernel parameter. It is not known beforehand which C and γ are the best for one problem; consequently some kind of model selection (parameter search) must be done. The kernel maps the input data points into a higher dimensional feature space.

There are number of kernels that can be used in Support Vector Machines models. These include linear, polynomial, Radial Basis Function (RBF) and sigmoid:

$$\begin{Bmatrix} x_i * x_j & Linear \\ (\gamma x_i^T x_j + t)^d & Polynomial \\ \exp(-\gamma |x_i - x_j|^2) & RBF \\ \tanh(\gamma x_i x_j + t) & Sigmoid \end{Bmatrix}$$

Here t, d and γ are Kernel parameters.

The RBF is by far the most popular choice of kernel types used in Support Vector Machines. The best separating hyper-plane that can be constructed by the SVM can be defined by:

$$f(x) = \sum_{i=1}^{N} \alpha_i \, y_i \, K(x_i, x) + b = 0 \tag{3.28}$$

3.4.2 SVM – Parameters, Over-fitting and Validation

The goal is to identify the best value so that the classifier can accurately predict unknown data (i.e., testing data) (Chih-Wei Hsu, 2008). The parameter C, if it is too large, provides a high penalty for non-separable points and we may store many support vectors and over-fit. If it is too small, we may have under-fitting.

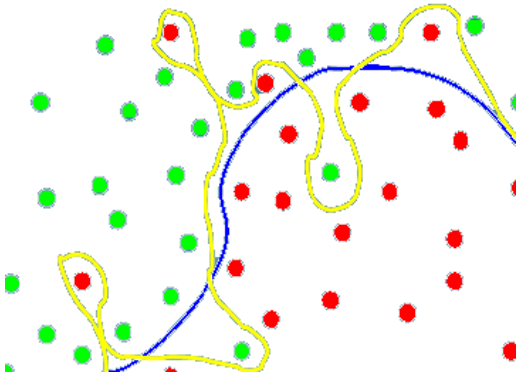


Figure 3.23: An Over-fitting Classifier. The Yellow line represents over-fitting classifier and the blue line represents the SVM classifier with a few misclassifications.

This literally means that the parameter C controls the trade-off between the misclassification errors on the training set and the margin between the two classes. Over-fitting means fitting too much of the training data and could result in too many errors (Vert, 2002) and an example for this is shown in Figure 3.23.

Classifiers can accurately predict training data whose class labels are indeed known. Therefore, the best way to achieve this is by separating the training data into two parts of which one is considered unknown in training the classifier. The classifier is trained by one half of the data set and then the prediction accuracy on the remaining set can be more precisely predicted (the other half of the training set not used for training). An improved version of this procedure is *cross-validation*. In v-fold cross-validation, we first divide the training set into v subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining $v-1$ subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. This cross-validation procedure can prevent the over-fitting problem.

There are two forms of cross-validation:

- The training set is divided into v subsets. One of them is used as the test set and the remaining $v-1$ sets are used for training to get the values for C and γ. This is repeated sequentially taking one subset as the test set while training the remaining subsets in order to get the best values for C and γ. Finally, the model is trained with the best parameters and test set is predicted. This process is adopted for experiments explained in Chapter 4.

- The entire dataset (training and test) is divided into m subsets each of the same size as the test set. The test set is predicted by training the remaining $m-1$ subsets by using the best values of for C and γ obtained by performing a fivefold cross validation on the set used as the training set. Sequentially this procedure is repeated for all the m subsets. Thus, each instance of the whole dataset is predicted once. This process is adopted for experiments explained in Chapter 5.

3.4.3 Steps involved in training the Support Vector Machine

The LIBSVM tool (Chang and Lin, 2001) can be used for SVM classification. The SVM can be trained in the following way:

1. Transforming the data to a format required for using the SVM software package - LIBSVM 2.86 (Chang and Lin, 2001).

2. Perform simple scaling on the data so that all the features or attributes are in the range [-1, +1].

3. Choose a kernel. Most often we use RBF, $k(x,y) = e^{-\gamma|x-y|^2}$ Kernel.

4. Perform fivefold cross validation with the specified kernel to find the best values of the parameter C and γ where C is the cost parameter.

5. Use the best parameter value of C and γ to train the whole training set.

6. Finally Test.

3.5 Discussion

Chapter 3 explains feature extraction of face images with Gabor filters and the various types of Dimensionality reduction techniques used and for reducing the high dimensional data set of images with various face expressions. Methods such as PCA, CCA, LDA and FLD; also, effect size and encoding power were also discussed. The true dimension estimation or intrinsic dimension of data set reduced by dimensionality technique such as CCA is also discussed. Classification is purely done by Support Vector Machines and has been discussed in detail in this chapter. The training of SVM, over-fitting and validation were all investigated.

CHAPTER FOUR

Recognizing Smiling and Neutral Expressions

4.1 Introduction

The previous chapter discussed many types of computational techniques used in face image processing and they included: feature extraction by Gabor filters, dimensionality reduction by PCA, CCA, LDA and FLD, and classification with SVM's. In this chapter, I explain how all these computational methods have been used on a face expression image dataset (FERET) (Philips *et al.*, 1998). Only two expressions: Smiling and Neutral are used in this experiment. This work shows that it is possible for a computational system to differentiate faces with a neutral expression from those with a smiling expression with high accuracy using these techniques.

4.2 Dataset Description

The FERET dataset is widely used, in many face recognition experiments as it provides a large appropriate data set (Rizvi *et al.*, 1998). It consists of face images of over 1200 individuals with multiple face images for each individual. The images are of grey scale and vary in pose, lighting angle, changes in expression, with or without glasses and some with beard and/or moustaches. Each individual has a number of expressions and in some cases have been photographed after a considerable time gap. The original images of the FERET dataset are of size 384×256 and included visible hair and clothing in some cases. The images used here were cropped to size 150×130 so that little or no hair is visible; further, histogram equalization was done to achieve uniformity, compensating for the various lighting conditions used for individual images.

The neutral faces were clearly labelled in the dataset description sheet, but the smiling faces were not labelled as such. Therefore I presented a selection of faces to a group of 5 people and where they all agreed that a face was smiling; it was placed in the smiling class.

A total of 120 faces were used for the experiment (30 male and 30 female) each with two classes, Neutral and Smiling expression (60 faces for each expression). Figure 4.1 shows an example set from the database and Table 4.1 explains the dataset used. With all faces aligned, based on their eye location, a 128×128 image was cropped from the original raw image of size 150×130 and further reduced to size 64×64 to reduce the computational complexity. Though they have been processed to exclude the external features of the face, since they have not undergone feature

extraction or dimensionality reduction, they are called **RAW** faces. Each individual is in both the smiling and neutral expression set.

Figure 4.1: Example images from the FERET dataset used for the experiment. The top row shows Neutral Images and bottom row shows smiling faces. This dataset includes various race, gender and age; however they are not equally balanced. This is a balanced dataset in terms of Expression and gender.

The training set was 80 faces (with 20 female, 20 male and equal numbers of them with Neutral and Smiling expression). Two test sets were created each with 20 faces. Test set A had easily discernible smiling faces and Test set B had smiling faces that were not so easily discernible to the experimenter. In both test sets the number of each type of face is balanced. For example, there were 5 smiling male faces, 5 smiling female faces, 5 neutral female faces and 5 neutral male faces.

Table 4.1 details the dataset used in this experiment. The Test set A and Test set B were different individuals; however, each person had a smiling and neutral expression.

Table 4.1: Description of the dataset used from the FERET database: A total of 80 images for training, 20 images for Test set A and 20 for Test set B.

Size of Total dataset: 120 faces	**Female**	**Male**
Neutral	30 (Training set -20, Test set A -5, Test set B- 5)	30 (Training set -20, Test set A -5, Test set B- 5)
Smile	30 (Training set -20, Test set A -5, Test set B- 5)	30 (Training set -20, Test set A -5, Test set B- 5)

4.3 Experiment

This experiment was carried out to compare SVM classification on these six models:

- **RAW** : raw face images (64×64)
- **RAWPCA** : raw faces reduced in dimensionality with PCA
- **RAWCCA** : raw faces reduced in dimensionality by CCA
- **GAB** : Gabor pre-processed images (64×64)
- **GABPCA** : Gabor pre-processed images reduced by PCA
- **GABCCA** : Gabor pre-processed images reduced by CCA

4.3.1Gabor Filters

A total of 40 Gabor filters were designed at five scales and eight frequencies to produce 40 image outputs (magnitude) for each image of size 64×64 from the FERET dataset. The filter bank uses the *L2 max norm superposition* principle to produce one image of size 64×64 from the 40 Gabor filter bank outputs of the same size. Using 40 filters covers all the frequencies and scales required to extract the important features of the face (Shen and Bai, 2006). In Section 3.2.1 of Chapter 3 the exact process of how feature extraction was done using Gabor filters was explained in detail. The Figures 3.8(a) and 3.8(c) in Chapter 3 showed an example FERET image from the dataset and the image after feature extraction using *L2 max norm superposition* principle respectively. The dataset of 120 images (each of size 64×64) used in this experiment produces a total of $120 \times 64 \times 64 \times 40$ final images. However, by using the *L2 max norm superposition* principle, the final output size from the Gabor filter bank is same as that of the input image set ($120 \times 64 \times 64$).

4.3.2 Principal Component Analysis

For PCA reduction we use the first few principal components of the maximum 120 components, which account for 95% of the total variance of the data, and project the data onto these principal components. This resulted in using 66 components of the raw dataset and 35 components in the Gabor pre-processed dataset. Figure 4.2 shows the first 5 Eigenfaces of the total dataset.

Figure 4.2: The first 5 Eigenfaces (left to right) of the whole set of faces (male and female with equal number of smiling and neutral faces).

Figure 4.3 shows a projection of the test and training data into the first two PCA components. The difficulty of the classification problem is obvious.

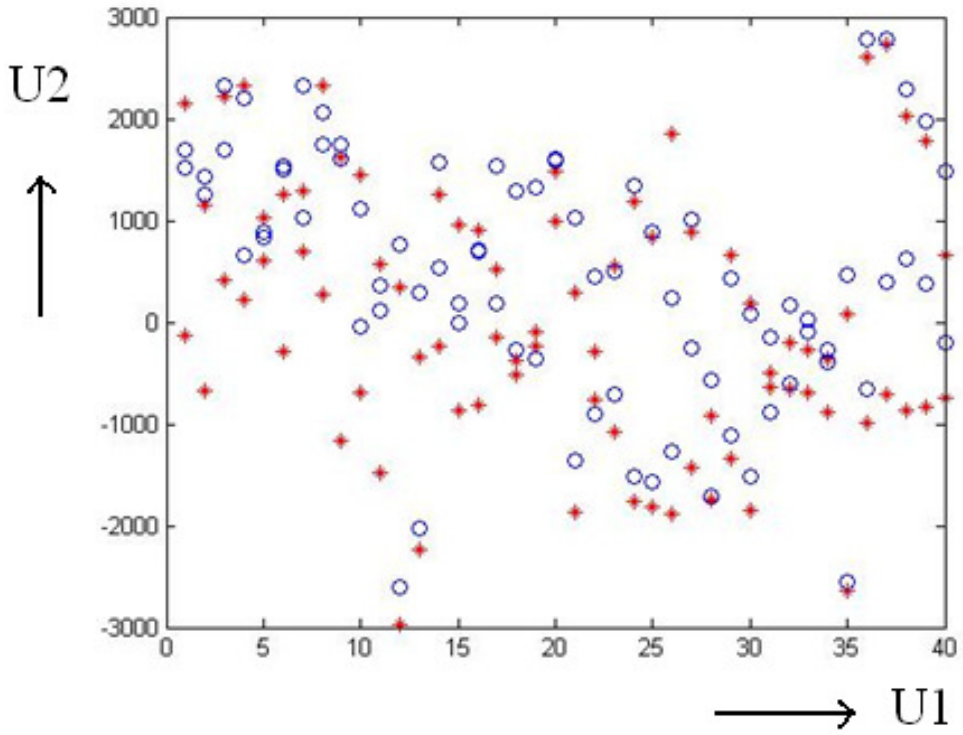


Figure 4.3: The PCA projection of the 120 examples from the dataset on a 2D plane. The red '*' and the blue 'o' represent the neutral and smiling data points respectively, after PCA projection of the training set. The PCA projection shows a very difficult classification problem and the results are reflective of this.

An important feature of PCA is that one can reconstruct any of the original images by combining the Eigenfaces. The original face image can be reconstructed from the Eigenfaces by adding up all the Eigenfaces (features) in the right proportion. The reconstructed original image is equal to a sum of all Eigenfaces, with each Eigenface having a certain weight. This weight corresponds to what degree the specific feature (Eigenface) is present in the original image. Figure 4.4 shows the original image on the left and the reconstructed images on the right. The reconstructed images use first 10, 25 and 66 (from left to right in the right column) Eigenfaces. The right most image of the reconstructed set uses 66 components and is much similar to the original face as compared to the left most reconstructed face which makes use of only 10 Eigenfaces. The steps involved in finding the PCA projection and the reconstruction of original images is detailed in Appendix A.

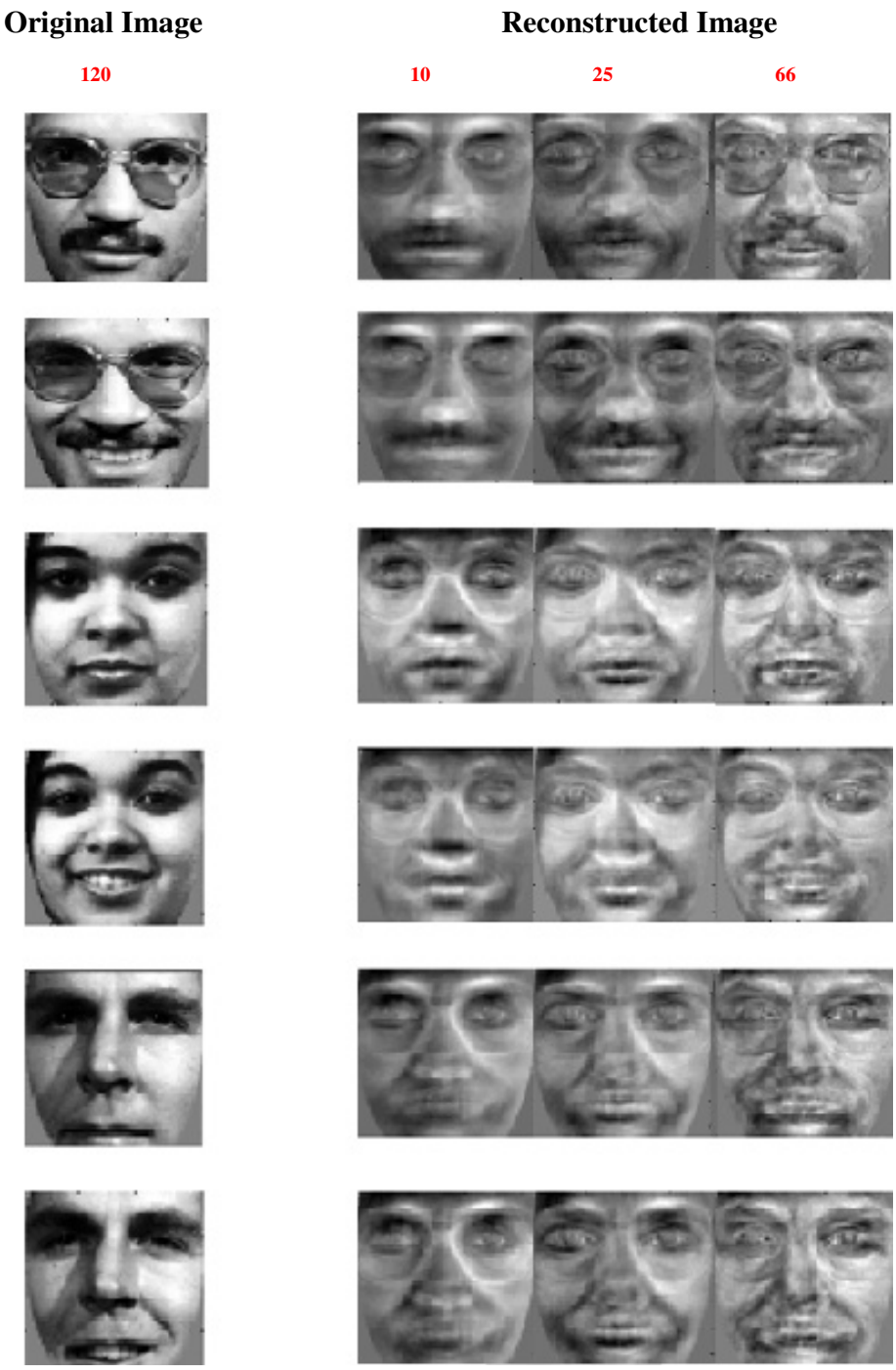


Figure 4.4: Figure showing original FERET face images on the left and the reconstructed images on the right. The reconstructed images use 10, 25 and 66 Eigenfaces (left to right) and the image on the extreme right is from just 66 Eigenfaces and is almost similar to original image. The left most image in the reconstructed set is least similar to the original and uses just 10 Eigenfaces for the reconstruction. In order to maintain 95% of the variance, 66 components need to be retained. The more principal components used, the more perfect reconstruction achieved.

4.3.3 Curvilinear Component Analysis

As described in Section 3.3.2 of Chapter 3, the problem with CCA is deciding how many dimensions the projected space should occupy, and one way of obtaining this is to use the *Intrinsic Dimension* of the data manifold. Figure 4.5 shows the $(dy - dx)$ plot of the CCA projection for the dataset and it shows that the smaller distances are well maintained and even at larger distances the scatter is low. The more dimensions used the better the graph with all distances almost on the $\boldsymbol{dy = dx}$ line.

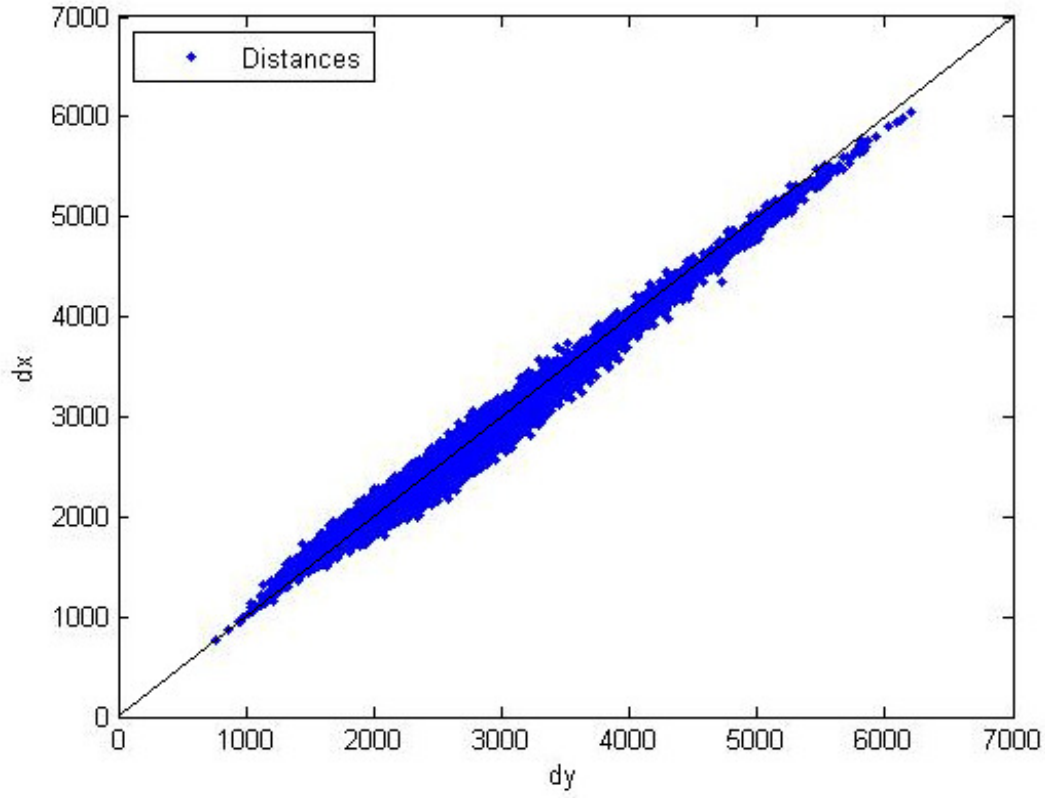


Figure 4.5: The $(\boldsymbol{dy - dx})$ plot of the CCA projection for the data set. If there is a good matching between input and output spaces and the data is linear, then all the distances would be on the line$(\boldsymbol{dy = dx})$. Here it shows that the data is non-linear in nature, however it has managed to do a very good projection as the original 4096 dimensions have been reduced to just 11 components.

4.3.4 Intrinsic Dimension

As described in Section 3.3.3 of Chapter 3, a plot of *log C(l)* against *log (l)* for the FERET dataset is shown in Figure 4.6. There are a number of non-linear and linear parts in the plot. Selecting the linear fit of the plot from the curve with the highest (maximum) slope, we obtain the Correlation Dimension. From the Figure 4.6, the largest slope is at the linear part marked with X and Y to correspond to the horizontal and vertical part of the slope.

When the Intrinsic Dimensionality technique is used, the CCA projected data is reduced to this Intrinsic Dimension. The Intrinsic Dimension of the CCA projection of raw faces was 14 and that of CCA projected Gabor pre- processed images was 11. These results are similar to what was obtained with experiments on Dimensionality Reduction for gender classification by Buchala et.al (2004b).

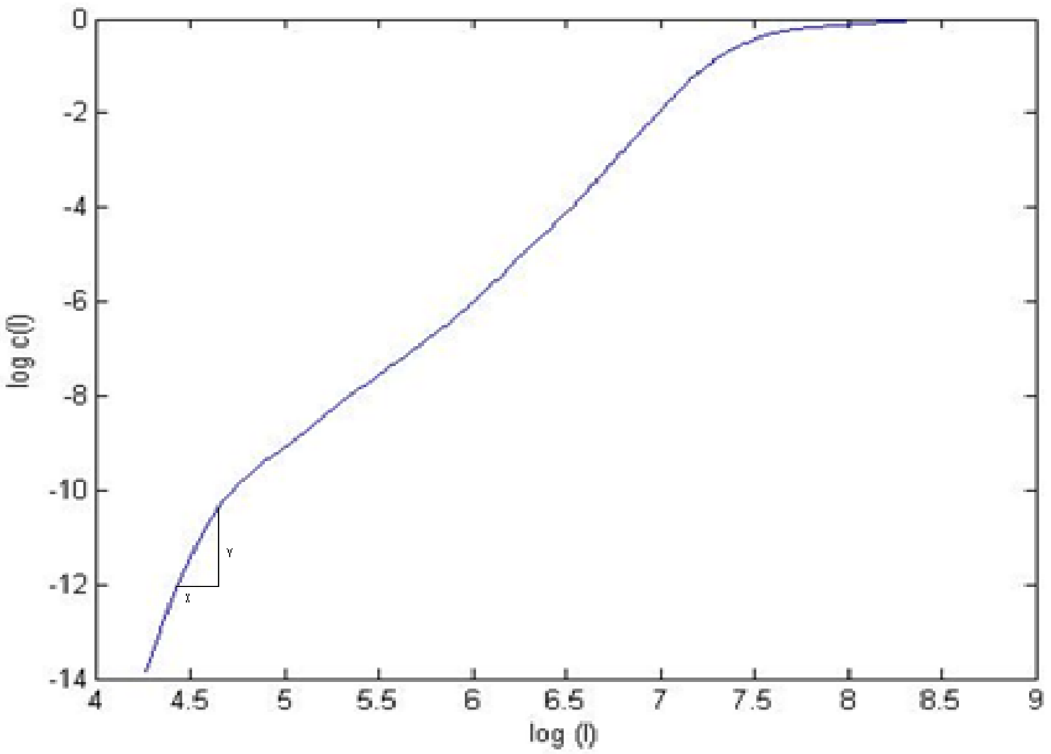


Figure 4.6: Correlation Dimension plot of Gabor filtered raw face images with CCA. The largest slope is in the most linear part of the graph and indicates the Intrinsic Dimension of the dataset and is the ratio of Y over X. In this case the maximum slope is estimated at 11.

4.3.5 Fisher Linear Discriminant Analysis and Classification

As described by Section 3.3.4 of Chapter 3, FLD can be used for classification purposes. The LDA projection of the dataset onto the fisher face was also shown in Figure 3.19 of Chapter 3. The LDA reduces the dataset to only one dimension. The two test sets namely, Test set A and Test set B were then classified by using the nearest neighbour in the test set in the projection space. The results are as in Table 4.2. The classification is best with only one misclassification with set A and five misclassifications with set B. The results with FLD are encouraging; however, the need to perform PCA before FLD for classification increases the computational complexity of the problem with high dimensional face images.

Table 4.2: Classification accuracy of raw faces using LDA

% Accuracy	Test set A	Test set B
LDA	19/20 (95%)	15/20 (75%)

4.4 Classification using Support Vector Machines

The dataset of 120 images included 80 images of the training set and 40 images of test set (Test set A- 20 images and Test set B – 20 images). An SVM was used for classification for all six models. The classification was performed as described in Section 3.4.3 of Chapter 3.

4.4.1 Classification Results

The Classification results for both the test sets used is shown in Table 4.3. The SVM classification results for both Test set A and Test set B show that the accuracy is good with raw faces and Gabor pre-processed images, but reduced with PCA. The raw faces are of size 64×64 (4096 dimensions) whereas the Gabor pre-processed image reduced with CCA has mere 11 components. The classification obtained with raw faces reduced by PCA, and Gabor pre-processed images reduced by PCA, was not as good in comparison to the rest of them.

Table 4.3: SVM Classification accuracy of raw faces and Gabor pre-processed images with PCA and CCA dimensionality reduction techniques.

SVM Results	Test set A	Test set B
RAW	19/20 (95%)	16/20 (80%)
RAWPCA66	18/20 (90%)	15/20 (75%)
RAWCCA14	18/20 (90%)	16/20 (80%)
GAB	19/20 (95%)	16/20 (80%)
GABPCA35	14/20 (70%)	12/20 (60%)
GABCCA11	19/20 (95%)	16/20 (80%)

The reason could be that the PCA, being a linear dimensionality reduction technique, might not have done quite as well as CCA. With CCA there was good generalization, but the key point to be noted here is the number of components used for the classification. The CCA makes use of just 14 components with raw faces and just 11 components with the Gabor pre-processed images to get good classification results, whereas the PCA used many components with lesser accuracy. This suggests that the Gabor filters are highlighting salient information which can be encoded in a small number of dimensions using CCA. Some examples of misclassifications are shown in Figure 4.7. The reason for these misclassifications is probably due to the relatively small size of training set. For example, the moustachioed face in the middle of the bottom row is misclassified as smiling. There are only four moustachioed faces (of two individuals) in the entire dataset. Although, a fivefold cross validation was done with the training set, no cross validation was done with both test sets.

Figure 4.7: Examples of the misclassified set of faces. The top row shows smiling faces wrongly classified as neutral. The bottom row shows neutral faces wrongly classified as smiling.

4.5 Discussion

In this chapter, all the computational techniques explained in Chapter 3 were implemented and results discussed. It should be noted that this data set is very small and all results are indicators only. The results show that further investigation of the classification of expressions using these techniques was justified. Identifying facial expressions is a challenging and interesting task. This experiment shows that identification from raw images can be performed very well. However, with a larger data set, it may be computationally intractable to use the raw images. It is therefore important to reduce the dimensionality of the data.

Performing classification using FLD was a trivial task and the result was very impressive. It is interesting to see the *Effect Size* for each pixel in the image. In other words which pixels discriminate most between smiling and neutral faces and the result of this analysis was shown earlier in Figure 3.7 of Chapter 3. The Creasing of the cheeks is diagnostic of smiling faces; teeth may also be an important indicator, though to a lesser extent.

A linear method such as PCA does not appear to be sufficiently tuneable to identify features that are relevant for facial expression characterization. Although the result of classification with FLD is impressive, for large datasets with face images, PCA needs to be done prior to the LDA. However, on performing Gabor pre-processing on the images and following it with the CCA,

there was good generalization in spite of the massive reduction in dimensionality. The most remarkable finding from the results of this experiment was that the facial expression can be identified with just 11 components found by CCA. The next step is to repeat the experiments with a larger dataset and with all the other expressions and compare them.

CHAPTER FIVE

Computational categorization of six prototypical human facial expressions

5.1 Introduction

Chapter 3 gave the necessary literature background for the computational methods that were used in dimensionality reduction and in feature extraction as a part of pre-processing of the FERET dataset; the experiments and results of which were discussed in Chapter 4. This chapter details the extension of the work explained in Chapter 4 with a larger dataset and with all six basic expressions (Ekman and Friesen, 1971). The BINGHAMTON BU - 3DFE database (Yin *et al.*, 2006) used here is a larger dataset with seven expressions namely: Happy, Angry, Fear, Sad, Surprise, Disgust and Neutral. All the experiments that were performed with the FERET dataset were repeated with this larger dataset and with all the expressions and the results are discussed in this chapter. The experiments were performed with a view to compare the human performance and the computational performance in facial expression classification. Hence, two sets of experiments were performed. One involved classification with computational models and the other involved human subjects. This chapter explains all the computational models that were tested. The human performance in classifying facial expressions is explained and discussed in detail in Chapter 6.

5.2 Dataset Description

The BINGHAMTON BU-3DFE dataset has 3D and 2D colour images of 100 subjects. Each subject, upon request, had performed the seven universal expressions: neutral, happiness, surprise, fear, sadness, disgust, and angry. The subject displayed the expression for a short period of time, during which four instant shots were taken, which captured four different degrees of the expression that ranged from low, middle, high and highest. The 2D images with the strongest expression were used in these experiments. It is a fairly large dataset consisting of 60% female and 40% male subjects, spanning a wide range of age groups and ethnic backgrounds including white, black, East Asian, Middle East Asian, Hispanic, Latino and others. The dataset used for the experiments is a balanced set in terms of gender, expression and includes all the ethnic groups mentioned above. The images in the original dataset have been validated by the individual participants and also by experts from the psychology department of the Binghamton University. The images in the dataset are already processed by cropping to show only the face

area to exclude any hair or clothing and are of size 256×256. To make images suitable for the experiments, these images had to be reduced to size 64×64 using an image editing tool named Irfanview (skiljan, 2009) and then cropped to size 63×63 in order to keep only the pure face region. The images were also converted into grey scale to help with the computational complexity.

The experiments were performed on a total of 616 face images (308 female and 308 male face images) of 88 individuals with seven basic expressions: happiness, angry, sadness, surprise, fear, disgust and neutral. Apart from neutral all other expressions were selected with the highest degree of intensity for that expression. The classification was done between neutral and one of the expressions at a time. For example: the model classified a test face image as neutral or happy if the classifier was trained for neutral and happy face image classification. Considering one of the six basic expressions (say for example angry) along with neutral, the dataset of 176 images (88 images of angry and 88 images of neutral set) was divided into 4 equal subsets of 44 images, balanced in terms of gender and expression. The SVM classifier was then trained with 3 subsets at a time and the left out set was used as the test set. A total of 22 male and 22 female face images was used in each set and was balanced, i.e., a person pictured in the neutral set was also present in the angry expression set. Hence at any time, the training set had 132 images. The accuracy was obtained by calculating the average of the classification accuracy for all four subsets used as test sets (when three subsets were used for training, the left out set was used as test set). Figure 5.1 shows examples of face images of four individuals. Each row corresponds to the expressions of one of the subjects. They are displayed from left to right in the order: neutral, happy, angry, fear, sad, surprise and disgust.

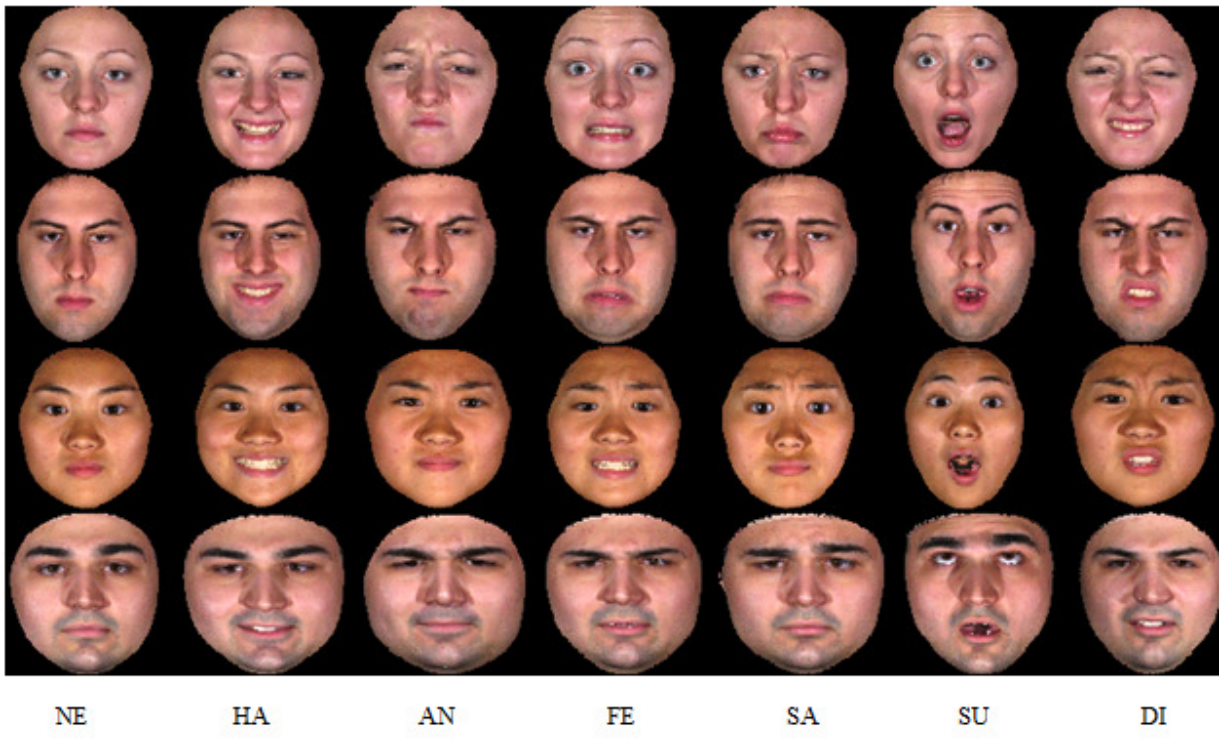


Figure 5.1: Examples face images from the BINGHAMTON BU-3DFE dataset. Each row is a subject showing various expression (left to right) neutral (NE), happy (HA), angry (AN), fear (FE), sad (SA), surprise (SU) and disgust (DI).

5.3 Experiments

A total of six experiments were performed with six computational models. Each experiment involved two expressions: one of them was neutral and the other was one of the six basic expressions.

The six models that were tested are:

- **RAW**: - Raw face images without any pre-processing or dimensionality reduction
- **RAWPCA**: - Raw face images without any pre-processing but reduced in dimensionality with PCA.
- **RAWCCA**: - Raw face images without any pre-processing but reduced in dimensionality with CCA.
- **GAB**: - Gabor pre-processed face images with no dimensionality reduction.
- **GABPCA**: - Gabor pre-processed face images reduced by PCA.
- **GABCCA**: - Gabor pre-processed face images reduced by CCA.

5.3.1Gabor Filtering

The pre-processing was done for feature extraction as with the FERET dataset and has been explained in Section 4.3 of Chapter 4. It used 40 filters at 5 scales and 8 directions and *L2 max norm superposition* principle to obtain the output from the filter bank.

5.3.2 Principal Component Analysis

Using PCA for the neutral and one of the expressions, in order to retain 95% of the total variance of that set, the number of components to which the PCA reduced the original data is detailed in the Table 5.1.

Table 5.1: Comparison of number of components used with PCA for raw and Gabor pre-processed face images for all expressions.

Number of components Reduced by PCA	Raw face images	Gabor pre-processed face images
Angry	97	22
Happy	100	23
Fear	99	23
Sad	96	22
Surprise	103	23
Disgust	101	23

5.3.3 Curvilinear Component Analysis

As discussed in Chapter 3 and Chapter 4, for CCA, the data was reduced to its Intrinsic Dimension. The Intrinsic Dimension of the raw faces images and Gabor pre-processed face images with neutral and one of the other basic expressions is detailed in Table 5.2. A wonderful reduction in dimensionality can be achieved using CCA. The best is just 5 components required for almost all of the Gabor pre-processed face images.

Table 5.2: Comparison of number of components used with CCA for raw and Gabor pre-processed face images for all expressions

Number of components Reduced by CCA	Raw face images	Gabor pre-processed face images
Angry	5	6
Happy	6	5
Fear	6	5
Sad	7	5
Surprise	6	5
Disgust	5	5

5.3.4 Fisher Linear Discriminant Analysis and Classification

As discussed in Chapter 3, FLD is performed to classify two classes, for example: happy from neutral. Here, each of the six basic expressions was classified against neutral. The Table 5.3 shows the classification accuracy obtained with FLD based on the Euclidean distance measure, for all six expressions. Each test set was tested separately and an average taken.

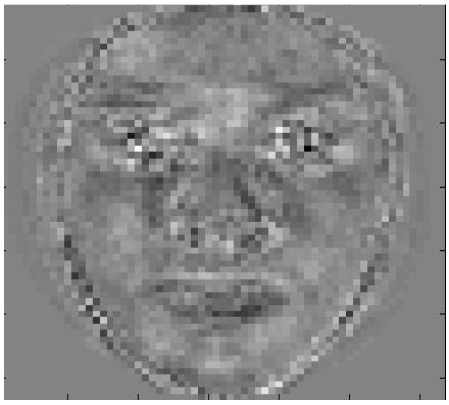
Table 5.3: FLD classification accuracy of raw faces

% Accuracy	LDA (Out Of/176)
Angry	104/176 (59%)
Happy	114/176 (65%)
Fear	106/176 (60%)
Sad	105/176 (60%)
Surprise	122/176 (69%)
Disgust	112/176 (64%)
Average	63%

The classification results have not been very encouraging. The best classification accuracy was with surprise and happy face images and the least classification accuracy was with sad, angry and fear face images; disgust being intermediate. The table which details the classification accuracy of each of the individual subsets is in Appendix C.

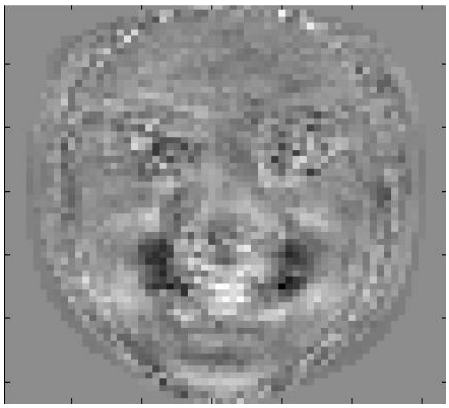
Interestingly, the psychological data shows that humans perform best on surprise and happy face expressions and least well with sad, anger and fear and is discussed in Chapter 6.

The projection of the dataset can be viewed as an image which is the *Fisher face*. Figure 5.2 shows the fisher face with respect to six basic expressions used. Each unique fisher face is the template reflective of the expressions it is associated with.

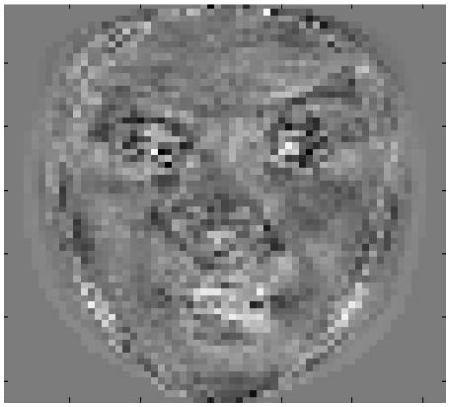


(a)

(d)

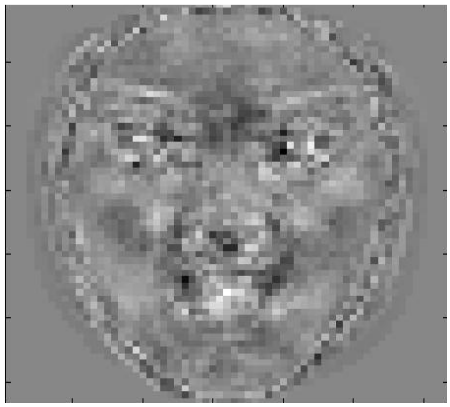


(b)



(e)

(c)



(f)

Figure 5.2: Figure shows fisher faces a) angry b) happy c) fear d) sad e) surprise f) disgust

5.3.4.1 Encoding power

With the PCA of two classes, the first components encode information common to both classes of faces, whilst the latter components encode information not so common between the two classes. The FLD can be used to estimate the encoding power of the various face properties such as expression, gender, age, identity and race. The LDA of faces also provides us with a small set of features that carry the most relevant information for the purpose of classification based on a property. The features are obtained through eigenvector analysis of scatter matrices with the objective of maximizing between-class variations and minimizing within-class variations. This was explained earlier in detail in Section 3.3.5 of Chapter 3. The experiments here were performed between two classes: one basic prototypical expression and the other neutral. One might suggest that all the early components could carry high expression information and by estimating the expression encoding power by FLD this can be decided. The expression encoding power of different components was estimated by Equation 3.14 and explained in Section 3.3.5 of Chapter 3. The expression encoding power of the components help to understand which of the components are important for each expression. It may be that some of the first few components are amongst the most significant when compared to the later ones or some of the initial ones may not be diagnostic for expression and may be important for other properties such as race, age, gender and identity (Buchala *et al.*, 2004c; Calder *et al.*, 2001; Belhumeur *et al.*, 1997; Kulikowski *et al.*, 1982). Every expression may have a different component as the most significant. Table 5.4 shows the most significant and the next most significant components for a particular expression. Note that a component which is the most significant for an expression may also be important for other expressions too. The plots of the discriminating power of the first components for all the expressions can be found in Appendix B and suggests that not all the first components are significant for expression encoding but the combination of first and second highest components are unique.

Table 5.4: Significant components for all expressions

Expression	First highest component	Second highest component	Magnitude of the highest component
Angry	26	3	0.16
Happy	7	6	0.35
Fear	7	14	0.20
Sad	26	14	0.10
Surprise	3	2	0.80
Disgust	26	13	0.18

It can be seen that 26th component is significant for angry, sad and disgust expression, 7th for happy and fear and 3rd for surprise. The plots also suggest that though all components important for expression are amongst the initial components, some of these components are not specifically diagnostic for the expression in question.

In comparing the magnitudes of these components with respect to each expression, they have the encoding power in order (highest to lowest) for surprise, happy, fear, disgust, angry and sad. This means the magnitude of the encoding power for expression surprise is highest and for sad is the least as can be seen in the last column of Table 5.4.

5.3.5 Effect Size

The Section 3.3.5 of Chapter 3 detailed how the effect size emphasizes the difference between the two classes. Here, the encoding face was obtained by applying the effect size to the pixels of the face image. Two classes were considered at a time: one of the basic expressions alongside the neutral expression. The discriminating pixels for different expressions are different. This result supports the evidence of variations in the facial appearance and movements of the facial muscle in response to the expression and in particular, emphasizes those parts of the face corresponding to each of the basic expression (Yacoob and Davis, 1994). The coloured images are shown on the right as they are clearer than their grey scales on the left. The research literature results are in the description given first followed by a comparison of these with the computational model.

Angry encoding face: Figure 5.3 shows the angry encoding face. The encoding face shows which pixels of the face discriminate most between the angry and neutral classes. Note the changes in the forehead, above and in between the eyebrows and changes in the lip and mouth area. Lowered eyebrows, which may be pulled together forming wrinkles in the skin of the forehead, tension in lips and mouth, all characterize the anger expression. Also, some people have their lowered eyelids tensed and the eyebrows pulled down and may have a glaring look. Others who have a closed mouth form of the angry expression will have a pushing up of the chin (Hager, 2006; Ekman and Friesen, 1975). All these areas described are indeed the parts of the angry encoding face that are highlighted showing that the computational model is emphasizing the same areas.

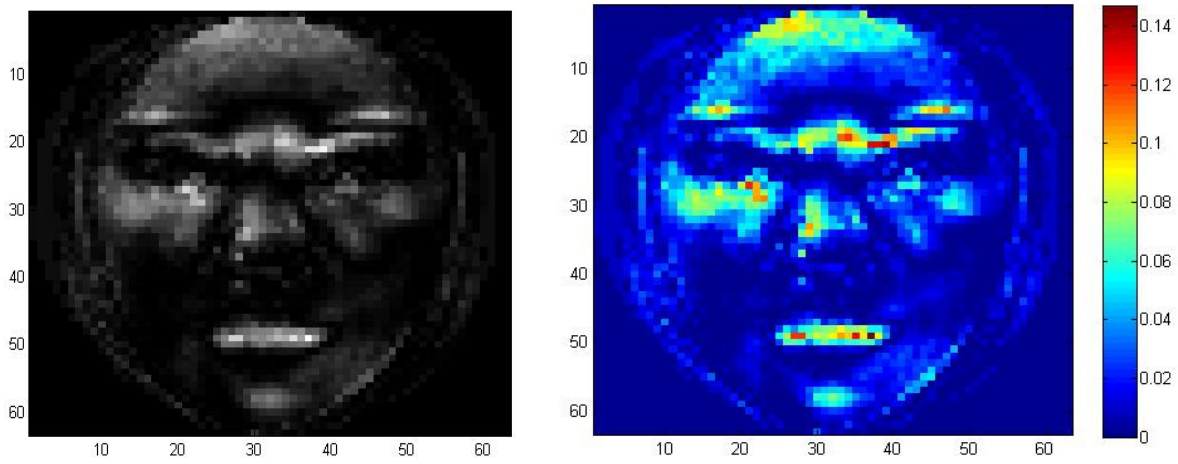


Figure 5.3: Angry encoding face

Happy encoding face: Figure 5.4 shows the happy encoding face. Note the changes in the cheeks and the lips. A happy face is normally recognizable with the smile. There is also normally an oblique raising of the lip corners and a wrinkling and creasing of the cheeks. These are defined as the characteristics of the happy expression (Hager, 2006; Ekman and Friesen, 1975). In addition to these there is a narrowing of the eyelids, crowfeet wrinkling at the corners of the eye and a raising of the upper areas of the cheeks indicating actual happiness. It may well be that since the dataset that is used here are posed expressions and are not spontaneous expressions; these areas are not very well highlighted.

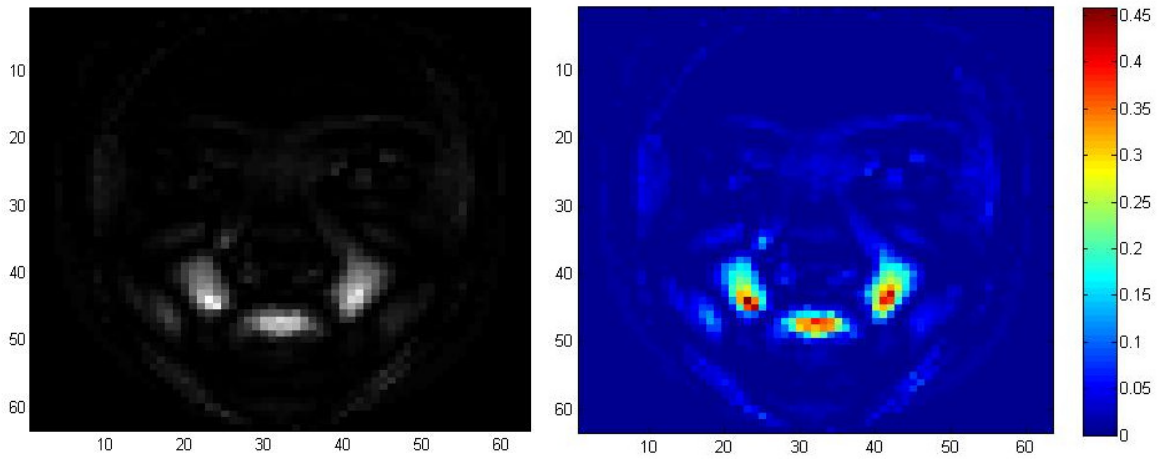


Figure 5.4: Happy encoding face

Fear encoding face: Figure 5.5 shows the fear encoding face. Note the changes in around the mouth, eyebrows, and eyelids. Normally, the fear expression shows raised upper eyelids, tensed lower eyelids, eyebrows pulled up, mouth open and jaw dropped. Sometimes, fear expressions are blended with surprise and may also cause a lateral pull on the corners of the lips causing it to stretch (Hager, 2006; Ekman and Friesen, 1975). These details match very well with the pixels highlighted for the fear expression.

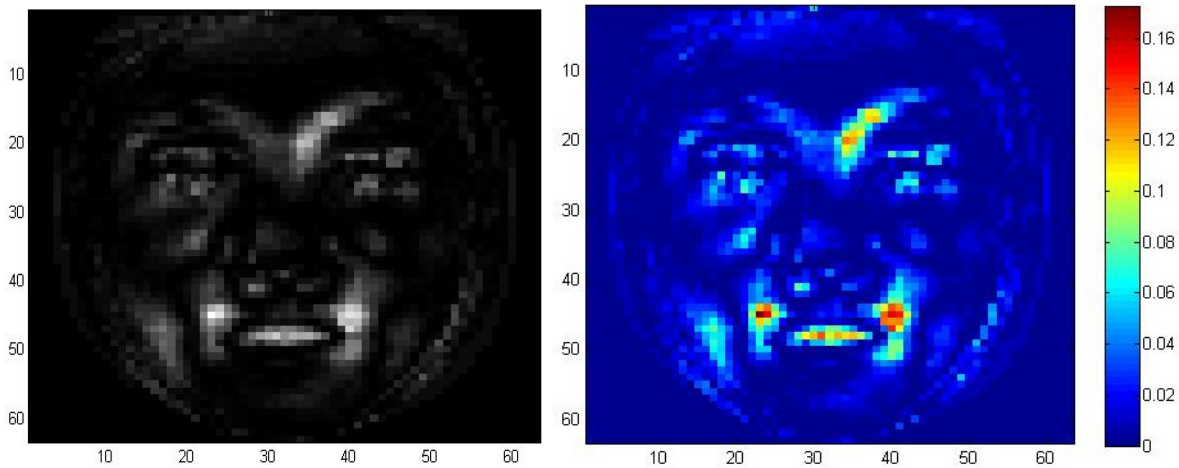


Figure 5.5: Fear encoding face

Sad encoding face: Figure 5.6 shows the sad encoding face. Note the changes in the space between the eyebrows, chin and the corners of the lips. The normal characteristics of a sad face would show narrowing of the eyes and raised cheeks, eyebrows pulled together and raised in the

centre of the forehead forming wrinkles. There is also the pushing up of the chin. Sometimes, there may be a lateral lip stretching, with a downturn lip corners and/or may have no raising of the eyebrows. The research literature descriptions of the sad expression match the highlighted areas of the encoding face very well (Hager, 2006; Ekman and Friesen, 1975).

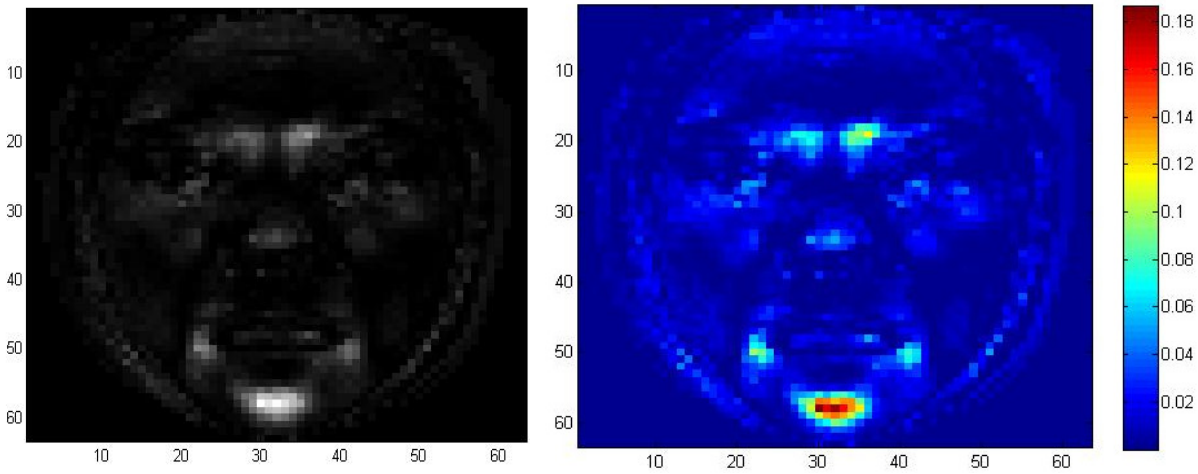


Figure 5.6: Sad encoding face

Surprise encoding face: Figure 5.7 shows the surprise encoding face. Note the changes in the overall shape of the face around the sides, the lines in the forehead, and mouth. A genuine surprise expression is characterized by slight raised eyebrows; horizontal wrinkles on the forehead, mouth opened by the jaw drop and relaxed lips. There may be a slight smile as well. Too much exaggeration could cause great amount of jaw drop with a very tense mouth opening (Hager, 2006; Ekman and Friesen, 1975). These variations are seen to some extent on the encoding face; however, as these are not genuine expressions, there may be some exaggerations.

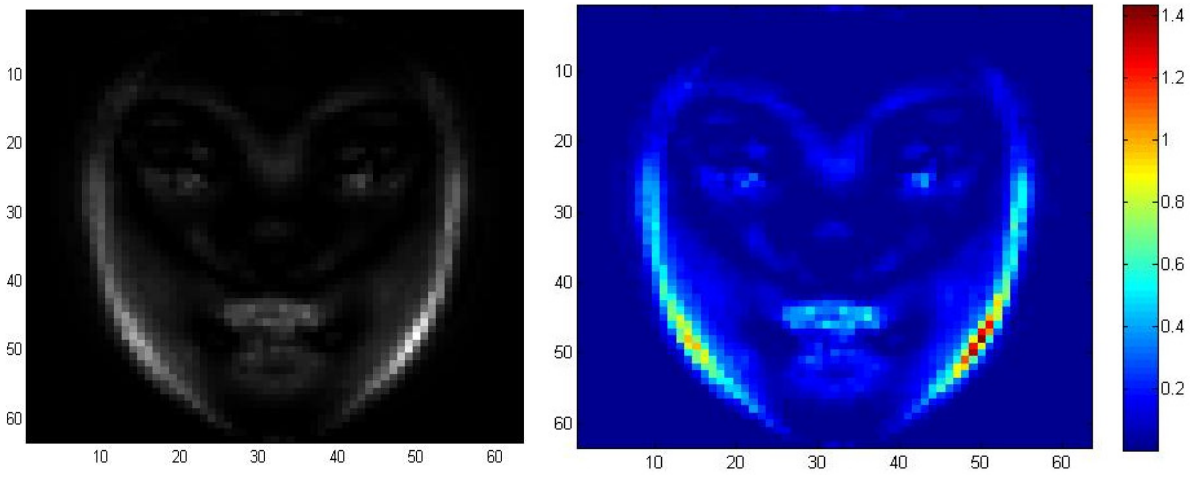


Figure 5.7: Surprise encoding face

Disgust encoding face: Figure 5.8 shows the disgust encoding face. Note the changes in the lower eyes, space between the eyebrows, forehead, nose, area around the nose and the mouth. A wrinkled nose with eyebrows pulled down and the upper lip drawn up, lower eyelid is tensed and the eye opening narrowed. In addition, the upper eyelids are normally relaxed and mouth would be open (Ekman and Friesen, 1975; Hager, 2006). These changes match with the changes highlighted in the disgust encoding face.

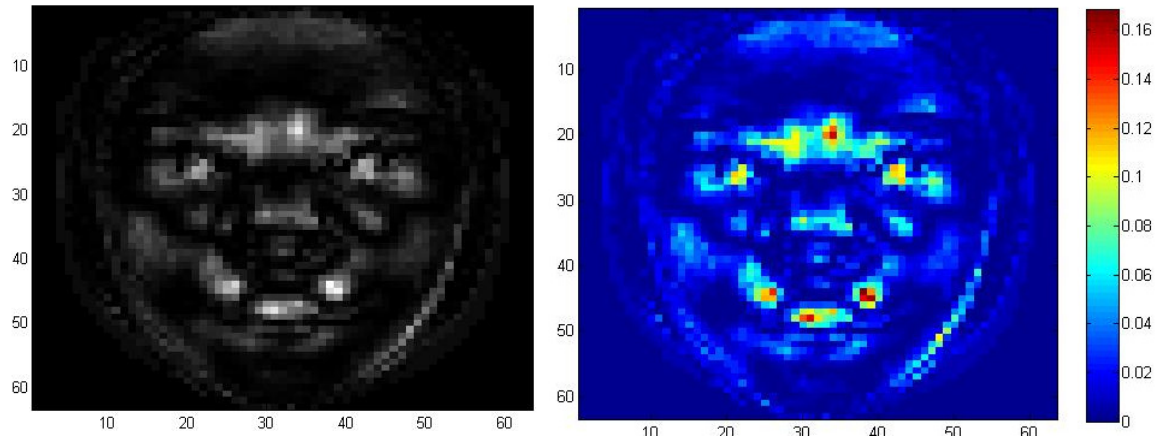


Figure 5.8: Disgust encoding face

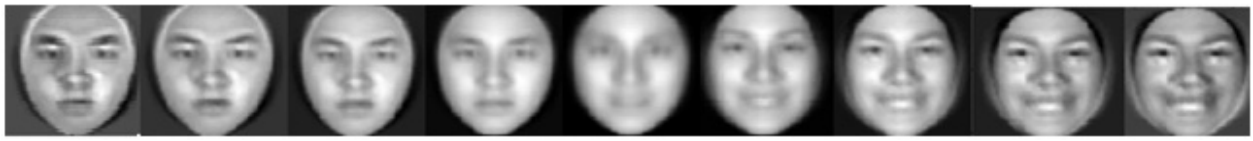
5.4 Morphing facial expressions using PCA

Upon performing LDA on the PCA projected data, one can find the important components of the expression. It is also clear from the previous section that a very few components encode information relevant to the property-expression. It would be interesting to see if the facial expressions over the faces can be morphed by extracting those components that are significant for that particular expression. Earlier work by Calder et al (2001) suggest that PCA can code facial expressions. In order to determine the information encoded in these significant components (for example - 26^{th} for angry with neutral), which are important for that specific expression, a series of reconstructed images were generated using the corresponding Eigenfaces. The original dataset was subjected to PCA and Eigenvectors (Eigenfaces for images) obtained were used for dimensionality reduction and used to project into the PCA space. The PCA projected data was used to obtain the scatter matrices (within and between classes) from which the encoding power of the components was found. A series of reconstructed images were obtained by altering the components in the following steps. The mean face is used for the reconstructions. Then additionally, the altered value of the 26^{th} component was used to reconstruct the faces along with the mean face. This was done progressively by adding or subtracting greater quantities of Eigenface 26 to the mean face in order to capture the effects for the angry expression. It was found that the 26^{th} component is also the most significant component for sad and also, for the disgust expression. The 7^{th} component is the most significant for expression happy and also, fear. The 3^{rd} component is the most significant for surprise. The image reconstruction was performed with the first and second significant components by repeating the steps.

Figure 5.9 shows these progressive changes over the prototype face. In Figure 5.9(a), the two classes used were angry and neutral; the middle face is the mean face. To the right of the mean (prototype) face are the reconstructions obtained by using the mean and subtracting 2 S.D of the 26^{th} component (the S.D was taken for the 26^{th} component of the entire dataset). Likewise, the reconstructions on the left were obtained by adding instead of subtracting. The similar procedure was adopted for the reconstructions in Figure 5.9 (b) and (c) but, with 7^{th} and 3^{rd} component respectively.

From the author's perception, the images are ordered in the obvious ever increasing featural changes in the expression. The images on the right from the prototypical image in the centre of Figure 5.9 (a) show obvious featural changes for the disgust expression. Figure 5.9 (b) show increasing changes for the happy face expression in the right. Figure 5.9 (c) show increasing featural changes for the surprise expression on the right of the prototypical image which is in the centre.

a) Component 26 (First component for ANGER, SAD, DISGUST)

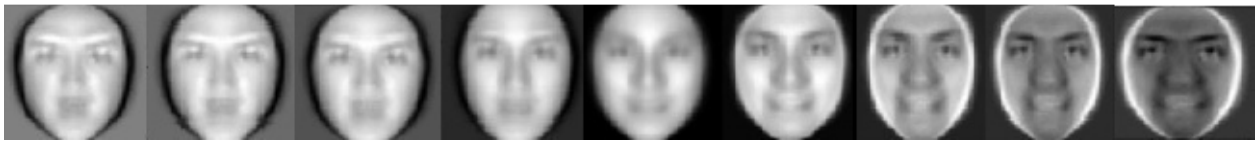


b) Component 7 (First component for HAPPY, FEAR)

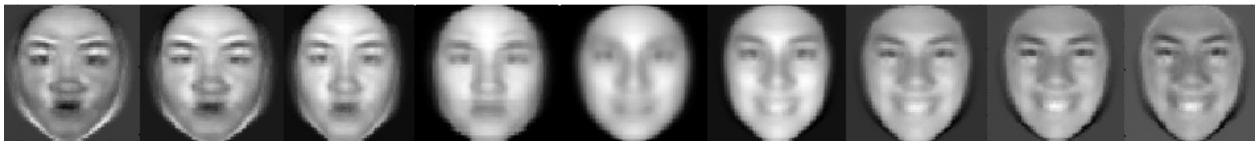
c) Component 3 (First component for SURPRISE, Second component for ANGER)

Figure 5.9: Reconstructed images using the altered components (a) 26th component – This is the first highest component for angry expression. It is also the highest component for expression sad and disgust against the neutral class (b) 7th component – It is the first highest component for happy and also for the expression fear (c) 3rd component - It is the first highest component for surprise and second highest for angry against neutral. The middle faces are the prototype face. The other faces were reconstructed by using the average face (obtained from the entire dataset - all expressions and the neutral face images) and adding the altered values of the respective component. Altering was done progressively by adding quantities of - 2S.D (right of the prototype) and + 2 S.D (left of the prototype face) of the 26th, 7th, 3rd to the prototype face. The reconstructions were obtained by altering 2 S.D, 4 S.D, 6 S.D and 10 S.D. Hence, for all sequences, the images shown here on the extreme left correspond to the average face altered by + 10 S.D and on the extreme right by -10 S.D. The images in between correspond to + 6 S.D, + 4 S.D, + 2 S.D, Average face, -2 S. D, - 4 S.D and - 6 S.D.

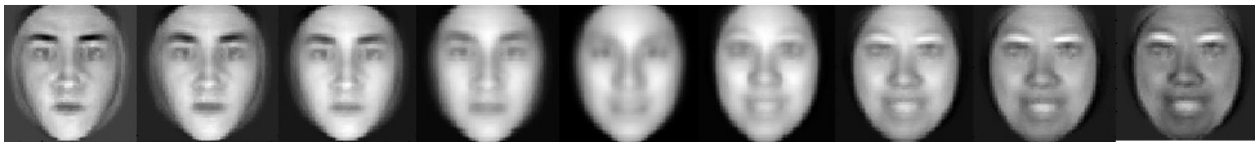
Figure 5.10 shows the reconstructions using the second significant component. Altering was done progressively by adding quantities of – 2 S.D and + 2 S.D of the 2nd, 6th, 14th and 13th component to the prototype face.



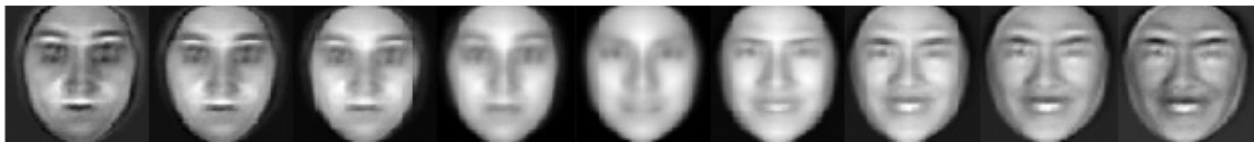
a) Component 2 (Second component for SURPRISE)



b) Component 6 (Second component for HAPPY)



c) Component 14 (Second component for FEAR, SAD)



d) Component 13 (Second component for DISGUST)

Figure 5.10: (a) 2nd component- second highest for surprise against neutral (b) 6th component- second highest for happy against neutral (c) 14th component- second highest for fear and sad against neutral (d) 13th component- second highest for disgust against neutral. The middle faces are the prototype faces (the mean face). The other faces are reconstructed by using the significant component and adding the altered values of the S.D of the respective component. Altering is done progressively by adding quantities of -2S.D and + 2 S.D of the 2nd, 6th, 14th and 13th component's mean to the prototype face and is shown in 5.10 (a), (b), (c) and (d) respectively. Figure 5.10 (a) and (d) has images on the extreme left which is altered by + 10 S.D and on the extreme right by – 10 S.D; The images in between correspond to + 6 S.D, + 4 S.D, + 2 S.D, average face, – 2 S.D, – 4 S.D and –6 S.D. Figure 5.10 (b) and (c) has images on the extreme left which is altered by – 10 S.D and on the right by + 10 S.D. The images in between correspond to – 6 S.D, – 4 S.D, – 2 S.D, average face, + 2 S.D, +4 S.D and +6 S.D.

All images are arranged such that the expression becoming ever increasing prominent is on the extreme right with the prototype face in the middle. Hence, Figure 5.10 (a) and (d) has images on the extreme left which is altered by + 10 S.D and on the extreme right by – 10 S.D; The images in between correspond to + 6 S.D, + 4 S.D, + 2 S.D, average face, – 2 S.D, – 4 S.D and –6 S.D.

Figure 5.10 (b) and (c) has images on the extreme left which is altered by – 10 S.D and on the right by + 10 S.D. The images in between correspond to – 6 S.D, – 4 S.D, – 2 S.D, average face, + 2 S.D, +4 S.D and +6 S.D.

Expressions angry, sad and disgust all have component 26 as the first significant component. Zucker et al (2007) experimented with human subjects to perform a six-way forced choice classification and found that angry expressions are very often confused with disgust and sometimes with expression sad. Expression disgust is often confused with angry. Expression sad is often confused with angry. This could suggest that from the results obtained here, the expressions angry, sad and disgust are encoded by the same component (26^{th}) and hence be some supporting evidence to these expressions being indeed confusing.

Susskind et al (2007) performed multidimensional analysis of human performance in similarity judgements of facial expressions. They found that by ordering the emotion clusters, angry exemplars were ordered between sad and disgust, surprise was between happy and fear, with expression sad at a large distance away from happy. These compliment the previous explanations for the confusion between expression angry, disgust and sad. Dailey et al (2002) have presented a multidimensional scaling (MDS) model of human response which reveals the dimensions of the emotions. The clusters for angry and disgust seemed to overlap, surprise was between happy and fear, and sad was close to angry and also positioned in between angry and fear. They suggest that humans find that fear expressions are difficult to classify and that they are often confused as surprise and never confused with happy (Calder *et al.*, 2001). Zucker et al (2007) also found fear expressions are confused with expression surprise. However, here component 7 is the first significant component for expressions happy and fear.

5.5 Comparison of dimensions used with PCA and CCA

The dimensionality reduction achieved by PCA on raw face images and Gabor pre-processed images for all expressions was detailed in Table 5.1 and the estimated intrinsic dimensions to which the CCA was reduced for raw and Gabor pre-processed face images was in Table 5.2. The important point to be noted here is that with PCA, the raw faces images for all expressions need at least 96 components in order to retain 95% of the total variance of the dataset. However, on performing Gabor filtering on the raw face images and then using PCA requires a mere 22 components to retain 95% variance without much significant information loss. This could be because of PCA being a linear dimensionality reduction method and Gabor filtering is a non-linear method highlighting expressive features of the face such as eyebrows or corners of the mouth which are involved while displaying any expression (Shen and Bai, 2006). The

explanation also holds good for results in Table 5.2 which indicate that the intrinsic dimensions estimated for the non-linear CCA (Demartines and Hérault, 1997b) does not make much difference with respect to raw and Gabor-preprocessed face image; as the facial features (Jarudi and Sinha, 2003) and Gabor filtering are both non-linear (Kruizinga and Petkov, 1999; Shen and Bai, 2006).

5.6 Comparison of classification results: FLD with PCA

Table 5.5 shows the results of classification by FLD in comparison to nearest neighbour classification of PCA projected raw face images and Figure 5.11 shows the plot for the same. When the training set is small, PCA can outperform FLD. When the number of samples is large and representative for each class, LDA outperforms PCA. With this dataset, the classification based on fisher faces yielded results that are just above average for all expressions. The average classification accuracy of 60% for PCA and 63% for FLD was obtained. It should also be noted that in order to obtain the FLD, the PCA is a prerequisite to overcome the problem with singular matrices and technically requires more processing.

Table 5.5: Comparison of classification accuracy of FLD and PCA

% Accuracy	FLD	PCA
Angry	104/176 (59%)	112/176 (64%)
Happy	114/176 (65%)	112/176 (64%)
Fear	106/176 (60%)	104/176 (59%)
Sad	105/176 (60%)	95/176 (54%)
Surprise	122/176 (69%)	102/176 (58%)
Disgust	112/176 (64%)	99/176 (56%)
Average	63%	60%

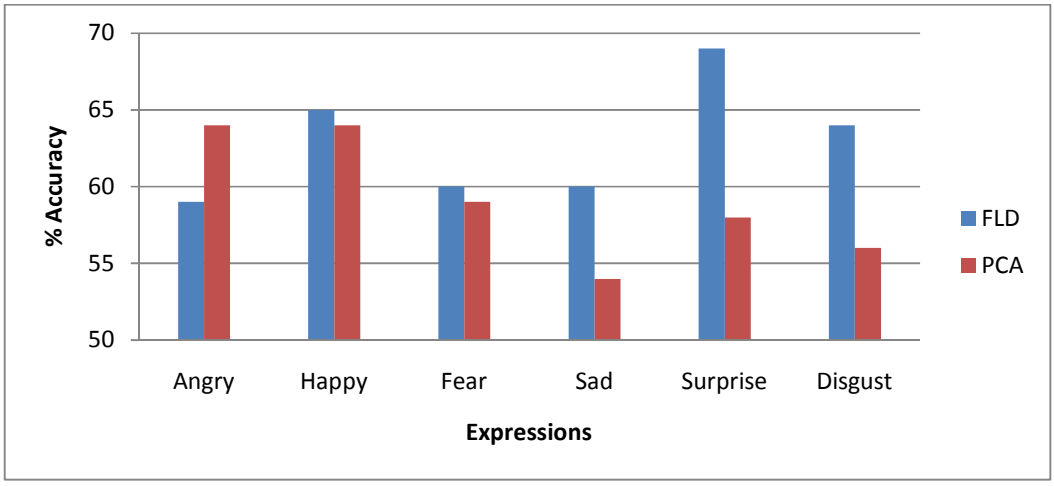


Figure 5.11: Classification accuracy of PCA and FLD for all expressions

Figure 5.11 charts the classification accuracy across various expressions for PCA and FLD on raw face images. As mentioned earlier in Chapter 3, PCA is commonly used for dimensionality reduction. It takes into consideration the greatest variance of the projected data; however, such projections may not be effective for classification since large and unwanted variations may be still be retained. On the other hand the LDA finds the projection such that there is a large between class scatter and little within class scatter. The steps necessary to perform the LDA needs PCA as pre-requisite. This overcomes the problem of singular matrices which often occurs with small number of data points in comparison with the large dimensions of the raw face images. Here, the LDA based classification is compared with the PCA in a similar manner to others (Belhumeur *et al.*, 1997; Kwak and Pedrycz, 2005).

Belhuemer and Kriegman (1997) developed a face recognition algorithm which is insensitive to gross variation in lighting direction and facial expression using Harvard and Yale databases. Here, they made a comparison in the performance of FLD with PCA for recognizing faces of two classes: one with variation in lighting intensities and the other with variations in the expressions, eye wear and lighting. They also performed a comparison on the classification of face images with/without glasses. They showed a comparatively better performance with FLD in comparison to PCA, and based on this they suggested classifications of facial expressions could have similar results, where the set of training images is divided into classes based on the facial expressions. PCA can significantly reduce the dimensionality of the original features without loss of much information in the sense of representation, but it may lose important information for discrimination between different classes (Deng H. B., 2005) and the accuracy may change with the size of the dataset. A frequently cited paper by Martinez and Kak (2001) used PCA and LDA for face recognition. By using varying sizes of dataset, they concluded that PCA might outperform LDA when the number of samples per class is small. They also report that several of their experiments have shown the superiority of PCA over LDA, while others show the

superiority of LDA over PCA indicating the classification accuracy depends on the classifier and the size of the dataset used.

5.7 Classification with Support Vector Machines

SVM based classification method has been described in detail in section 3.4 of Chapter 3 and again with reference to FERET face image classification in section 4.4 of Chapter 4. SVM classification was performed by using a 5 fold cross validation on each of the four subsets (described earlier in section 5.2) and the average accuracy is calculated. The individual tables pertaining to each of the expressions and for all the models are in Appendix C. All the raw faces and Gabor pre-processed face images have a dimension of size 3969 (63×63); whereas the PCA and CCA dimensionality reductions have lesser dimensions; the details of which are in Table 5.2 and 5.3.

5.7.1 Comparison of classification accuracy – by Models

Table 5.6 shows the accuracy obtained for each expression and also the average accuracy of each model across all the expressions; for example - the average accuracy for RAW models of all expressions is considered. Figure 5.12 plots the average classification results detailed in Table 5.6.

Table 5.6: Average SVM classification accuracy for all models across all basic expressions

% Accuracy	Angry	Happy	Fear	Sad	Surprise	Disgust	Average
RAW	84.09%	99.43%	83.52%	77.27%	94.89%	90.34%	88.26%
RAWPCA	70.45%	89%	82.39%	74.43%	89.20%	80%	80.91%
RAWCCA	63.64%	87.50%	73%	62.50%	93.75%	69.89%	75.05%
GAB	75.57%	89.77%	75.00%	70.45%	95.45%	73.30%	79.92%
GABPCA	72.16%	86.93%	79.55%	71.02%	90.34%	76.68%	79.45%
GABCCA	66.48%	61.36%	55%	58.52%	84.09%	60.80%	64.38%

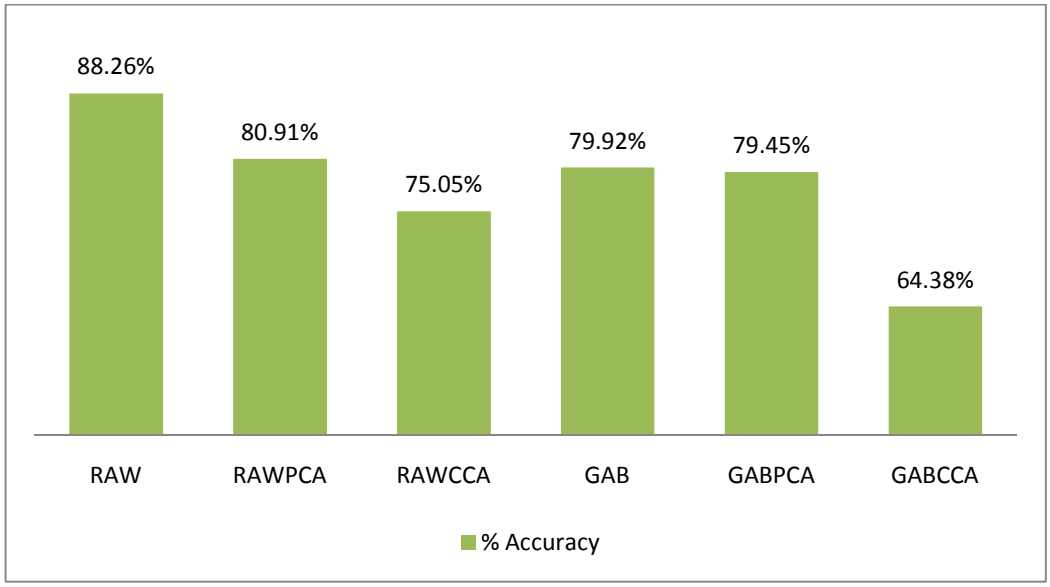


Figure 5.12: Average classification percentages (last column of Table 5.6) for each of the six models: RAW, RAWPCA, RAWCCA, GAB, GABPCA, GABCCA for all expressions

The average classification accuracy of the RAW model has an outstanding performance in comparison to the rest of them. The RAW model performs best with the happy dataset (99.43%) and the least with expression sad (77.27%). An average of 88.26% for the RAW model is the best in comparison with the other models; GABCCA being the worst (64.38%). The point here to be noted is that the RAW model did well as predicted due to the high dimensionality and no information loss, unlike other models that have undergone pre-processing (Gabor filtering) and dimensionality reduction (PCA or CCA). In all cases, the PCA reduces the dimensionality to between 96 and 103 (least for sad and maximum for surprise) whilst the CCA has the most reduction to a mere 5 components (for both angry and disgust).

Figure 5.13 charts the accuracy of classification for all models and for all expressions. The best classification accuracy of all the models across all the expressions is happy – RAW model (99.43%) and the worst of all is the fear – GABCCA model (55%) approximated to 5 components.

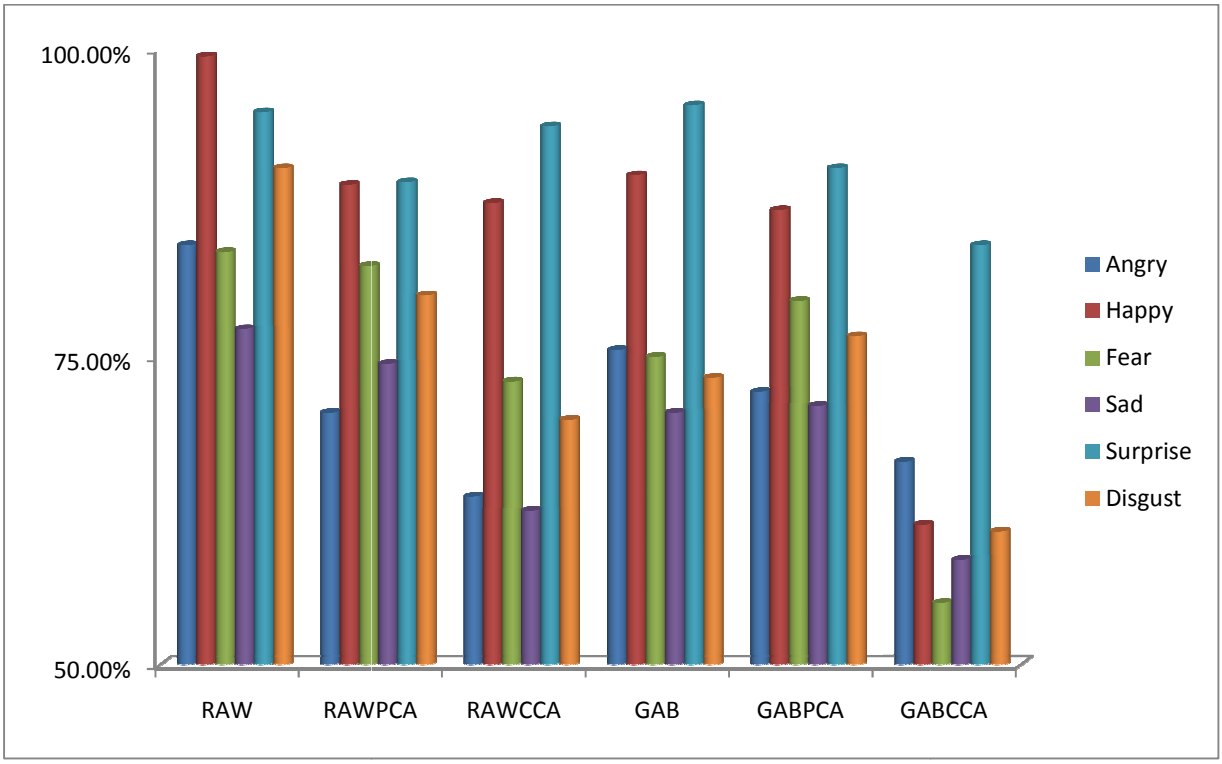


Figure 5.13: Classification accuracy of all models for all expressions – angry, happy, fear, sad, surprise, disgust

5.7.2 Comparison of classification accuracy – by Expression

The results are analyzed for all the expressions:

Angry expression: Figure 5.14 charts the results of angry expression for all models. The RAW model does very well with angry face images followed by GAB model at 84.09% and 75.57% respectively. On comparing, PCA and CCA, GABPCA does better than RAWPCA, and GABCCA does better than RAWCCA.

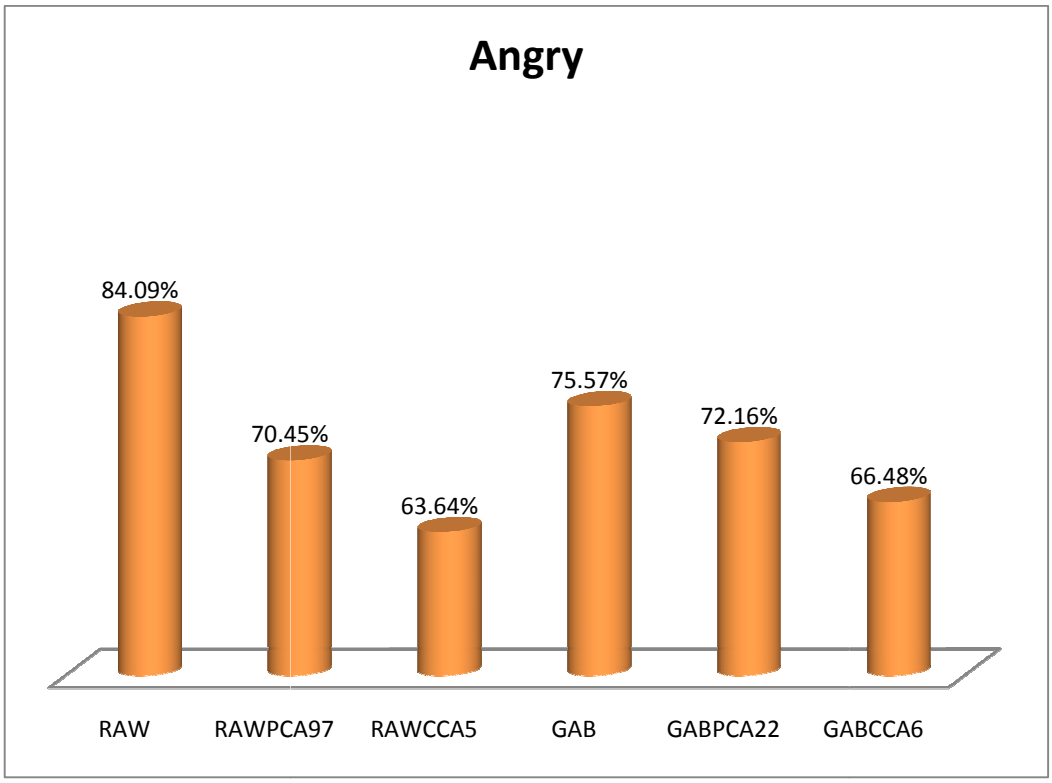


Figure 5.14: Classification accuracy of all models for – angry expression (RAW and GAB are the best)

Happy expression: Figure 5.15 charts the results of happy expression for all models. The RAW model does very well with happy face images followed by Gabor model at 99.43% and 89.77% respectively. It can be seen that PCA on raw images does better than PCA on Gabor pre-processed images and CCA on raw dimensions is much better than CCA on Gabor pre-processed images. In addition, RAWCCA with just 6 components has managed to get better results than GABPCA.

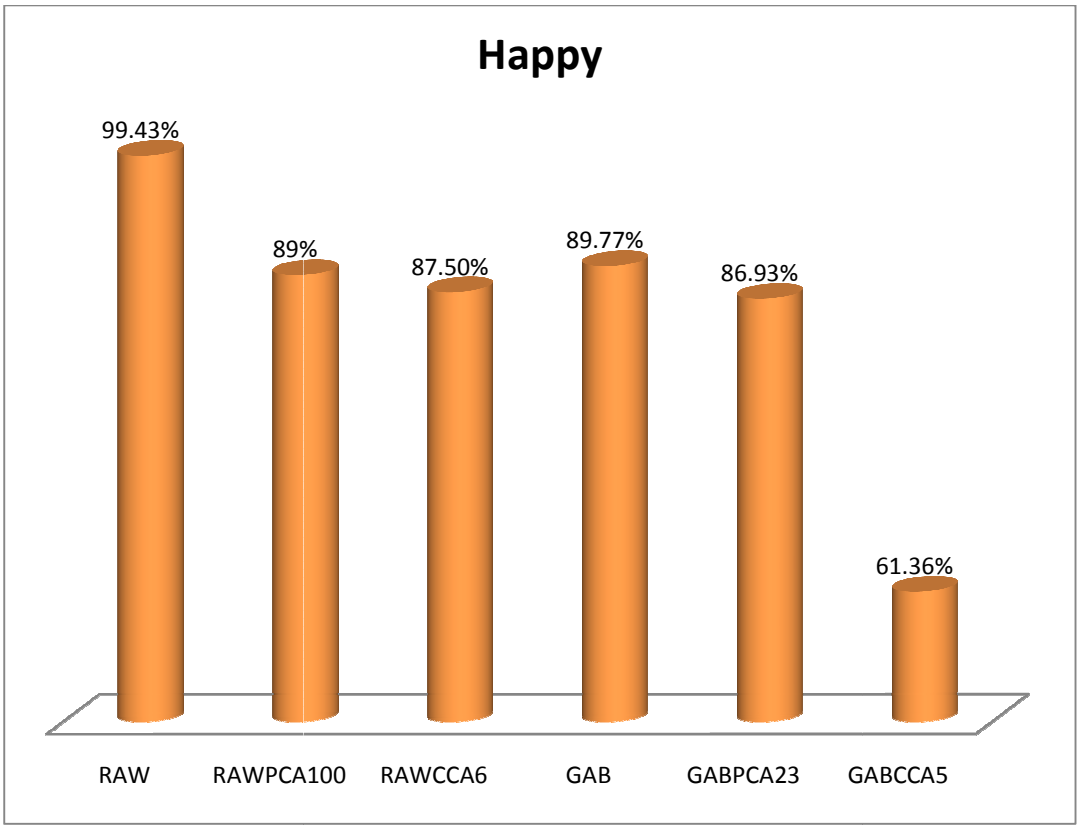


Figure 5.15: Classification accuracy of all models for – happy expression (RAW and GAB are the best)

Fear expression: Figure 5.16 charts the results of fear expression for all models. The RAW model does very well with fear face images followed by RAWPCA99 model at 83.52% and 82.39% respectively. It can be seen that PCA on Gabor pre-processed images with 23 components seems to do better than CCA on Gabor pre-processed images with 5 components. In addition, RAWCCA with just 6 components has managed to get better results than GABCCA with 5 components.

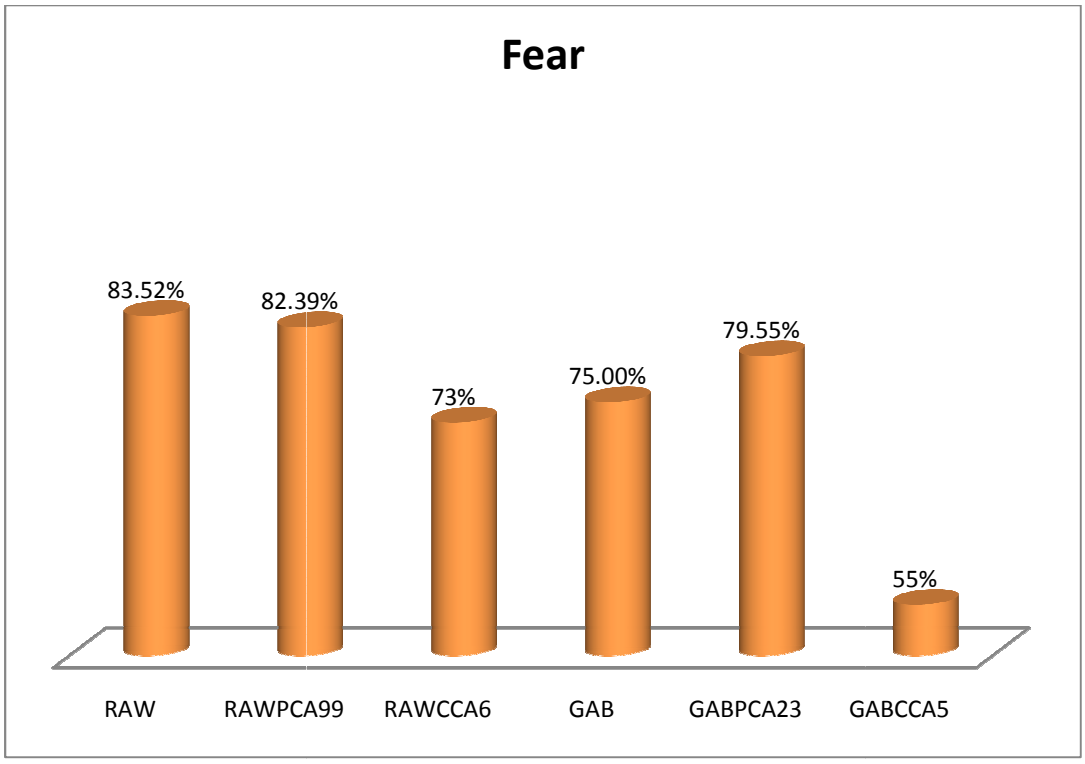


Figure 5.16: Classification accuracy of all models for – fear expression (RAW and RAWPCA are the best)

Sad expression: Figure 5.17 charts the results of sad expression for all models. The RAW model does quite well with sad face images closely followed by RAWPCA96 model at 77.27% and 74.43% respectively. It can be seen that PCA on Gabor pre-processed images with 22 components seems to do slightly better than classification with only Gabor pre-processed face images without any dimensionality reduction. In addition, RAWCCA with just 7 components has managed to get slightly better results than GABCCA with 5 components.

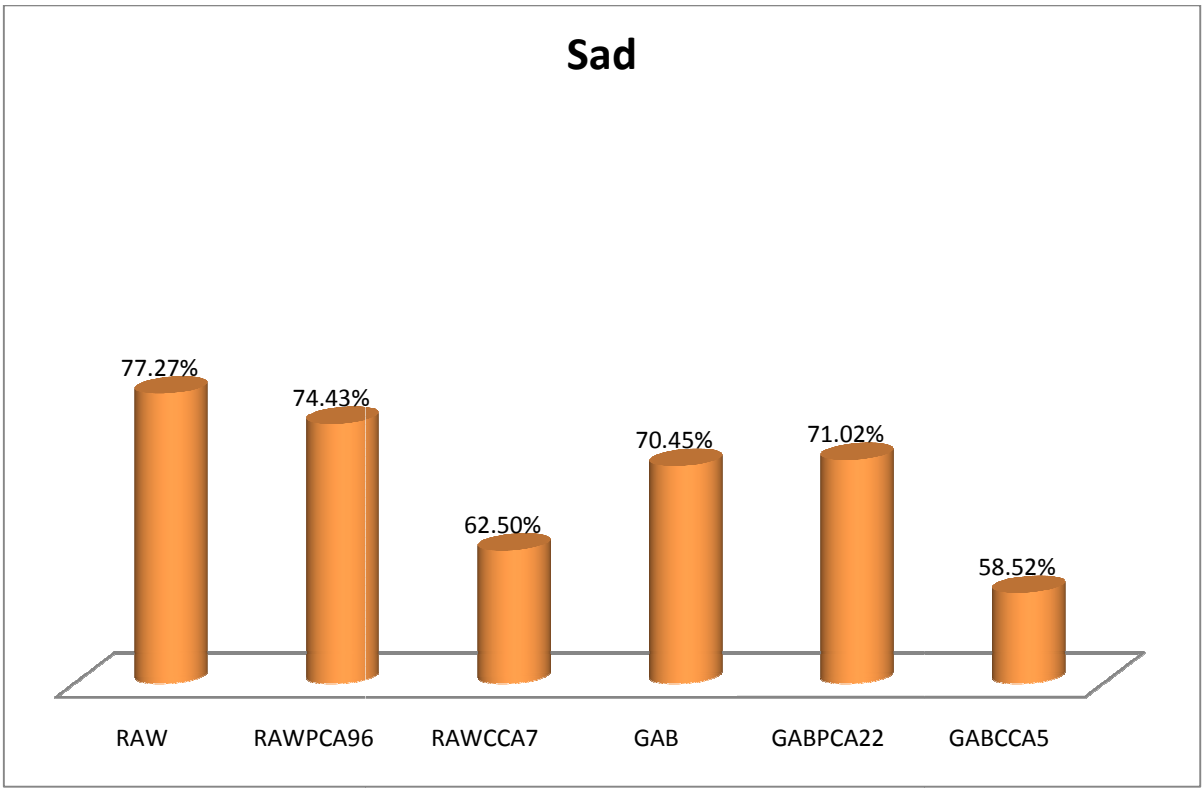


Figure 5.17: Classification accuracy of all models for – sad expression (RAW and RAWPCA are the best)

Surprise expression: Figure 5.18 charts the results of surprise expression for all models. The Gabor pre-processed model with no dimensionality reduction method performs better than all of the rest of the models unlike other expressions where the RAW model has always done better. This is however closely followed by the RAW model (GAB - 95.45% and RAW - 94.89%). On similar lines, PCA on Gabor pre-processed images with 23 components seems to do slightly better than classification of raw face images with PCA at 103 components. The RAWCCA model with just 6 components has managed to get slightly better results than GABCCA with 5 components. All models have remarkable results in comparison to other expressions.

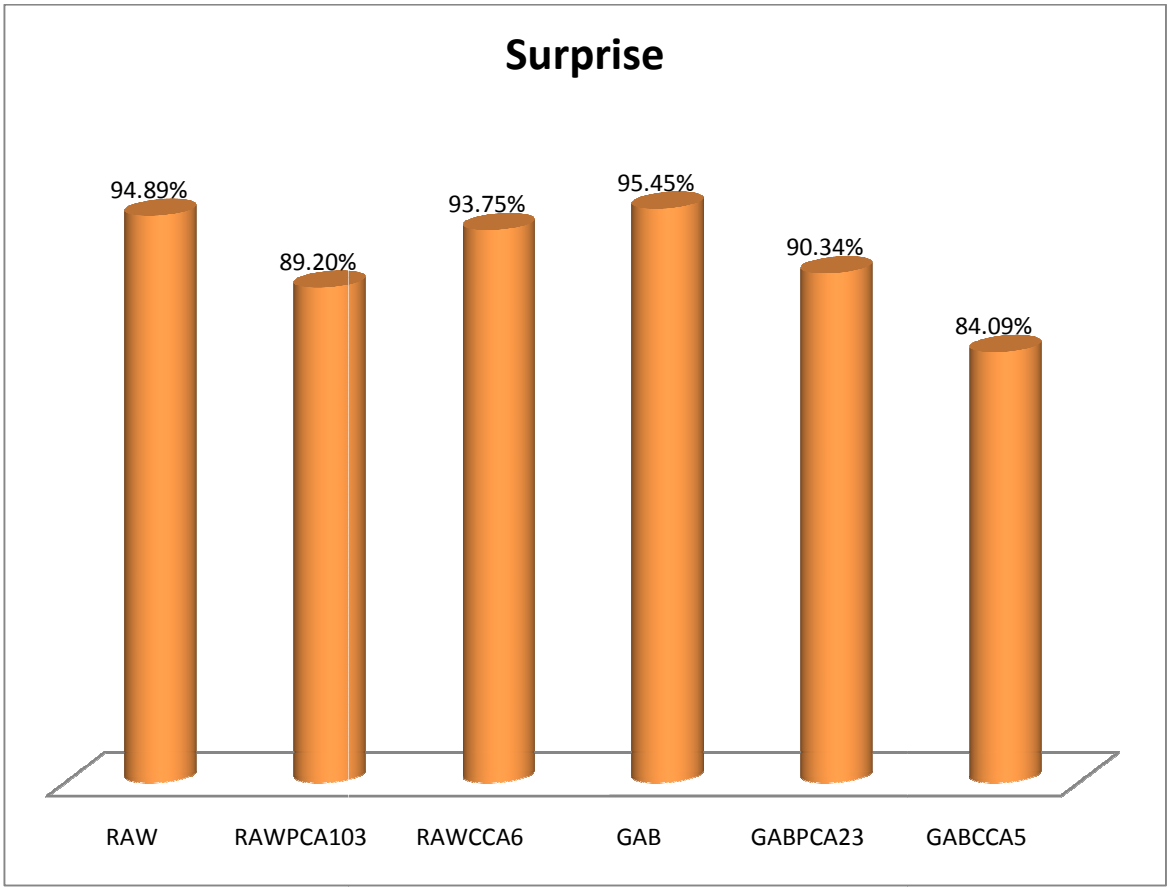


Figure 5.18: Classification accuracy of all models for – surprise expression (GAB and RAW are the best)

Disgust expression: Figure 5.19 charts the results of the disgust expression for all models. The RAW model with no dimensionality reduction method performs better than all the models at 90.34%. This is closely followed by RAWPCA101 model at 80%. PCA on Gabor pre-processed images with 23 components seems to do slightly better than the Gabor pre-processed face images at 76.68% and 73.30% respectively. RAWCCA with just 5 components have managed to get slightly better results than GABCCA with 5 components.

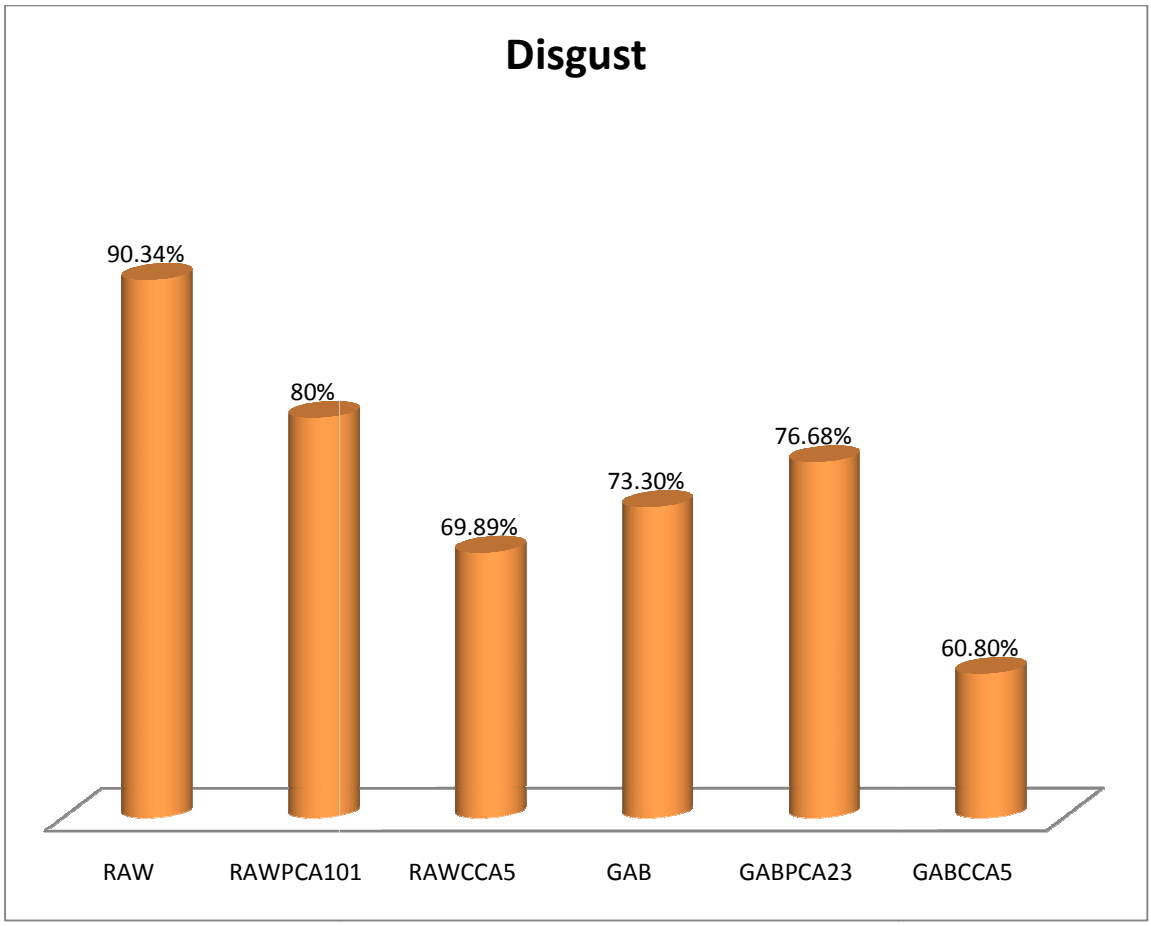


Figure 5.19: Classification accuracy of all models for – disgust expression (RAW and RAWPCA are the best)

Table 5.7 and Figure 5.20 show the average classification results for every expression averaged across all the models.

Table 5.7: Classification accuracy for all expressions averaged across all models

Angry	Happy	Fear	Sad	Surprise	Disgust
72.07%	85.67%	74.74%	69.03%	91.29%	75.17%

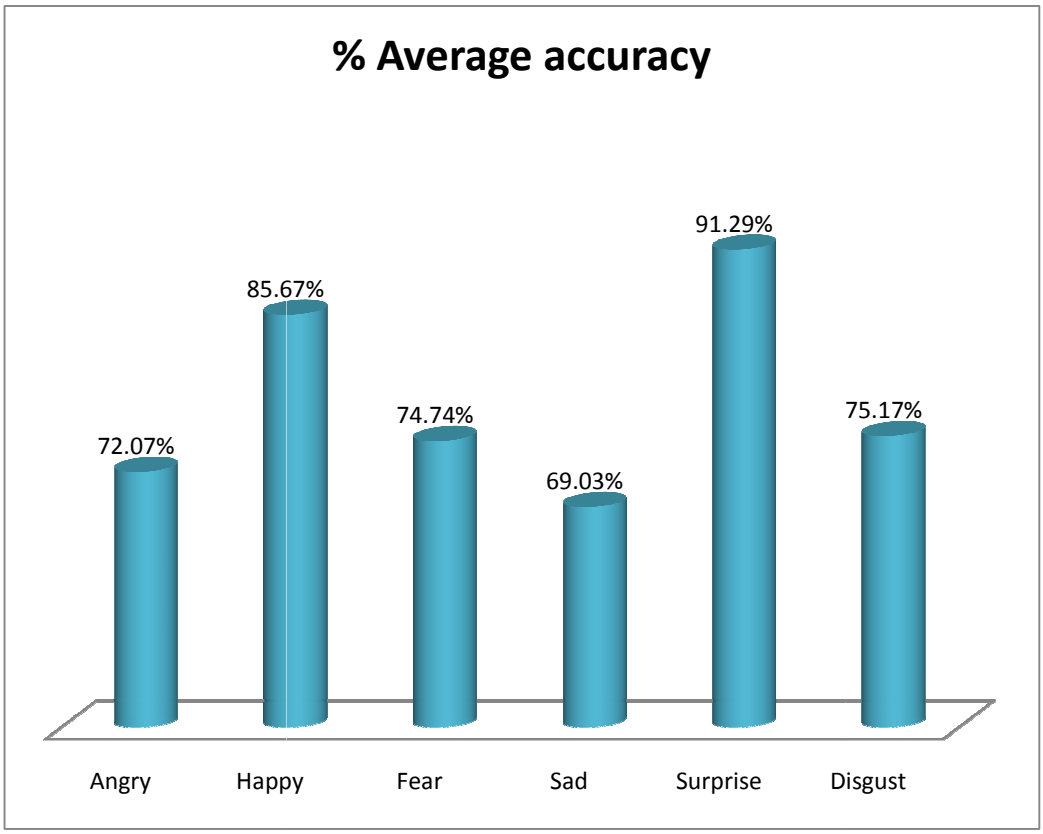


Figure 5.20: Classification accuracy of all expressions - averaged across all models

The averaged results in Table 5.7 and the plots in Figure 5.20 show that surprise, happy and disgust expression recognition is easier than fear, angry and sad.

5.8 Discussion and analysis of the results - Model wise and Expression wise

The recognition rates of this system seem to be really encouraging in comparison with other results by various researchers over the past few years. Unfortunately, these results cannot be directly compared because there have been large differences in the datasets used, the methods adopted for feature extraction, dimensionality reduction and also the type of classifier used. The lack of literature on all similar models that are used here makes the comparisons even more difficult. However, a sincere effort has been made to compare the systems which are nearest to these and analyzed.

Automatic facial expression systems attempt direct interpretation of facial display of emotion and indirect interpretation using facial expression dictionaries. Research based on classification of facial emotions is discussed here.

A very recent work by Liejun et al (2009) studied facial expression classification using SVM by modifying various kernels. Their results look very impressive. However, on further investigation a number of important issues can be highlighted. They used the JAFFE dataset of 210 images of 10 individuals each with 7 expressions (including neutral) and 3 images per expression. The comparison is shown in Table 5.8. The average accuracy of their model which is similar to my RAWPCA model gave an accuracy of 94.8%. The average accuracy of my RAWPCA model (for all expressions) is 80.91%. Their training set uses 70 images from the 210 mentioned earlier and this implies that there are 30 test images for each expression. They used PCA for feature extraction to retain only 28 components. The test set is 210 images and it included the images of the training set and thus would indeed increase the accuracy considerably. In comparison to that, I have used a unique set of 88 individuals with no repeating images for any expressions. The total training set used has 176 face images (for any expression alongside neutral at any time). This was divided into 4 subsets of 44 images each. The training set used is 132 images and test set is 44 images. The classification was performed 4 times by considering the test set as one of the 4 subsets while using the rest as the training set. Finally, the average was obtained. Hence, the test set can be thought of as a set of unique 176 images and is much larger than the dataset used by Liejun et al. In addition, the size of the images they have used is large (256×256) and I have used a smaller size of (64×64) in order to compensate for the larger size of the number of samples (88 face images per expression) used. The results of their experiments suggest that angry and disgust expression was the easiest to be identified, followed by surprise, happy, sad and finally, fear. My RAWPCA model finds surprise and happy to be the easiest, followed by fear, disgust and sad and found angry face image classification hard.

Table 5.8: Comparison with Leijun's model

	Total number of images	Individuals	No. of expressions	Size of Training Set	Size of the Test Set	Average
Leijun's model	210 (10 × 3 images/ expression)	10 (all female)	7	70	210	94.8%
Average RAWPCA Model	616 (88 × 7 expressions)	88 (44 male, 44 female)	7 (One expression at a time against neutral)	176	176	80.91%

Liu and Wang (2006) studied facial expression recognition based on a fusion of multiple Gabor feature extraction. Though this work aims to compare a NN based classification with a PCA based classification of the pre-processed face images by Gabor filters, the results of GABPCA are comparable to the work explained in this thesis. They used the JAFFE dataset of 10 subjects; two images per expression and 7 expressions including the neutral. From a total of 219 face images, 140 are used for training and 79 are used for the test set. Thirteen channels are used to accommodate the 5 scales and 8 orientations. Each channel is a group of different Gabor filters that have the same scale or orientation at specific fiducial points. Liu and Wang perform Gabor filtering using all the channels and regard the maximum of all channel features as the vector. They perform a number of classifications including PCA and neural networks. There is thus some comparison with my GABPCA. My GABPCA performs best with surprise, followed by happy, fear, disgust, angry and sad. Their model also recognizes surprise with good accuracy and the rest in the order sad, fear, disgust, angry and happy. Their results are similar except for the happy and sad expression accuracy rates.

Lyons et al (1999) report achieving 75-92% recognition accuracy using Gabor wavelets with Elastic bunch graph method for feature extraction followed by LDA + PCA + classification. Using RBF based neural networks of features selected by optical flow method results in 88% accuracy (Rosenblum *et al.*, 1996). Padgett and Cottrell's (1996) research on facial expression with PCA and NN has been able to achieve a classification accuracy of 86%. Essa and Pentland (1997) obtained 98% accuracy with feature extraction using optical flow coupled by a physical muscle model that described the skin and texture to extract features followed by a motion energy model for classification. Lanitis et al (1997) obtained a 74% accuracy with a dataset of 690 images (300 in test set and 390 in training set) that used appearance based feature extraction that followed with mahalanobis distance based classification. Using expert rules for classification of emotional displays where feature extraction was by multiple feature detection resulted in a 91%

success rate (Pantic and Rothkrantz, 2000; Pantic and Bartlett, 2007). Dailey et al (2002) experimented on the POFA (Ekman and Friesen, 1976) dataset by performing Gabor filtering at specific grid points and using PCA for dimensionality reduction and followed it with LDA for classification and obtained an average of 90% accuracy. They found that fear was the expression that was most difficult to be recognized out of the basic expressions. Though the methods employed are different from those I have used for feature extraction, the closest model is my RAWPCA model which also recognizes fear and sad expressions with the least recognition accuracy in comparison to other expressions. The happy expression recognition was also the easiest of all.

A research by Buciu et al (2003) on the JAFFE dataset (Lyons *et al.*, 1998) using Gabor filters for feature extraction and SVM (linear kernel) for classification resulted in an accuracy of 95.18%. These results are good in comparison to the 79.92% that I have obtained for the GAB model with SVM classification. However, the dataset used is small consisting of 213 images of 10 individuals in 3-4 poses for one of the 7 different expressions. Also, the Gabor filters used are tuned for 3 frequencies and 4 orientations which result in a total of 12 filters. When the original image of size 80×60 is convolved with the filters, it results in a size $80 \times 60 \times 12$. They use down sampling of the Gabor filtered image to obtain a matrix of size $20 \times 15 \times 12$ and thereby a feature vector of size 1×3600. This is subjected to linear SVM classification. The computational complexity of the problem with larger dataset cannot be underestimated and would be an interesting to extend this to a larger dataset. They also report that the since the database is limited, the recognition rate is measured over identity using a leave-one-out strategy which makes maximal use of the available data for training. These results were averaged over the subjects and classes. They too report that fear is one of the most difficult expressions to be judged along with the expression sad.

Black and Yacoob (1997) report 83-100% recognition rate with video sequences, extracting features by local motion modelling and classification by expert rules. Wang and Yin (2007) used the Cohn Kanade dataset (Kanade *et al.*, 2000) and MMI dataset (Pantic, 2005). They used a topographic analysis technique for feature extraction. The topographic context is used for facial expression classification. The facial topographic surface is obtained for various regions of the face and it is labelled to form a terrain map; the statistical details for all regions are put together for the entire face to obtain a topographic feature vector. Classification is performed by LDA and SVM apart from other methods which are of lesser relevance. The LDA provided an average recognition rate of 82.68% and SVM resulted in 77.68%. They also report that expressions surprise and happy were well detected by the LDA classifier. Their results compliment Cohen's system (Cohen *et al.*, 2003) though the database consisted of only video sequences and the recognition rate was best at 81.80%.

Some researchers use classification of facial action using Facial Action Coding Scheme (FACS) (Ekman and Friesen, 1976). Research based on these systems is discussed here. Littlewort et al

(2006) have suggested that machine learning when combined with appearance based feature extraction are highly robust for expression recognition. Many machine learning methods have been applied and aim to achieve a high accuracy with an automatic facial expression recognition system. They include Adaboost, SVM, and LDA.

The datasets used by them are:

- Cohn-Kanade (Kanade *et al.*, 2000) which has 313 sequences of frames that has expressions changing from neutral to one of the six basic expression with maximum intensity.
- Pictures of facial affect (POFA) (Ekman and Friesen, 1976) have 110 images from 14 subjects.

A combination of Gabor filtering for feature extraction, and best filter selection is done by Adaboost followed by SVM classification in seven-way forced choice (six expressions and neutral) resulted in the best accuracy. It is an automated FACS recognition system and hence, the results are facial action labels. The results were 93.3% and 97% correct on these two publicly available datasets.

Bartlett et al (1999) obtained up to 96% using difference images and Gabor jets for feature extraction followed by nearest neighbour using ICA for classification. Fasel and Luettin (2003) reported a maximum expression recognition rate of 83% by difference image for feature extraction and ICA + Euclidean distance based classification. Cohn et al (1997) used Hidden Markov Model for classification of features extracted using feature point tracking and achieved a 86% recognition rate. Using expert rules for classification of facial actions using FACS where extraction of features is by multiple feature detection results in 89% recognition rates (Pantic and Rothkrantz, 2000; Pantic and Bartlett, 2007).

Buenaposada et al (2008) used the Cohn-Kanade dataset (Kanade *et al.*, 2000) with video sequences to classify facial expressions. Only those sequences that have clearly identifiable prototypical expressions are used. This is possible with only 333 sequences. Each image begins with a neutral expression and ends with an expression that is labelled by FACS. They make use of a tracker system for feature extraction and dimensionality reduction by LDA. Here, a facial expression is represented as a set of samples that model a low dimensional manifold in the space of deformations generated by the tracker parameters. An image sequence is considered as a path in the space of deformations. Using the nearest neighbour technique, the probability of occurrence of an image is estimated. A recursive Bayesian procedure is adopted to combine these probabilities and assign a target sequence to the facial expression with maximum probability. This resulted in an average recognition rate of 89.13%.

A brief summary of some of the current state of the art research in the field of facial expressions using static images is presented here:

- Classification by Facial emotions where the output is one of the six expression classes (Zheng *et al.*, 2009):

  - Cohen et al (2003) obtained an recognition rate of 66.53 % and 73.22% with the Cohn-Kanade (Kanade *et al.*, 2000) and Ekman-Hager datasets respectively. They used shape models and Gabor wavelets for feature extraction followed by a Linear Discriminant Classifier (LDC).

  - Fasel et al (2004) used gray-level intensities for feature extraction followed by neural networks for classification on the Cohn-Kanade dataset to obtain a classification accuracy of 38-68%.

  - Gunes and piccardi (2005a) used the FABO dataset (Gunes and Piccardi, 2005b) with shape features and optical flow for feature extraction followed by Bayesian network for classification that resulted in 80-100%.

  - Ioannou et al (2005) use a facial animation parameter technique for feature extraction followed by neurofuzzy network to obtain an accuracy of 78%.

  - Lee and Elgammal (2005) used the pixel intensities of the face region for feature extraction followed by decomposable models for classification on the Cohn-Kanade dataset with an accuracy of 61.85%.

  - Pantic and Rothkrantz (2004a) used frontal and profile points for feature extraction followed by rule based and case based classification on the MMI dataset (Pantic, 2005) to obtain a classification accuracy of 83%.

  - Sebe et al (2004) used motion units for feature extraction and k-means nearest neighbour for classification on the Cohn-Kanade dataset and obtained a recognition rate of 93%. They too obtained 95% for their dataset but it should be noted that their dataset was much smaller than the Cohn-Kanade dataset.

  - Wang et al (2006) used 3D surface labels for feature extraction followed by LDA for classification on the Binghamton BU-3DFE dataset (Yin *et al.*, 2006) to get an accuracy of 83.6%.

- Using Haar features extraction followed by Adaboost for classification by Whitehill and Omlin (2006) resulted in 92.35%.

- Classification by Facial actions where the output is in terms of the Action Unit (Zheng *et al.*, 2009):

- Lucey et al (2007) performed feature extraction by active appearance model (AAM) followed by SVM based classification that resulted in 95% accuracy with the Cohn-Kanade dataset.

- Bartlett et al (2005) used Gabor wavelets for feature extraction and Adaboost and SVM for classification which resulted in 93.4% accuracy when Cohn-Kanade and Ekman-Hager dataset was used together.

- Pantic and Rothkrantz (2004) obtained an accuracy of 86% by extracting features using frontal and profile facial points and classifying by expert system rules.

Most of these studies using different databases, different feature extraction methods and various classification methodologies seem to recognize happy and surprise with ease and also find fear and sad difficult. This may be because they involve subtle changes in appearance (Buenaposada *et al.*, 2008). The various models that have been explained in this thesis (RAW, RAWPCA, RAWCCA, GAB, GABPCA, GABCCA) compliment these results though with different classification accuracies. Some studies have failed to use six basic prototypical expressions and some have not used a balanced set and all these issues have an impact on the classification results. Although the database used most often is the one by Cohn-Kanade, the sequences that are used for training and testing are not the same and this means comparisons can be difficult.

5.9 Conclusions

All automatic facial expression systems focus on six basic prototypical expressions. This is based on the research by Ekman and Friesen (1971) and also by Izard (1977) who proposed that there are emotion specific facial expressions and compliments the work of Darwin. In our everyday life, however, occurrences of such prototypic expressions on their own are relatively rare. Instead, emotions are often communicated by subtle changes in the facial features such as creasing of the cheeks, wrinkles in the forehead or dropping of the jaw, just to name a few and may also be a combination of more than one emotion such as angry - sad or a happy- surprise

(pleasant surprise). To design a system that is really capable of detecting these expressions and classifying them is not a trivial task.

Other important factors in designing a good computational model are robust and precise detection of the facial features, independent of gender, identity, race, shape of the face, texture, colour, presence of facial and scalp hair (Tian *et al.*, 2005). Expressions are a very important aspect of communication and there have not been any systems developed that use the facial emotion to convey meaning (Schwaninger *et al.*, 2006). The system should be capable of identifying any micro expressions which are very rapid and are missed very easily (Lisetti and Schiano, 2000). Spontaneous or posed expressions, still and video images are some of the factors that affect the recognition of facial expressions. Mostly, the images are expected to be in frontal view however, in reality or in spontaneous expressions there could be a lot of rigid head motions. The system should be robust despite changes in hair-style, changes in lighting conditions, and other distractions such as glasses or facial hair. A human visual system easily fills in gaps in the areas that are occluded. Hence, the ideal system should also be capable of doing this. Eye openings and contrast between iris and sclera differ among various individuals of different ethnic background, which could result in difficulty to track eye movements or even facial features (Tian *et al.*, 2005).

Recent advances towards the emotion recognition include voice or audio based recognition systems (Zheng *et al.*, 2009). Though a number of facial expressions occur during a conversation rather than on its own, none of the approaches so far have dealt with it (Fasel and Luettin, 2003). Though none of the methods enabled a one-to-one comparison to the results of my computational models, an honest attempt has been made to critically evaluate these results with current research in the field.

CHAPTER SIX

Facial expression recognition by humans

6.1 Introduction

Despite an enormous amount of input from various researchers in recent years in the area of automatic recognition of facial expression, no consensual results have emerged from these studies. There have been studies investigating generation of facial expressions and tasks relevant to recognition of facial expression and studies in these domains have spanned over a century. The oldest articles date back to 1844 even before Darwin (1872); Bell (1844) studied facial expressions and reported differences between the positive and negative emotions in terms of the muscle movements on the face. A very recent study by Matsumoto and Willingham (2009) compared the expressions of blind and non-blind individuals and their findings provide sufficient evidence that the production of spontaneous facial expressions of emotion is not learned. They conclude that something genetically wired is responsible for the facial expressions of emotions. This suggests that recognition of facial expressions is a trivial task for humans. This chapter concentrates only on tasks involving recognition of facial expression by humans. There have been a limited number of studies comparing the performance of human subjects with computational models for facial expression recognition. This chapter reports an experiment that involved human participants performing tasks of expression recognition and compares the performance with the computational models that have been described in Chapter 5.

6.2 Background research

Earlier research in similar areas has included human performance in facial expression similarity studies (Susskind *et al.*, 2007), classification accuracy (Stathopoulou and Tsihrintzis, 2007; Ekman and Friesen, 1976; Zucker *et al.*, 2007; Jinghai *et al.*, 2006) and studies to find the minimum presentation time for accurate identification of facial expressions, danger and threat detection (Milders *et al.*, 2008; Ohman *et al.*, 2001). Some studies have concentrated on all basic expressions; whereas others have concentrated only some of the expressions such as anger and fear.

According to Ohman, Lundqvist and Esteves (2001), danger and threat are processed faster due to the evolutionary benefit. This would suggest that the response time is shorter for fear and

anger expression recognition than for other expressions (Hansen and Hansen, 1988). However, Kirita and Endo (1995) have shown that the response time for happy faces was shorter than sad faces. Carvajal, Vidriales, Rubio, and Martin (2004) found happy expression identification were easiest in comparison to angry and neutral expression and were detected faster. They concluded that the facial expression of happiness is the easiest one to identify, and that it could be attributed to the higher prevalence of this expression in social circumstances. Milders, Sahraie and Logan (2008) suggest an advantage in processing happy expressions and support earlier studies suggesting a bias towards facial expressions of positive valence.

Wagner, MacDonald and Manstead (1986) examined whether participants can accurately distinguish spontaneous facial expressions for the seven affective states (six emotional and one neutral). Happy, angry, and disgust expressions were recognized at above-chance rates, whereas surprised expressions were recognized at rates that were significantly worse than chance. Other studies that involved classification accuracy tasks have resulted in different accuracies and they are not consistent. Stathopoulou and Tsihrintzis (2007) found sad expressions were hardest to recognize, followed by angry, disappointment, disgust, scream and smiling ; whereas surprise was easiest. Jinghai, Zilu and Youwei (2006) found that humans found classifying happy and surprise expression recognition easy. They found anger, disgust and sadness more difficult to classify and expression fear being the hardest of all. Zucker, Radig and Wimmer (2007) found happy and surprise relatively easy to recognize, followed by anger, disgust, sadness and fear was the hardest.

Some comparisons between human and computational performance in various facial expression recognition tasks have been conducted by Dailey, Cottrell, Padgett, and Adolphs (2002), Susskind, Littlewort, Bartlett, Movellan, and Anderson (2007), Jinghai, Zilu, and Youwei (2006), Calder, Butron, Miller, Young and Akamatsu (2001) and by Milders, Sahraie, and Logan (2008). The results of the empirical work reported in this thesis are compared with these studies in the later sections of this chapter.

Here, the classification performance of the computational models that has been described in Chapter 5 is compared with the human performance in the classification of facial expression. Two types of relevant analysis were performed:

- Bi-Variate Correlation analysis
- Signal Detection Theory (SDT)

For the purposes of comparison, the response time (RT) for human subjects in judging the facial expression of a given face was considered to be analogous with the distance measure from the hyper-plane for the computational models for that face. It can be reasonably argued that both are indicators of how ‘easily’ the classification was made. The analyses therefore focus on

examining the relationship between the average RT of the humans in responding to the stimuli and the distance measure from the hyper-plane for the computational models.

The natures of the analyses are described below:

- Bi-Variate Correlation – finds the strength of the relationship between two variables. The value of the correlation coefficient varies between +1 and -1. When the value of the correlation coefficient is close to ± 1, it means a perfect degree of association between the two variables. If the value is around 0, the relationship between the two variables is considered to be very weak. There are three types of correlation in use in statistics: Pearson correlation, Kendall rank correlation and Spearman correlation. Here, only Pearson correlation is used as both types of data are interval and a linear relationship is sought, i.e. is it the case that the faster the response times of humans the greater the distance from the hyper-plane? If so, a significant negative correlation would be expected. The strength of any such relationship would be taken to indicate that one set of data is mirroring the other. The value of the correlation between the two measures decides the strength of the relationship. In this case the two variables in question are, average RT and the distance measure. In addition to the strength of the relationship described by the correlation coefficient, another parameter for analysis is the significance of the relationship. This indicates how unlikely it is that the correlation coefficient is the result of chance factors such as noise in the data and is expressed as a probability value. The results are considered to be significant at the level of 0.05 or less.

  The larger the correlation coefficient, the stronger is the relationship and the smaller the p-value i.e. the more significant the relationship.

- Signal Detection Theory (SDT) - Signal detection involves the perception of some information from the environment (the signal) and a decision process for categorizing that information as either being or not being the target signal (Abdi, 2007). It is suited to data where speed and accuracy may be traded off against one another i.e. where error data is informative. The presence of a face image with prototypical expression or the presence of a neutral face image and the response to that can be best described by the following four possibilities:

  - "Hit" (correct acceptance) = the signal is present, and it is detected.
  - "False Alarm" (incorrect acceptance) = the signal is absent, but it is detected.
  - "Miss" (incorrect rejection) = the signal is present, but it is not detected.
  - "Correct Rejection" = the signal is absent, and it is not detected.

In this case the presence of a face image with basic expression is taken as signal present. A parameter d' (d - *prime*) is used to describe the strength of this signal. This is the difference between the means of two distributions (distance measure from the hyper-plane of the SVM for the computational model and the average RT for the human subjects) which are often thought to be Gaussian and corresponds to the effect of the signal. It is best described by the Equation 6.1.

$$d' = |z(H) - z(FA)| \tag{6.1}$$

where $z(H)$ and $z(FA)$ are the z-scores of the instances of Hits and False Alarms.

Considering only the entries that account for Hit (H) and False Alarm (FA) from the computational models and the human subjects, the d' is calculated. The frequencies of these four responses are dependent on one another. For example when the signal is present, the proportion of hits and the proportion of misses add up to one (because when the signal is present the subject can say either Yes or No). Similarly, in the event of signal being present, the proportion of FA and the proportion of Correct Rejection will add up to one.

The larger the d' value, the better is the performance. A d' value of zero means that the ability to distinguish between the two trials (presence of signal or not) is least and a value close to 4.6 indicates a nearly perfect ability to distinguish between two tasks (Oliva *et al.*, 2005).

6. 3 Method

6.3.1. Participants

Thirty one healthy individuals took part in the study. All the individuals were within the age group of 18 to 59 years of age and ranged across various ethnic backgrounds and included 18 males and 13 females. The participants were from various professions and participated on request. This experiment was undertaken adhering to the ethics guidelines in the university and approved by the ethics committee.

6.3.2. Design

A single variable was manipulated – expression. There were seven levels of this variable, angry, happy, fear, sad, surprise, disgust and neutral. The dependent variables were time taken to identify the expression and errors made in identification. Due to the time consuming and somewhat onerous nature of the task, not all participants agreed to take part in all sessions so a repeated measures design, although desirable, was not achievable.

6.3.3. Materials

A total of 644 unique face images from the BINGHAMTON BU-3DFE database (Yin *et al.*, 2006) was used for this study. The dataset is the same that had been used with the computational models described in detail in Section 5.2 of Chapter 5. A total of 588 of the face image dataset were used for testing and the remaining 56 images were used in the practice session. The test set has 7 expressions and is balanced in terms of gender (294 male and 294 female). There are 84 face images for every expression in the test set and 8 face images for every expression in the practice session. The test set was balanced in terms of expression and gender- angry and neutral. The images in the dataset are already processed by cropping to show only the face area to exclude any hair or clothing and are of size 256×256 . No further processing or image reductions are done on the original images and hence are of good quality for perception.

6.3.4. Procedure

Every session included classification of one prototypical expression from neutral. Twenty sessions were conducted for every expression classification against neutral. As there are six expressions, there were 120 sessions. In each session a total of 168 images were shown to the participant. It consists of 84 neutral images and 84 images belonging to one of the six basic expressions. These 168 images were shown to the subjects as a set of six blocks with 28 images in each block. They were equally balanced in terms of gender and expressions (14 neutral images and 14 face images that belonged to one of the six basic expressions in each block). A preview block was used as the practice session to enable the participants to get used to the procedure. The preview block in each session (for each expression) had 16 images of 8 neutral and 8 images of one of the basic expressions and they were randomly shown to the subject. Of the 31 participants, some kindly agreed to attend six sessions corresponding to the six expressions. Others attended at least one of the sessions. No individual participated in more than one session for the same expression.

A tool called TESTBED (Taylor, 2003) was used to document the responses of the participants. This is a response test generator program which records the response time (RT) and the classification of a face image by individual subjects.

In every session, the face images were randomly shown on the computer screen from one block at a time and the participants responded by pressing the mouse button. Participants were given the practice session with the preview block to identify the expression on the face image. The images were displayed in a random order and a new image appeared on the screen when the response to the previous image was recorded. The participant was asked to click the left (for neutral face image) or the right mouse (for the basic expression in that session) button as soon as he/she identifies the expression of the face displayed. The click on the left mouse button in response to the neutral face was a correct guess and vice-versa. The TESTBED recorded the response time and also the correctness of the judgment. The participants were allowed to take an interval period between evaluations of every block. Some of the participants kindly agreed to take part in all sessions, and some have just taken part in one session; in other words, some have judged all six expressions and some have judged only one expression but all blocks in it. None of the participants had seen any of the images prior to the experiment. The data obtained from all the participants was stored in a single file and imported to a statistical package for further analysis.

6.3.5 Results

The results of human performance in the classification of static facial expression images can be compared in terms of response time (RT) and the classification accuracy for each expression. Table 6.1 shows the average RT for classification by the participants for every expression that is properly classified. Table 6.2 shows the average classification accuracy for each expression obtained from all participants.

Table 6.1: Average Response time (RT) for each correctly identified expression

Expression	Average RT in seconds (S.D in brackets)
Angry	1.09 (0.1925)
Happy	0.85 (0.1187)
Fear	0.99 (0.1636)
Sad	1.04 (0.1849)
Surprise	0.74 (0.1078)
Disgust	0.85 (0.1190)

The cut-off for RT to remove the outliers was calculated as 3*SD + actual average RT. We assume the smaller the RT for a correct judgment means the easier the classification task was for the subject. For the analysis, the RT for the entries with wrong guesses was removed from the result. It seems that the expression surprise is recognized faster by humans, followed by happy, disgust, fear, sad and the expression angry was the hardest. A one way ANOVA has been performed to confirm that the differences are significant ($F(5, 1007) = 133.113$, $p<0.01$). The Post Hoc Scheffe comparisons showed RT was not significantly different for expressions angry and sad RT ($p=0.121$), for expressions happy and disgust ($p=1.000$) and for expressions fear and sad ($p=0.123$). However, it showed very good significant differences for other expression comparison with $p<0.001$.

The results of the task of classifying facial expression in digital images by human subjects are shown in Table 6.2. When all the face images were shown to all the participants, the accuracy obtained by the twenty participants was averaged for individual expression was taken.

Table 6.2: Results of human performance in classification of facial expressions

Expressions	Human Performance (% Accuracy)
Angry	82.9% (139/168)
Happy	94.6% (159/168)
Fear	87.9% (148/168)
Sad	82.9% (139/168)
Surprise	97.7% (164/168)
Disgust	92.8% (156/168)
Average	89.8% (151/168)

The results in the Table 6.2 show that the expressions surprise, happy and disgust were recognized with a very good accuracy. Recognizing the expression fear was a bit difficult, with sad and angry being equally hard. The average accuracy for all expressions was 89.8%. The average accuracy for every image obtained by the twenty participants was obtained. Then, the average accuracy for all 168 face images was obtained. A one way ANOVA confirms that the differences are significant.

The results in Table 6.1 and Table 6.2 when compared, suggests that the expression surprise seems to be recognized fastest of all other expressions and also with the best classification accuracy. Angry expression recognition was the slowest and was not recognized easily.

6.4 Analysis

The response time and the classification accuracy recorded by the TESTBED are analyzed.

6.4.1 Response Time

Hansen and Hansen (1988) found that an angry face could be detected faster than an happy face in the crowd and hence concluded that facial expressions that are threatening are processed better than the others. Later further research (Ohman *et al.*, 2001; Fox *et al.*, 2000) suggested that angry faces were indeed processed faster. Contradictory results were obtained when these experiments were repeated by Hampton, Purcell, Bersine, Hansen and Hansen (1989) and Byrne

and Eysenck (1995) and recently by Carvajal, Vidriales, Rubio, and Martin (2004) who found that happy expressions were easiest for identification in comparison to angry and neutral expression and were detected faster. They concluded that the facial expression of happiness is the easiest one to identify, and that it could be attributed to the higher prevalence of this expression in social circumstances. In this study, happy expressions were indeed faster to be recognized than angry expression as can be seen from the data in Table 6.1. Kirita and Endo (1995) also have shown that the response time for happy faces was smaller than the sad faces. A study suggests that people in positive moods are faster in recognition of happy faces as compared to people who not (Leppänen and Hietanen, 2003). If this is to be believed then recognition rates and accuracy could well be affected by the moods of the participants. They also suggest that in general positive expressions such as happiness is recognized faster than negative expressions such as disgust or sad and my results complement these findings to some extent. A very recent study by Bannerman, Milders, Gelder and Sahraie (2009) supports earlier studies of Ohman et al and suggests expressions such as fear or threat are detected faster using neuropsychological evidences based on eye movements. A study by Yang, Zald and blake (2007) also provides evidence that fearful expressions are recorded faster by the brain than others and happy expressions are slower to be recognized than even neutral expressions. They suggest that happy expressions signal safety to the brain and hence require no attention. It also suggests that faster recognition of fear expressions could have emerged from the evolutionary survival mechanism and could signal threats in the environment. As one can see, none of the evidence converges on support for any one expression being recognized faster. This could depend on number of factors such as the methods of experiments, database used, the number of male and female face images in the database, the number of males and females in the participants and the debate, and research, goes on. This is an issue constantly discussed in all facial expression recognition tasks which makes comparisons harder and also results in the inconsistencies (Schwaninger *et al.*, 2006; Lisetti and Schiano, 2000; Fasel and Luettin, 2003).

6.4.2 Accuracy

A study by Wagner, MacDonald and Manstead (1986) examined whether spontaneous facial expressions participants can distinguish accurately the seven affective states (six emotional and one neutral). Happy, angry, and disgusted expressions were recognized at above-chance rates, whereas surprised expressions were recognized at rates that were significantly worse than chance. However, in the current study case surprise and happy were identified with better accuracy than other expressions. In their case, they also noted that female subjects were found to be significantly better in displaying facial expression than male. However, although they found that neither gender was found to be better at perceiving facial expressions, female subjects were better at accurately perceiving expressions on the female face than on the male face. The found

that female face images displayed neutral and surprised expressions much more accurately than male face images. Men were found to be good with angry facial expression recognition. Goos and Silverman (2002) found that men were good in posing angry expressions and these would be perceived much more accurately compared to the angry expressions posed by females. Also, they suggest that the angry expression posed by females is not perceived by females any more than males as was previously thought. In another study (Williams *et al.*, 2008) showed that happy faces were detected significantly better than other expressions.

The experiments with Wagner et al used 6 subjects to display facial expression which was recorded when they responded to emotionally loaded photographic slides. Their expressive face in response to the slides was videotaped and shown to a total of 53 participants (15 male and 38 female) to judge the expressions. There is a large bias towards female participants and could be the basis for such conclusive evidence. In this thesis, an equally balanced set of face images in terms of gender and expressions was used i.e., 22 unique female face images and 22 unique male face images. However, the number of female and male participants was not same, with 18 male and 13 female subjects. Based on the other findings of Wagner et al that were mentioned earlier, these discrepancies could be a factor in obtaining different accuracies for individual expressions. Wagner et al found happiness as the easiest followed by disgust and anger and found fear to be one of the difficult ones.

The results of a study by Wimmer, Zucker and Radig (2007) found happy expressions were detected with best accuracy followed by surprise, anger, disgust, sad and the hardest was fear with an average of 64%. They used Cohn-Kanade dataset and the stimuli were video sequences. The POFA dataset by Ekman and Friesen (1976) results in an average accuracy of 90% with happy expression detected best followed closely by surprise and disgust, sadness, anger and the fear was detected with the least accuracy. Bassili (1979) suggests that for a trained person or a face expert, classification accuracy for the six basic facial expressions is 87%. However, he also points to the fact that this accuracy could depend on a number of factors such as the face being familiar, being an expert in recognizing expressions, the intensity of the emotion on the face, the face image as such or even the ethnicity of the participant and the ethnicity of person whose expressions are being categorized (Altarriba *et al.*, 2003). Stathopoulou (2006) created a database in order to help researchers develop better automatic facial expression classifiers. They also measured the human performance in classifying facial expression. The expressions included were: surprise, smile, scream, sad, disgust, disappointment and angry. They found that surprise was more correctly recognized, followed by smile, scream, disgust, disappointment, angry and sad. A study by Wang, Hoosain, Lee, Meng, Fu and Yang (2006e) that involved only Chinese participants and performed a forced choice labelling technique. This study resulted in the following conclusions - consistent results with earlier studies that show that fear and disgust are difficult to recognize, whilst happiness was easiest followed by surprise (Susskind *et al.*, 2007).

Calder, Burton, Miller, Young and Akamatsu (2001) obtained an average classification accuracy of 82%. The best to worst recognized expressions were: happy, surprise, disgust, fear, sad and

angry. The evidence from various studies does not seem to converge and merit further investigations relevant to all issues. Here again, just as with the experiments with automatic facial expression recognition, the evidence does not converge on a single conclusion. Studies with human subjects have also failed due to reasons such as cultural differences, race and social differences which seems to affect the way and ability to recognize facial expressions (Altarriba *et al.*, 2003).

6.5 Comparison of human performance with computational models in expression recognition

The human performance in the classification of facial expressions was compared with that of the computational models described in this thesis by two types of analysis. The results of these are discussed below. While comparing the results of the computational models with the human participants, only the responses to 84 face images (for every expression) that are common for both experiments were used. Hence to maintain uniformity, although with computational models used 88 face images with each expression; only results corresponding to the same face images used with human subjects and computational models were taken for analysis.

6.5.1 Results of the Bi-Variate correlation analysis

The result of the Bi-Variate correlation analysis for all expressions is shown in Table 6.3.

Table 6.3: Results of Bi-Variate correlation between average RT of human subjects and the distance measure of the hyper-plane for the SVM classifier used with all computational models for correct responses. The numbers in red font indicate significant levels and their corresponding correlation values.

Expression	RAW		RAWPCA		RAWCCA		GAB		GABPCA		GABCCA	
	S	r	S	r	S	r	S	r	S	r	S	r
Angry	0.024 N=141	-0.191	0.645 N=119	+0.043	0.126 N=108	-0.148	0.016 N=128	-0.212	0.065 N=120	-0.169	0.597 N=110	-0.051
Happy	0.090 N=167	-0.132	0.069 N=149	-0.149	0.259 N=146	-0.094	0.043 N=151	-0.165	0.537 N=146	-0.052	0.786 N=103	0.027
Fear	0.018 N=140	-0.199	0.287 N=138	-0.091	0.242 N=122	-0.107	0.250 N=124	-0.104	0.016 N=132	-0.209	0.306 N=92	-0.108
Surprise	0.005 N=159	-0.224	0.598 N=149	+0.044	0.004 N=157	-0.232	0.000 N=160	-0.281	0.032 N=151	-0.174	0.012 N=140	-0.211
Sad	0.086 N=129	-0.152	0.067 N=126	-0.164	0.746 N=104	-0.032	0.455 N=118	-0.069	0.347 N=119	-0.087	0.418 N=98	-0.083
Disgust	0.080 N=152	-0.143	0.946 N=134	0.006	0.683 N=116	-0.038	0.047 N=125	-0.178	0.890 N=128	-0.012	0.053 N=102	-0.192

Significance Value : S , Correlation Value: r

The correlation analysis was performed only on data for correct responses. As expected there seems to be a negative correlation value for average RT versus distance measure for the entries where the models got the classifications correct. This indicates that the images that needed longer time for the participants to classify had a larger average RT and that these images were closer to the classifying hyper-plane of the computational model and had a smaller distance measure.

As can be seen, the right half of the Table 6.3 has more numbers in red font indicating significant correlations. The right hand side of the table has entries for the models with Gabor filters. It suggests that more number of Gabor based computational models have significant correlations with human subjects in comparison to the number of RAW models (without any feature extraction by Gabor filters) with significant correlations. For expression sad, none of the models showed significant values. However, the expression surprise seems to be good even with all models except with RAWPCA. So, does this mean surprise expressions are perceived in a different way than others? A study by Lee and Elgammal (2005) shows that a 3D plot of six

basic expression vectors has the surprise expression far away from the other expressions which could be due to very distinguishable visual motions on the face posing surprise. Also, angry, fear, sad and disgust expressions are located closer to one another compared to the other expressions, and distinguished visually with more subtle motions. As can be recalled from the results of computational models developed, only expression surprise had best classification accuracy with GABOR model as shown in Figure 5.18 of Chapter 5. For all the other expressions, the RAW model gave excellent results.

The association between the two described variables is said to be perfect when the correlation value is very close to -1 (negative sign for the negative correlation). Here, the GAB model for the expression surprise has the largest correlation value of - 0.281 when compared to all other models and it has a significance value of 0.000.

Similarly the misclassifications can be used for correlation analysis and though is not an important analysis; interested readers are directed to the Table 1 in Appendix D.

Table 6.4 details the expressions and the computational models that have the best association between the RT and distance measure.

Table 6.4: Levels of association for various models and expressions for response time (RT) and distance measure

Expression	Model	Significance Level (S)	Correlation Value (c)
Angry	GAB	0.016	-0.212
Happy	GAB	0.043	-0.165
Fear	GABPCA	0.016	-0.209
Surprise	GAB	0.000	-0.281
Sad	None	-	-
Disgust	GAB	0.047	-0.178

As can be seen the model from the results in Table 6.4, the computational models based on GABOR filters had a significance values between the two variables (response time and distance measure) for all expressions, except for sad. With the expression sad none of the models suggests any association between the response time RT and the distance measure.

6.5.2 Results of the SDT Analysis

The result of the Signal Detection theory for all expressions is shown in Table 6.5.

Table 6.5: Signal Detection Theory results (d') for all expressions

Expression	Computational Models						Human Subjects
	RAW	RAWPCA	RAWCCA	GAB	GABPCA	GABCCA	
Angry	2.03	1.14	0.78	1.47	1.17	0.84	1.91
Happy	0.99	0.77	0.74	0.8	0.74	0.23	0.89
Fear	0.67	0.64	0.45	0.48	0.57	0.095	0.76
Surprise	3.23	2.53	3.02	3.36	2.55	1.96	3.98
Sad	0.54	0.5	0.24	0.4	0.42	0.17	0.66
Disgust	0.81	0.6	0.38	0.49	0.52	0.21	0.86

Note: The red font shows the highest value (absolute value) of *d-prime* for that expression and the numbers in blue font shows the second highest

In Table 6.5, numbers in red are the largest d' for that expression and the numbers in blue are the second largest. As discussed earlier, the values closer to zero indicate the ability to distinguish between the presences of a signal or not is least and a larger value of d' indicates perfect ability. It is interesting to note from Table 6.5 that the highest and the second highest d' are either the RAW models or the human subjects. The highest absolute value of d' for each expression and the model is shown in Table 6.6.

Table 6.6: Highest absolute values of d' for all expressions

Expression	**Model**	***d'* value**
Angry	RAW	2.03
Happy	RAW	0.99
Fear	Human Subjects	0.76
Surprise	Human Subjects	3.98
Sad	Human Subjects	0.66
Disgust	Human Subjects	0.86

The *d' (d-prime*) determines how well the model or the human subjects are able to select the correct stimuli while avoiding the incorrect ones, i.e. the ability to distinguish the expressive face from that of a neutral face. The values from Table 6.6 suggest that human subjects find it easy to distinguish surprise, fear, sad and disgust expressions from neutral expression in comparison to the other models. The RAW computational model seem to be the better of all the models in distinguishing angry and happy face images from the neutral face images. The highest value of *d' (d-prime*) is for the expression surprise and the least value is for expression sad. From the values of *d'* expression surprise seems to be easily distinguished from neutral compared to sadness or fear as found by others (Torro-Alves *et al.*, 2009).

6.6 Discussion

The comparisons of the human performance with that of the computational models led to interesting results. The models can be compared in terms of the overall performance for all expressions or for individual expressions. When human subjects performed the same type of classification as that of the models, the classification seemed to be exactly similar to that of the models. In terms of accuracy, the expressions surprise, happy and disgust were easier for classification while fear, angry and sad were harder. The average response time (RT) for the human subjects in classifying the different expressions is analogous to the distance measure of the data points from the classification hyper-plane. This indicates that the harder an expression on the face is to classify by human subjects, the closer it is to the classifying hyper-plane of the classifier. This result was obtained by performing a bi-variate correlation analysis between the average RT for human subjects and the distance measure of the face images from the hyper-plane of the classifier of the computational models. Here, a linear negative correlation was obtained for those entries which had this relationship with a significance level below or equal to 0.05. The significant p-values are shown in Table 6.4.

The other findings were that the surprise expression behaves differently to the other expressions from bi-variate analysis results. Here, irrespective of whether the images are pre-processed by Gabor filters or are RAW images, there seems to be a similarity in the ease/difficulty with which humans and models classify facial expressions.

For all expressions except sad, the results of the bi-variate analysis in Table 6.3 showed that the correlation between average RT of humans and the distance measure of the hyper-plane of the classifier for the computational models was significant mostly for models with Gabor filters. Out of the 11 models that are significantly correlated, 7 models are GABOR based and the remaining 4 are models. This suggests more similarities between computational models that use Gabor filtering for pre-processing and human subjects in terms of difficulty or ease of recognizing a facial expression.

The average classification accuracy of each of the six computational models described in this thesis is now compared with the accuracy obtained for human subjects across all expressions. Table 6.7 shows this result for comparison.

Table 6.7: Comparing performance – Six computational models versus human subjects

Expression	Average Accuracy (%)
RAW	88.26%
RAWPCA	80.91%
RAWCCA	75.05%
GAB	79.92%
GABPCA	79.45%
GABCCA	64.38%
Human Subjects	89.8%

The results in the table suggest that human subjects are better in facial expression recognition than any of the six computational models. The accuracy obtained by the RAW model is very close to the accuracy of human subjects. The accuracy obtained from other computational models such as RAWPCA, GAB and GABPCA have intermediate results. All classification accuracies are above chance.

The main point to be noted here is the dimensionality reduction methods used in the thesis such as the PCA and CCA in combination with Gabor pre-processing can reduce the original image dimensions to just a few components in comparison to the RAW models. This saves a lot of computational time and also memory space when handling larger databases. Although the raw images have managed to do better in classification accuracy this should be obvious as there has been no dimensionality reduction which could result in information loss. However, when the number of images increases dimensionality reduction will be a necessity and hence methods such as CCA with Gabor pre-processing may become more useful.

6.7 Conclusions

In this chapter, the experiment involving human subjects is explained and compares human performance with the performances of various computational models.

Susskind, Littlewort, Bartlett, Movellan, and Anderson (2007) compared human performance with computational models based on similarity and dissimilarity judgments. Most often, the average human accuracy is compared with the computational model. They realized that this could flaw the performance level as there is a variation within subgroups of human subjects in specific expressions. They also found that fear was poorly recognized by both humans and computer models. Since an average is taken collectively over the entire human subject group, it could masks the variations in individuals with different ethnic background, intellectual levels, age, gender, context and situation, familiarity, socio economic status, personality, attention, motivation, personal ability and emotional intelligence within the group of human subjects (Elfenbein *et al.*, 2002).

In another comparison test that compares the performance of human subjects with a computational classifier, Wimmer et al suggests that humans are not as good as some computational models (Wimmer *et al.*, 2007). They think that the poor performance by humans could be due to the database used. They used the Cohn-Kanade dataset and consider its posed expressions are the reason for this. Posing the happy expression is easier, but people are not sure how to pose expressions of fear, angry or disgust. They also conclude that lack of social circumstance or environment is a disadvantage as they think people's expressions change in response to the social communication and that is lacking in posed expressions. Hence, they think human subjects are not accurate in facial expression recognition.

Likewise, Dailey, Cottrell, Padgett and Adolphs (2002), found that the relative level of difficulty for the six prototypical expressions for their models was highly correlated to human performance. They found humans are good in classifying happy faces as are the models which complement my work. They suggest that the smile on the face aids faster detection, and the model finds it easy to detect smiles because of visual features for the happy expression that are obvious. They found that fear is one the most difficult expression for both humans and their computational model which is similar to the results of my work. However, as they use forced choice classification method, they also perceived that humans quite often confuse it with surprise and so does their neural network based computational model. They suggest that expression fear is often found to be difficult for classification because the perceptual similarity to other expressions and inherently difficulty to classify from other five expressions (Katsikitis, 1997; Ekman and Friesen, 1976).

Happiness and surprise were best detected by both humans and computers when Jinghai, Zilu and Youwei (2006) experimented with both. Complementing my work, they also found anger,

disgust, fear and sad were more difficult to classify. However, they also found that accuracy by humans was higher than computers.

On comparing the performance of human subjects with that of the computational models, there seems to be a lot of similarity. Surprise, happy and disgust were easier for classification, fear; angry and sad were harder for both humans and computational models. When the models were compared in terms of the classification performance, RAW performed the best for all models except for surprise. As there is no dimensionality reduction or information loss, it is not a surprising for the RAW model to perform very well. The other models, RAWPCA, GABPCA and GAB model perform equally well and that the RAWPCA uses just 97 components in comparison to the GABPCA, which uses a mere 22 components and manage to get reasonably good classification. The performance of RAWCCA and GABCCA are quite similar to one another and both do not do as well as the rest of the models, although they are way above chance results. However, RAWCCA uses 5 and GABCCA uses only 6 components.

From a direct comparison of the classification results, the GAB model seems to perform exceptionally well with expression surprise than with other expressions. Overall performance of surprise expression classification with all models has been extremely good. Although the GABCCA model uses just 5 components, the accuracy result is as high as 84.09%. This expression seems to be different from others in that it can be easily detected by any of the models and with very good accuracy.

From the bi-variate correlation analysis, the surprise expression seemed to have significance levels with almost all models. However, sad expression did not have significance levels for any of the computational models. There was a significant anti-correlation between the average RT of the human subjects and the distance measure of the classifier indicating that the images that needed longer time for the participants to classify had a larger average RT and that these images were closer to the classifying hyper-plane of the computational models and had a smaller distance measure.

Also in general, the results of the bi-variate correlation analysis suggests more number of GAB based models have significant correlation values when compared to the RAW models. This could mean that when models used images which are pre-processed by Gabor filters, they have a more similarities with human subjects in terms of difficulty or ease of recognizing a facial expression. The results from the SDT analysis show that humans are very good with classifications of surprise, disgust, fear and sad expression classification. The RAW model performs very well with surprise and angry.

Table 6.8 shows the order or rank of the scores obtained for different models. This table summarizes the rank of each model with respect to expression in the classification accuracy.

Table 6.8: Comparing Models – Rank or Order of the models in classification

Models / Expression	Human Subjects	RAW	RAWPCA	RAWCCA	GAB	GABPCA	GABCCA
Angry	5	4	6	5	3	5	2
Happy	2	1	2	2	2	2	3
Fear	4	5	3	3	4	3	6
Surprise	1	2	1	1	1	1	1
Sad	5	6	5	6	6	6	5
Disgust	3	3	4	4	5	4	4

In this thesis, although all datasets were balanced in terms of gender and all subjects' performance was very well recorded; in order to make a gender based comparison, it is impossible to model a female or a male computational system. Hence, no such comparisons can be made to study the effect of gender in the classification task.

A critical comparison with other similar studies does not provide complementary results in every aspect. A number of factors which have already been discussed make this evaluation more difficult. Recent work studied non frontal views for expression recognition (Hu *et al.*, 2008) which has not been explored before. Their experiment showed that non-frontal view is better than the frontal view for a computer to recognize facial expressions where the facial features points are manually marked. The best performance was at 45° for all expressions, except sad for which 60° gives the best accuracy.

In real life situations face to face communication is expected as non frontal view communication is considered to be impolite. Most of the datasets use more frontal face images in comparison to the number of non frontal face images. This however, could result in human perception bias which suggests that humans seem to be more sensitive to changes in the features of frontal face images than non frontal face views. This is a new area of research that is being explored.

Though ongoing research has concentrated on cognition and perception by humans, how humans recognize facial expressions is still not clear. With more and more biologically plausible computational techniques being developed, analyzing them in comparison to human performance can bring us a step closer to this understanding. Healthy humans are indeed still the sole winners when it comes to facial expression detection, they can fill in the gaps with obscured areas of the face and still detect the expression in a way that is difficult for any computational system.

Humans are quite robust and precise in detecting facial features, and can detect the expressions in spite of changes in identity and gender, race, shape of the face, texture, colour, with or without glasses or with variations in facial and scalp hair (Tian *et al.*, 2005). Computational systems on the other hand are still far less robust and do not have the capabilities to fill in the gaps or areas of the face if they are obscured and makes recognition harder. The systems intended to do accurate expressions recognition should take these factors into consideration.

CHAPTER SEVEN

Conclusions and Future Work

7.1 Introduction

This chapter summarizes the major findings and contribution of the work presented and also discusses future work. The thesis presented is inter-disciplinary and hence, facial expression recognition is discussed from social psychology and computational perspective. It has been a challenging experience to bring together the research and studies in these two different domains.

7.2 Summary

In Chapter 2, the psychological and computational aspect with relevance to facial expression recognition was discussed. The literature review in Chapter 2 discussed the universality of six prototypical facial expressions and suggested that expressions are innate. Psychological studies relevant to facial expression generation and the process of recognition were also discussed. Earlier studies have shown the existence of six universally accepted prototypical expressions: anger, happiness, fear, sadness, surprise and disgust. However, culture and regional variations do affect the process of exhibiting expressions. Emotional expressions can be controlled by the expresser, as they are voluntary in nature, but often we are not so good doing this (Ekman, 1973). The well known psychological model by Bruce and Young (1986) that explains separate pathways for facial identity and facial expression recognition was discussed in Chapter 2. This model was also compared with the neuropsychological model by Haxby et al (2000). Conflicting evidence from research work demonstrates both categorical and continuous perception of facial expressions by humans. Holistic processing and feature based processing involved in facial expression recognition has also been outlined. The neuropsychological perspective, the effect of brain injuries and trauma, lesions or disease on the recognition of facial expressions recognition has been reviewed. Sometimes, the effect is on the entire range of expression recognition or only on specific expressions depending on the area of lesions or injury or the disease. Different areas of the brain that are involved with processing of some expressions and the diseases that cause impairment of specific expressions have been studied (Adolphs *et al.*, 2000).

The second half of the Chapter 2 was devoted to computational models of facial expression recognition. Methods for feature extraction commonly used and classification were studied. Issues relevant to producing an ideal automatic facial expression recognition system were also

discussed. Factors relevant to producing an ideal database are also mentioned. However, the reader is directed to the area of neuropsychology that processes images in general and how the psychology and neuropsychology can be better understood by developing biologically plausible models for feature extraction. This chapter concluded with a discussion of the requirements of an automatic facial expression recognizing computational model that ideally matches human performance in recognizing facial expressions.

In Chapter 3, the computational methodologies used in this thesis were described in detail. A biologically plausible technique for feature extraction, in the form of Gabor filters, which are thought to mimic the simple cells of the pre-processing technique, was used for feature extraction. As face images are often of higher dimensionality, dimensionality reduction methods: PCA, CCA and LDA were also presented in Chapter 3. These methods may remove redundancies in the dataset by using the correlations within the data. When using a PCA projection the number of dimensions to which the original dataset is reduced is such that it retains 95% of the total variance of the dataset. However, with CCA the true dimensionality of the data called as Intrinsic Dimension which may be much lesser than the original dimension needs to be estimated. Finally, classification using an SVM was also studied.

Using an effect size analysis it is possible to identify those pixels in a face image that show a high discrimination between any two expressions. The method for performing this was discussed in this Chapter. This compliments research that describes the regions of the face that is associated with different expressions. An analysis of these methods with actual datasets was discussed in Chapter 4 and Chapter 5.

Chapter 4 uses a small dataset with only two facial expressions, neutral and happiness. The methods described in Chapter 3 were applied to this dataset. The results were interesting and are summarized below.

- The best models were: RAW, GAB and GABCCA11. These gave error rates of only one from 20 in Test Set A and 4 from 20 in Test Set B.
- GABCCA11 did remarkably well as it used only 11 components.
- The PCA based models did not perform well.
- The LDA based classifier did not perform well.

These results encouraged extending these experiments to larger dataset and with all six universally accepted prototypical expressions.

In Chapter 5 the BINGHAMTON dataset was introduced. This is a larger dataset and includes all six prototypical expressions. The first experiment performed was an *Effect Size* analysis to identify those pixels that most clearly discriminate between two expressions. Here, the analysis was performed on one of the basic prototypical expressions and the neutral one. The next experiment involved identifying those principal components that encoded particular expressions. It was found that some components were significant for more than one expression. Using these

components it was possible to morph a face from a neutral expression to an extreme prototypical expression. The third experiment reported here was an extensive analysis of representation and classification of these six expressions. The techniques described in Chapter 3 were applied to all six expressions and all six models. The major findings were as follows:

- The easiest expression to recognize was surprise and the hardest were sad and angry.
- The RAW model performed best.
- The RAWPCA was the best model with reduced dimensionality.
- The RAWCCA model did not do quite as well but only used 5 components.
- With the exception of surprise, the Gabor based models did not do so well.

This chapter is concluded by a discussion in the current research in this field and makes a critical evaluation of the performance in classifying facial expression.

The final experiment in the thesis takes the BINGHAMTON dataset and a set of human subjects who undertook a forced choice expression recognition task.

In Chapter 6, a study with human subjects in classification of facial expressions is reported. The application TESTBED enabled recording the response time and the classification accuracy. This was used to compare the performance of various computational models with that of human subjects. A bi-variate correlation analysis and signal detection theory was used to analyze and compare the results from computational model and human subjects. This section concluded with a critical evaluation of existing and current studies on performance of human subjects in facial expression recognition. The major findings were as follows:

- The human subjects found the expression surprise the easiest to identify and the angry expression classification the hardest.

- The human accuracy was similar to the best of the computational models.

- The Bi-variate analysis indicated that the Gabor based models showed greatest similarity to human performance.

- There was a negative correlation between the average RT of human subjects and the distance measure for the computational models. This suggests that the harder the image classification is for humans, the closer it is to the classification boundary.

- The d' analysis did not provide any consensus; however, the SDT analysis suggested that human subjects distinguished surprise, disgust, fear, and sad from neutral expressed better than the other models. The RAW models seem to distinguish angry and happy faces from neutral expression better than other models.

7.3 Contribution

The contributions in the field of facial expression recognition made by this thesis are the following:

- The thesis confirms that the surprise and happy expressions are the easiest to identify for both humans and computational models. When human subjects performed the same type of classification as done by the models, the performance across the different expressions seemed to be similar to that of the models. Surprise, happy and disgust were easier for classification, fear, angry and sad were harder.

- The bi-variate correlation analysis suggests that Gabor based computational models may be more similar to human subjects in facial expression classification. More number of GAB based models showed significant levels in comparison to RAW models and suggests correlation with human subjects in terms of difficulty or ease of recognizing a facial expression.

- For expression surprise, the almost all RAW and Gabor based models showed significant correlations.

- The PCA and CCA can reduce the original dataset to a very small dimension and still produce effective classification. The RAW model performed the best for all expressions except for surprise. It can also be noted that the RAWPCA, GABPCA and GAB model perform equally well and that the RAWPCA uses 97 components in comparison to the GABPCA, which uses a mere 22 components. The performance of RAWCCA and GABCCA are quite similar to one another and both do not do as well as the rest of the models, although they are way above chance results. The main point to be noted is that the classification results are obtained with just few components - RAWCCA uses 5 and GABCCA uses only 6.

- The GABCCA model did not perform well on the BINGHAMTON dataset but performed well with the FERET dataset and hence it is very hard to make general conclusions.

- The pixel based *Effect size* analysis showed for the first time those areas of the face that actually discriminate a particular expression from a neutral face. This analysis may enable us to better understand the human facial features involved and the generation of the expressions.

- The PCA analysis showed that different principal components encoded the various expressions. Some components encoded more than one expression and perhaps, it could be suggestive of the confusions in classification of these expressions by human subjects.

- Using PCA components it was shown that a neutral face can be morphed to an extreme prototypical expression.

- The facial features are non linear (Jarudi and Sinha, 2003) and a non linear CCA method reduced the dimensions of the face images better than a linear technique such as PCA (Buchala *et al.*, 2005; Buchala *et al.*, 2004a). In addition a non linear Gabor filtering method (Kruizinga and Petkov, 1999; Shen and Bai, 2006) combined with non linear CCA has also managed to get a very small number for the ID.

- On comparing the classification accuracy for every expression across all models, surprise, happy and disgust expression recognition seemed to be easier than fear, angry and sad.

- The GAB model performs well with expression surprise than with other expressions. Overall performance of surprise expression classification with all models has been extremely good. By using just 6 components with GABCCA model, the accuracy result is as high as 84.09%. This expression seems to be different from others in that it can be easily detected by any of the models, by very good accuracy.

- The hypothesis that the average response time (RT) for the human subjects in classifying the different expressions is analogous to the distance measure of the data points from the classification hyper-plane was verified. This means the harder an expression on the face is to classify by human subjects, it is closer to the classifying hyper-plane of the classifier. This is obtained by using bi-variate correlation analysis. Here, a linear negative correlation was obtained for those entries which had this relationship had a significance level below 0.05.

- The signal detection theory (SDT) or the *d-prime* determined how well the model or the human subjects are in making the classification of an expressive face from a neutral one. On comparison, human subjects are better in classifying surprise, disgust, fear, and sad expressions. The RAW computational model provides better able to distinguish angry and happy expressions.

7.4 Future Work

7.4.1 Morphing of facial expressions using PCA

In Chapter 5, PCA and LDA has been used together to find the component capable of coding specific expressions. By finding the encoding powers of the components, the 26th component has found to code angry, sad and disgust expressions. Likewise, the 7th component is significant for happy and fear expressions and 3rd component is significant for surprise. By reconstructing face images using different proportions of these components, facial expression morphing was achieved. This could well be used to study changes in facial expressions such as micro expressions, lie detection and threat detection in humans.

7.4.2 Psychological plausibility of computational models

Calder et al (2001) have shown that their system based on PCA has a lot of similarity to human performance in a forced-choice experiment and later Dailey et al (2002) have shown that their model based on Gabor filtering and PCA was more biologically plausible computational model and not only shows similarity to human forced choice performance but also supports both categorical and multidimensional theories of facial expression recognition and perception. These types of experiments could be extended to other dimensionality reduction methods such as Independent Component Analysis, ISOMAP, and may be even in combination with classifiers such as SVM, Linear Discriminant Analysis and compare them with human subject's performance on the facial expression related experiments.

7.4.3 Gabor filtering methods

This thesis involved experiments that have used Gabor filtering for pre-processing. The Gabor filters were applied across the entire face and later, *L2 max norm superposition* method was used to produce the output of the filter bank. Though this has been commonly used, it would be interesting to see the results by averaging the output of all the filters and follow it up with any dimensionality reduction methods such as PCA or CCA. A holistic approach has been followed here; however, an expert on facial expression recognition could select fiducial points that would

enable better recognition. The Gabor filters could be applied only at these points and it would be interesting to see the performance of this pre-processed data set.

7.4.4 Gender based expression dataset

The studies with human subject in tasks related to facial expressions have revealed some interesting results. Wagner, MacDonald and Manstead (1986) have found that female subjects are better in displaying facial expression than male. They found that female face images displayed neutral and surprised expressions much more accurately than male face images. Goos and Silverman (2002) found that men pose angry expression better than females. Could this influence the results of the experiments with human subjects? Does that mean that all datasets researchers use need to be a balanced set as used in this thesis for all experiments with human subjects. The effects of an unbalanced set may not change the performance of a computational model as it is difficult to model a gender based computational system. Whilst performing all the experiments a balanced set (in terms of gender) has been used and the analysis can be repeated to obtain gender based classification results for human subjects.

7.4.5 Effects of Age on facial expression recognition

There have been experiments conducted to study the effect of age in the recognition of facial expression. Studies by Suzuki, Hoshino, Shigemasu, Kawamura (2007) have showed that age affects the perception of facial expressions and emotions as such. In particular, they found that there is age-related decline in sadness recognition and age-related improvement in disgust recognition. Vasiliki and Louise (2008) report age related impairments in recognition of negative expressions in particular. This could be an affect of socio-environmental factors and hence, the age of the participants in human subject experiments could influence the average performance accuracy.

7.4.6 Dynamic Expression database

The experiments in this thesis used static grey scale images and hence, it would be interesting to repeat the experiments with a dynamic dataset. This dataset should include image sequences of the individuals in the dataset who change the facial expression from neutral to one of the basic prototypical expression. The performance of the human subjects can be then be compared to the

classification accuracy of the computational models. As recent research has suggested (Zheng *et al.*, 2009; Fasel and Luettin, 2003), the expression recognition could include voice or audio based recognition. Experiments using datasets that portray an actual social environment such as facial expression that occur during a conversation rather than on its own could be interesting.

7.4.7 Other expressions

It would be interesting to include other expressions such as deceit and contempt.

The suggestion for future work in this chapter can be extended in various dimensions and can lead to further PhD work in its own right.

References

Abdi, H. (2007) Signal Detection Theory (SDT). Dallas, University of Texas.

Adolphs, R. (2002) Recognizing Emotion From Facial Expressions: Psychological and Neurological Mechanisms. *Behavioral and Cognitive Neuroscience Reviews,* 1**,** 21-62.

Adolphs, R., Damasio, H., Tranel, D., Cooper, G. & Damasio, A. R. (2000) A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *Journal of Neuroscience,* 20**,** 2683-2690.

Adolphs, R., Damasio, H., Tranel, D. & Damasio, A. R. (1996) Cortical systems for the recognition of emotion in facial expressions. *Journal of Neuroscience* 16**,** 7678-7687.

Adolphs, R., Travel, D., Damasio, H. & Damasio, A. R. (1994) Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature,* 312.

Altarriba, J., Basnight, D. M. & Canary, T. M. (2003) Emotion representation and perception across cultures. *Online Readings in Psychology and Culture (Unit 4, Chapter 5).* Center for Cross-Cultural Research, Western Washington University, Bellingham, Washington USA W. J. Lonner, D. L. Dinnel, S. A. Hayes, & D. N. Sattler (Eds.).

Aluport, F. H. (Ed.) (1924) *Social Psychology,* Boston, Houghton Mifflin.

Asirvatham, A. (2002) Script Segmentation of Multi-script Documents. .

Bannerman, R. L., Milders, M., Gelder, B. D. & Sahraie, A. (2009) Orienting to threat: faster localization of fearful facial expressions and body postures revealed by saccadic eye movements. *Proceedings of the Royal Society B*.

Barlow, H. B. (1989) Unsupervised Learning. *Neural Computation,* 1**,** 295-311.

Bartlett, M. S., Hager, J. C., Ekman, P. & Sejnowski, T. J. (1999) Measuring facial expressions by computer image analysis. *Psychophysiology,* 36**,** 253-263.

Bartlett, M. S., Littlewort, G., Braathen, P., Sejnowski, T. J. & Movellan, J. R. (2005) A Prototype for Automatic recognition of spontaneous facial actions. *Advances in Neural Information processing systems,* 15**,** 1271-1278.

Bassili, J. (1979) Emotion recognition: the role of facial motion and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology,* 37**,** 2049-2059.

Belhumeur, P. N., Hespanha, J. P. & Kriegman, D. J. (1997) Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on patternAnalysis and Machine Intelligence,* 19**,** 711-720.

Bell, C. (Ed.) (1844) *Anatomy and Philosophy of Expression as Connected With the Fine Arts. (7th ed., rev.)* London: G. Bell & Sons.

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

Blanz, V., Sch¨Olkopf, B., B¨Ulthoff, H., Burges, C., Vapnik, V. & Vetter, T. (1996) Comparison of view-based object recognition algorithms using realistic 3D models. *Proc. Int. Conf. on Artificial Neural Networks* 251-256.

Browndyke, J. N. (2002) Neuropsychosocial Factors in Emotion recognition: Facial expression. *Neuropsychology Central.*

Bruce , V. & Young, A. W. (1986) Understanding face recognition *British Journal of Psychology,* 77 305-327.

Buchala, S., Davey, N., Frank, R. J. & Gale, T. M. (2004a) Dimensionality reduction of face images for gender classification. *International IEEE conference on Intelligent Systems.* Varna, Bulgaria.

Buchala, S., Davey, N., Frank, R. J. & Gale, T. M. (2004b) Dimensionality reduction of face images for gender classification. *Intelligent Systems, Proceedings. 2004 2nd International IEEE Conference,* 1**,** Page(s): 88 - 93 Vol.1.

Buchala, S., Davey, N., Frank, R. J. & Gale, T. M. (2005) Analysis of linear and nonlinear dimensionality reduction methods for gender classification of face images. *International journal of Systems Science,* 36**,** 931-942.

Buchala, S., Davey, N., Frank, R. J., Gale, T. M., Loomes, M. & Kanargard, W. (2004c) Gender Classification of Face Images: The Role of Global and Feature-Based Information. *International Conference on Neural Information Processing* Calcutta, India.

Buciu, I., Kotropoulos, C. & Pitas, I. (2003) ICA and Gabor representation for facial expression recognition. *IEEE International Conference on Image Processing.*

Buck, R. W. & Duffy, R. J. (1980) Nonverbal behavior and the theory of emotion: The facial feedback hypothesis. *Journal of Personality and Social Psychology,* 38**,** 811-824.

Buenaposada, J. M., Enrique Mu˜Noz, E. & Baumela, L. (2008) Recognising facial expressions in video sequences. *Pattern Analysis and Application,* 11**,** 101-116.

Byrne, A. & Eysenck, M. W. (1995) Trait anxiety, anxious mood, and threat detection. *Cognition and Emotion* 9**,** 549-562.

Calder, A., Young, A., Perrett, D., Etcoff, N. & Rowland, D. (1996) Categorical perception of morphed facial expressions. *Visual Cognition,* 3**,** 81-117.

Calder, A. J., Butron, M., Miller, P., Young, A. W. & Akamatsu, S. (2001) A Principal component analyisis of facial expressions. *Vision Research,* 41**,** 1179-1208.

Calder, A. J., Young, A. W., Keane, J. & Dean, M. (2000) Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance,* 26**,** 527-51.

Calder, A. J., Young, A.W., (2005) Understanding the recognition of facial identity and expresssion. *Neuroscience Nature Reviews,* 6.

Camastra, F. & Vinciarelli, A. (2001) Intrinsic Dimension estimation of data: An approach based on grassberger-procaccia's algorithm. *Neural processing letters,* 14**,** 27-34.

Cao, Y., Zhengi, W., Zhao, L. & Zhou, C. (2005) Expression Recognition Using Elastic Graph Matching. *Lecture Notes in Computer Science - Affective Computing and Intelligent Interaction* 3784**,** 8-15.

Carvajal, F., Vidriales, R., Rubio, S. & Martin, P. (2004) Effect of the changes in facial expression and/or identity of the model on a face discrimination task. *Psicothema,* 16**,** 587-591.

Chan, V. (2009) The perception and recognition of emotions and facial expressions. *Journal of Undergraduate Life Sciences,* 3.

Chang, C., C. & Lin, C. J. (2001) LIBSVM: a library for support vector machines.

Chih-Wei Hsu, C.-C. C., and Chih-Jen Lin (2008) A Practical Guide to Support Vector Classification.

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

Cohen, J. (1988) *Statistical power analysis for the behavioural sciences,* Hillsdale , New Jersey, Lawrence Earlbaum Associates.

Cohn, J. F., Zlochower, A., Lien, J., Wu, Y. & Kanade, T. (1997) Automated face coding: a computer - vision based method of facial expression analysis. *European Conference on facial expression measurement and meaning.* Salzburg, Austria.

Comon, P. (1994) Independent Component Analysis- A new Concept? *Signal Processing,* 36**,** 287-314.

Cortes, C. & Vapnik, V. (1995) Support Vector Networks. *Machine Learning,* 20**,** 273-297.

Crocker, V. & Mcdonald, S. (2005) Recognition of emotion from facial expression following traumatic brain injury. *Brain Injury,* 19**,** 787-799.

Dailey, M. & Cottrell, G. (1999) PCA=Gabor for facial expression recognition. Institution UCSD.

Dailey, M., Cottrell, G. & Adolphs, R. (2000) A six-unit network is all you need to discover happiness. *Annual Conference of the Cognitive Science Society.* Mahwah, NJ. Erlbaum.

Dailey, M., Cottrell,G., Padgett, C., Adolphs, R. (2002) EMPATH: A Neural Network that Categorizes Facial Expressions. *Journal of Cognitive neuroscrience,* 14**,** 1158-1173.

Darwin, C. (Ed.) (1872) *The expression of emotion in man and animals,* New York, D. Appleton and Co.

Daugman, J. G. (1980) Two-Dimensional Spectral Analysis of Cortical Receptive Field Profile. *Vision Research,* 20.

Daugman, J. G. (1985) Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two dimensional visual cortical filters. *Journal of Optical.Society of America A,* 2.

Demartines, P. & Hérault, D. J. (1997a) Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets *IEEE Transactions on Neural Networks,* 8**,** 148-154.

Demartines, P. & Hérault, D. J. (1997b) Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets *IEEE Transactions on Neural Networks,* 8**,** 148-154.

Deng H. B., J., L.W., Zhen., Jian-Cheng Huang (2005) A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA. *International Journal of Information Technology* 11

Derpanis, K. (2007) Gabor Filters. York, York Univeristy.

Donato, G., Bartlett, S., Hager, C., Ekman, P. & Sejnowski, J. (1999) Classifying facial actions. *IEEE Transaction on Pattern Analysis and Machine Intelligence,* 21**,** 974-989.

Donoho, D. L. (2000) High-dimensional data analysis: The curses and blessings of dimensionality. Los Angeles.

Drakos, N. & Moore, R. (1999) Document Segmentation using Texture Segmentation

Edwards, G. J., Cootes, T. F. & Taylor, C. J. (1998) Face Recognition Using Active Appearance Models. *Proceedings of the European Conference on Computer Vision,* 2**,** 581-695.

Ekman, P. (Ed.) (1973) *Cross-cultural studies of facial expression. ,* New York, Academic.

Ekman, P. (2003) Darwin, Deception, and Facial Expression. *Annals New York Academy of Sciences.* San Francisco, California 94143, USA, Department of Psychiatry, University of California, San Francisco.

Ekman, P. & Friesen, W. V. (1971) Constants across cultures in the face of the emotion. *Journal of Personality and Social Psychology,* 17.

Ekman, P. & Friesen, W. V. (1975) *Unmasking the face. A guide to recognizing emotions from facial clues.,* New Jersey, Prentice-Hall.

Ekman, P. & Friesen, W. V. (1976) Pictures of facial affect. . Palo Alto, CA: , Consulting Psychologists Press.

Ekman, P. & Friesen, W. V. (1978) *Manual for the Facial Action Coding System*, Palo Alto: Consulting Psychologists Press.

Ekman, P., Friesen, W. V. (1986) A new pan-cultural facial expression of emotion. *Motivation & Emotion,* 10 159-168.

Ekman, P. & Oster, H. (1979) Facial expressions of emotion. *Annual Review of Psychology,* 30**,** 527-554.

Elfenbein, H. A., Marsh, A. & Ambady, N. (2002) Emotional Intelligence and the recognition of emotion from the face.

Ellis, H. D. (1975) Recognizing faces. *British Journal of Psychology,* 66**,** 409-426.

Escobar, M. J. & Ruiz-Del-Solar, J. (2002) Biologically-based face recognition using Gabor filters and log-polar images. *International Joint Conference on Neural Networks* Honolulu, USA.

Essa, I. A. & Pentland, A. P. (1997) Coding, Analysis, Interpretation, and Recognition of Facial Expressions *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 19**,** 757-763.

Etcoff, N. L. & Magee, J. J. (1992) Categorical perception of facial expressions. *Cognition,* 44**,** 227-240.

Etemad, K. & Chellappa, R. (1997) Discriminating analysis for recignition of human face images. *Journal of Optical.Society of America A,* 14**,** 1724-1733.

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

Fasel, B., Monay, F. & Gatica- Perez., D. (2004) Latent Semantic Analysis of Facial action codes for automatic facial expression recognition *ACM international Workshop multimedia information retrieval*

Fellenz, A. W., Taylor, J. G., Tsapatsoulis, N. & Kollias, S. (1999) Comparing template-based, feature-based and supervised classification of facial expressions from static images. *Computational Intelligence and Applications.*

Fisher, B. (2001) Fisher linear discriminant and dataset transformation.

Fisher, R. A. (1936) The use of mutliple measures in anatomical problems. *Annals of Eugenics,* 7**,** 179-188.

Fodor, I. K. (2002) A survey of dimension reduction techniques. Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, L-560, Livermore, CA 94551.

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A. & Dutton, K. (2000) Facial Expressions of Emotion: Are Angry Faces Detected More Efficiently? *Cognition and Motion,* 14.

Fridlund, A. (1994) *Human Facial Expression: An Evolutionary View,* San Diego, CA: Academic Press.

Gisalason, S. (2007) The Brain centre. *Human Brain.*

Goos, L. M. & Silverman, I. (2002) Sex related factors in the perception of threatening facial expressions. *Journal of Nonverbal Behavior,* 26**,** 2.

Grassberger, P. & Proccacia, I. (1983) Measuring the strangeness of strange attractors. *Physica D,* 9.

Grigorescu, S. E., Petkov, N. & Kruizinga, P. (2002) Comparison of Texture features based on Gabor filters. *IEEE Transactions on Image Processing,* 11.

Gunduz, A. & Krim, H. (2003) Facial feature extraction using topological methods. *International Conference on Image Processing.*

Gunes, H. & Piccardi, M. (2005a) Affect recognition from face and body : early fusion versus late fusion. *IEEE International Conference on Systems, Man and Cybernetics.*

Gunes, H. & Piccardi, M. (2005b) A Bimodal face and body gesture database for automatic analysis of human nonverbal affective behaviour *International conference of pattern recognition.*

Gunopulos, D. (2001) Dimensionality Reduction Techniques. *DIMACS Summer School Tutorial on New Frontiers in Data Mining.*

Hager, J. C. (2006) Dataface. *Human face.*

Hampton, C., Purcell, D. G., Bersine, L., Hansen, C. H. & Hansen, R. D. (1989) Probing "pop out" : Another look at the face in the crowd effect. *Bulletin of the Psychonomic Society,* 27**,** 563-566.

Hansen, C. H. & Hansen, R. D. (1988) Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology,* 54**,** 917-924.

Hargrave, R., Maddock, R. J. & Stone, V. (2002) Impaired Recognition of Facial Expressions of Emotion in Alzheimer's Disease *Journal of Neuropsychiatry and clinical neurosciences,* 14.

Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. (2000) The distributed human neural system for face perception. *Trends in Cognitive Science* 4**,** 223-233.

Hong, H., Neven, H. & Malsburg, C. V. (1998) Online Facial Expression Recognition Based on Personalized Galleries. *International Conference on Automatic Face and Gesture Recognition.*

Hu, Y., Zeng, Z., Yin, L., Wei, X., Tu, J. & Huan, T. S. (2008) A Study of Non-frontal-view Facial Expressions Recognition. *IEEE International Conference on Pattern Recognition*.

Hubel, D. H. & Wiesel, T. N. (1968) Receptive fields and functional architecture of the monkey striate cortex. *Journal of Physiology,* 195**,** 215-243.

Hubel, D. H. & Wiesel, T. N. (1995) Eye, Brain and Vision. http://hubel.med.harvard.edu/b17.htm#simp.

Hyvärinen, A. & Oja, E. (2000) Independent Component Analysis: Algorithms and Applications. *Neural Networks,* 13(4-5)**,** 411-430.

Ioannou, S., Raouzaiou, A., Tzouvaras, V., Mailis, T., Karpouzis, K. & Kollias, S. (2005) Emotion recognition through facial expression analysis based on a neurofuzzy method. *Neural Networks,* 18**,** 423-435.

Izard, C. E. (Ed.) (1977) *Human Emotions,* New York, Plenum.

Izard, C. E. (Ed.) (1978) *On the ontogenesis of emotions and emotion-cognitive relationships in infancy.,* New York: Plenum.

Jain, A. K. (1988) *Fundamentals of Digital Image Processing*.

Jain, A. K. & Farrokhnia, F. (1991) Unsupervised texture segmentation using Gabor filters. *Pattern Recognition,* 24.

Jarudi, I. N. & Sinha, P. (2003) Relative Contributions of Internal and External Features to Face Recognition. Massachusetts institute of technology — Artificial Intelligence laboratory.

Jenness, A. (1932) The recognition of facial expressions of emotions. *Pyschological Bulletin* 29.

Jinghai, T., Zilu, Y. & Youwei, Z. (2006) The contrast analysis of facial expression recognition by human and computer. *International Conference on Signal Processing.*

Jolliffe, I. T. (2002) Principal Component Analysis. 2 ed. New York, Springer- Verlag.

Jones, P. & Palmer, L. (1987) An evaluation of the two-dimensional gabor filter model of simple receptive fields in the cat striate cortex. *Journal of Neurophysiology,* 58.

Kanade, T., Cohn, J. F. & Tian, Y. (2000) Comprehensive Database for Facial Expression Analysis *IEEE International Conference on Automatic Face and Gesture Recognition* 46–53.

Katsikitis, M. (1997) The classification of facial expressions of emotion: A multidimensional scaling approach. *Perception*.

Kirita, T. & Endo, M. (1995) Happy face advantage in recognizing facial expressions. *Acta Psychologica,* 89.

Kirkpatrick, S. W., Bell, F. E., Johnson, C., Perkins, J. & Sullivan, L. A. (1996) Interpretation of facial expressions of emotion: the influence of eyebrows. *Genetic, Social, and General Psychology Monographs* 122**,** 405-423.

Kobayashi, H. & Hara, F. (1997) Facial Interaction between animated 3D face robot and human beings. *International Conference on Systems, Man and Cybernatics.* Orlando.

Kohonen, T. (Ed.) (2001) *Self Organizing Maps*, Springer-Verlag.

Kruizinga, P. & Petkov, N. (1999) Non linear operator for oriented texture. *IEEE Transactions on image processing,* 8.

Kulikowski, J. J., Marcelja, S. & Bishop, P. O. (1982) Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biological Cybernetics* 43**,** 187-198.

Kwak, K. C. & Pedrycz, W. (2005) Face recognition using a fuzzy fisherface classifier. *Pattern Recognition,* 38 1717 – 1732.

Langfeld, H. S. (1918) The Judgment of Emotions from Facial Expressions. *Journal of Abnormal and Social Psychology,* 13**,** 172-184.

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

Lee, C. & Elgammal, A. (2005) Facial expression analysis using decomposable generative models. *IEEE International workshop on analysis and modelling of faces and gestures.*

Leloglu, U. M. (1994) Artificial versus natural stereo depth perception. *Life to Artificial Intelligence.*

Leppänen, J. M. & Hietanen, J. K. (2003) Affect and Face Perception: Odors Modulate the Recognition Advantage of Happy Faces. *Emotion,* 3**,** 315-326.

Ley, R. & Bryden, M. (1979) Hemispheric differences in recognizing faces and emotions. *Brain and Language,* 7**,** 127-138.

Liejun, W., Xizhong, Q. & Taiyi, Z. (2009) Facial Expression Recognition Using Improved Support Vector Machine by Modifying Kernels *Information Technology Journal* 8 595-599.

Lien, J. (1998) Automatic recognition of facial expression using hidden markov models and estimation of expression intensity. *The Robotics Institute.* CMU.

Lisetti, L. C. & Schiano, D. J. (2000) Automatic facial expression Interpretation : Where Human-Computer Interaction, Artificial Intelligence and Cognitive Science Intersect. *Pragmatics and Cognition (Special issue on Facial Information Processing: A Multidisciplinary Perspective),* 8**,** 185-235.

Littlewort, G., Bartlett, M. S., Fasel, I., Susskind, J. & Movellan, J. (2006) Dynamics of facial expression extracted automatically from video. *Journal of Image and Vision Computing,* 24**,** 615-625.

Liu, C. & Wechsler, H. (2003) Independent component analysis of Gabor features for face recognition. *IEEE Transactions on Neural Networks,* 14.

Liu, W. & Wang, Z. (2006) Facial Expression Recognition Based on Fusion of Multiple Gabor Features. *IEEE International Conference on Pattern Recognition*

Lucey, S., Ashraf, A. B. & Cohn, J. F. (Eds.) (2007) *Invesitigating spontaneous facial action recognition through AAM representations of the face*, I-Tech Education and publishing.

Lyons , M., Akamatsu , S., Kamachi , M. & Gyoba, J. (1998) Coding Facial Expressions with Gabor Wavelets. *International Conference on Face and Gesture Recognition.* Nara, Japan.

Lyons, M., Budynek, J., Akamatsu, S. (1999) Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 21.

Marcelja, S. (1980) Mathematical Description of the Responses of Simple Cortical Cells. *Journal of Optical Society of America A,* 70.

Martinez, A. M. & Kak, A. C. (2001) PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 23**,** 228-233.

Mase, K. (1991) Recognition of facial expressions from optical flow. *IEICE Transactions (Special Issue on Computer Vision and its Applications),* 74**,** 3474-3483.

Matsumoto, D. (Ed.) (2001) *Culture and Emotion,* New York, Oxford University Press.

Matsumoto, D. (2007) Culture, context, and behavior. *Journal of Personality,* 75**,** 1285-1319.

Matsumoto, D., Keltner, D., O'sullivan, M. & Frank, M. G. (2007) What's in a face? Facial expressions as signals of discrete emotions.

Matsumoto, D. & Willingham, B. (2009) Spontaneous Facial Expressions of Emotion in Congenitally and Non-Congenitally Blind Individuals. *Journal of Personality and Social Psychology,* 96.

Mcculloch, D. (2005) A Investigation into Novelty Detection. *Engineering Mathematics.* University of Bristol.

Milders, M., Sahraie, A. & Logan, S. (2008) Minimum presentation time for masked facial expression discrimination. *Cognition and Emotion,* 22**,** 63-81.

Movellan, J. R. (2002) Tutorial on Gabor Filters.

O'toole, A. J., Deffenbacher, K. A., Valentin, D. & Abdi, H. (1994) Structural Aspects of face recognition and the other race effect. *Memory and Cognition,* 22**,** 208-224.

Ohman, A. (Ed.) (1993) *Fear and anxiety as emotional phenomenon: Clinical phenomenology, evolutionary perspec tives, and informat ion-processing mechanisms.,* New York, Guildford Press.

Ohman, A., Lundqvist, D. & Esteves, F. (2001) The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology,* 80**,** 381-396.

Oliva, A., Balas, B. & Kemp, C. (2005) Signal Detection Theory.

Oster, H. (Ed.) (1978) *Facial expression and affect development.,* New York, Plenum.

Padgett, C. & Cottrell, G. (Eds.) (1996) *Representing face images for emotion classification,* Cambridge, MIT press.

Padgett, C. & Cottrell, G. (1998) A simple neural network models categorical perception of facial expressions *Conference on Cognitive Science.* Mahwah, NJ. Erlbaum.

Padgett, C., Cottrell, G. & Adolphs, R. (1996) Categorical perception in facial emotion classification. *Proceedings of the 18th Annual Conference of the Cognitive Science Society.* . Hillsdale, New Jersey.

Pantic, M. (2005) MMI Database.

Pantic, M. & Bartlett, M. S. (2007) Machine analysis of Facial expressions. *Face Recognition.*

Pantic, M. & Patras, I. (2006) Dynamics of facial expression: recognition of facial actions and temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man and Cybernetics Part B,* 36**,** 433-449.

Pantic, M. & Rothkrantz, L. J. M. (2000) Automatic Analysis of Facial Expressions: The State of the Art *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22.

Pantic, M. & Rothkrantz, L. J. M. (2004) Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man and Cybernetics Part B,* 34**,** 1449-1461.

Pantic, M. & Rothkrantz, L. J. M. (2004a) Case based reasoining for user profile recognition of emotions from the face images. *ACM international conference on multimedia*

Philips, P. J., Wechsler, H., Huang, J. & Rauss, P. (1998) The FERET evaluation methodology for face recognition algorithms. *Image and Vision Computing,* 16**,** 295-306.

Pollen, D. A. & Ronner, S. F. (1981) Phase relationships between adjacent simple cells in the visual cortex. *Science,* 212**,** 1409-1411.

Rizvi, S., Philips, P. J. & Moon, H. (1998) A Verification protocol and statitical performance analysis for face recognition algorithms. *Computer Vision and Pattern Recognition.*

Rosenblum, M., Yacoob, Y. & Davis, L. (1996) Human expression recognition from motion using a radial basis function network architecture. *IEEE Transactions on Neural Networks,* 7**,** 1121-1138.

Russell, J. A. (1980) A circumplex model of affect. *Journal of Personality and Social Psychology,* 39**,** 1161-1178.

Russell, J. A. & Bullock, M. (1986) Fuzzy concepts and the perception of emotion in facial expressions. *Social Cognition,* 4**,** 309-341.

Russell, J. A., Lewicka, M. & Niit, T. (1989) A cross-cultural study of a circumplex model of affect. *Journal of Personality and Social Psychology,* 57.

Schiano, D. J., Ehrlich, S. M. & Sheridan, K. (2004) Categorical Imperative NOT: Facial Affect is Perceived Continuously. *ACM Conference on Human Factors in Computing Systems.* Vienna, Austria.

Schwaninger, A., Wallraven, C., Cunningham, D. W. & Chiller-Glaus, S. D. (2006) Processing of facial identity and expression: a psychophysical, physiological, and computational perspective. *Progress in Brain Research,* 156.

Schwartz, G. M., Izard, C. E. & Ansul, S. E. (1985) The 5-month-old' s ability to discriminate facial expressions of emotion. *Infant Behavior and Development,* 8**,** 65-77.

Sebe, N., Lew, M. S., Cohen, I., Sun, Y., Gevers, T. & Huang, T. S. (2004) Authentic facial expression analysis. *IEEE International conference on automatic face and gesture recognition.*

Serrano, J. M., Iglesias, J. & Loeches, A. (1992) Visual discrimination and recognition of facial expressions of anger, fear, and surpris e in 4- to 6-month-old infants. *Developmental Psychobiology,* 25**,** 411-425.

Shen, L. (2005) Recognizing Faces- An approach based on Gabor Wavelets. University of Nottingham.

Shen, L. & Bai, L. (2004) Gabor Wavelets and Kernel direct discriminant analysis for face recognition. *IEEE International conference on pattern recognition*.

Shen, L. & Bai, L. (2006) Review on Gabor wavelets for face recognition *Pattern Analysis Application,* 9**,** 273-292.

Shimamura, A. P., Ross, J. G. & Bennett, H. D. (2006) Memory for facial expressions : The power of a smile. *Pyschonomic Bulletin and Review,* 13**,** 217-222.

Shlens, J. (2005) A tutorial in Principal Component Analysis. Center for Neural Science, New York University andSystems Neurobiology Laboratory, Salk Insitute for Biological Studies, La Jolla,CA.

Sinha, P., Balas, B., Ostrovsky, Y. & Russell, R. (2006) Face Recognition by Humans: Nineteen Results Researchers know About. *Proceedings of the IEEE,* 94.

Skiljan, I. (2009) Irfan View.

Smith, L. I. (2002) Tutorial on Principal Component Analysis. Cornell University.

Sprengelmeyer, R., Rausch, M., Eysel, U. T. & Przuntek, H. (1998) Neural structures associated with recognition of facial expressions of basic emotions. *Proceedings of the Royal Society of London. Series B: Biological Sciences,* 22**,** 1927-1931.

Stathopoulou, I. O. & Tsihrintzis, G. A. (2006) Facial Expression Classification: Specifying Requirements for an Automated System. *Lecture Notes in Artificial Intelligence : Knowledge-Based Intelligent Information and Engineering Systems***,** 1128 – 1135.

Stathopoulou, I. O. & Tsihrintzis, G. A. (2007) NEU-FACES: A Neural network based face image analysis system. IN Al, B. B. E. (Ed. *International Conference on Adaptive and Natural Computing Algorithms.*

Strack, F., Martin, L. & Stepper, S. (1988) Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology,* 54**,** 768-777.

Susskind, J. M., Littlewort, G., Bartlett, M. S., Movellan, J. & Anderson, A. K. (2007) Human and computer recognition of facial expressions of emotion. *Neuropsychologia,* 45.

Suzuki, A., Hoshino, T., Shigemasu, K. & Kawamura, M. (2006) Disgust-specific impairment of facial expression recognition in Parkinson's disease. *Brain*

Suzuki, A., Hoshino, T., Shigemasu, K. & Kawamura, M. (2007) Decline or improvement?: Age-related differences in facial expression recognition *Biological Psychology,* 74.

Swets, D. & Weng, J. (1996) Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 18**,** 831-836.

Taylor, N. (2003) Developing with Authorware- Test bed. *Association of the Technical Staff in Psychology Conference* University of Hertfordshire, Hatfield, UK.

Teo, W. K., De Silva, L. C. & Vadakkepat , P. (2004) Facial Expression Detection and Recognition System. *Journal of The Institution of Engineers,* 44.

Terzopoulos, D. & Waters, K. (1993) Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15**,** 569-579.

Tian, Y., Kanade, T. & Cohn, J. F. (2005) Facial expression analysis.

Torro-Alves, N., Aznar-Casanova, J. A. & Fukusima, S. S. (2009) Patterns of brain asymmetry in the perception of positive and negative facial expressions. *Laterality: Asymmetries of Body, Brain and Cognition,* 14**,** 256-272.

Tsoi, D. T., Lee, K. H., Khokhar, W. A., Mir, N. U., Swalli, J. S., Gee, K. A., Pluck, G. & Woodruff, P. W. (2008) Is facial emotion recognition impairment in schizophrenia identical for different emotions? A signal detection analysis. *Schizophrenia Research,* 99**,** 263-269.

Turk, M. & Pentland, A. (1991) Eigenfaces for recognition. *Journal of Cognitive neuroscrience,* 3**,** 71-86.

Vapnik, V. (1979) Estimation of Dependences Based on Empirical Data

Vasiliki, O. & Louise, H. P. (2008) Effects of Age and Emotional Intensity on the Recognition of Facial Emotion. *Experimental Aging Research,* 34**,** 63-79.

Veksler, O. (2006) Pattern Recognition.

Vert, J. P. (2002) Support Vector Machines (SVM) in bioinformatics Bio-informatics centre, Tokyo, Japan.

Vezhnevets, V., Soldatov, S., Degtiareva, A. & Park, I. K. (2004) Automatic extraction of frontal facial features. *Asian Conference on Computer Vision (ACCV04).* Jeju, Korea.

Viola, P. & Jones, M. (2004) Robust real time face detection. *Journal of Computer Vision,* 57**,** 137-154.

Wagner, H. L., Macdonald, C. J. & Manstead, A. S. R. (1986) Communication of individual emotions by spontaneous facial expressions. *Journal of Peronality and Social Psychology,* 50**,** 737-743.

Wang, J. & Yin, L. (2007) Static topographic modeling for facial expression recognition and analysis. *Computer Vision and Image Understanding*.

Wang, J., Yin, L., Wei, X. & Sun, Y. (2006) 3D Facial expression recognition based on primitive surface feature distribution *IEEE international conference on Computer Vision and Pattern Recognition.*

Wang, K., Hoosain, R., Lee, T. M. C., Meng, Y., Fu , J. & Yang, R. (2006e) Perception of Six Basic Emotional Facial Expressions by the Chinese. *Journal of Cross-Cultural Psychology,* 37**,** 623.

Welling, M. (2005) Fisher Linear Discriminant. Department of Computer Science,University of Toronto, Toronto, M5S 3G5 Canada.

Whitehill, J. & Omlin, C. W. (2006) Haar Features for FACS AU recognition. *IEEE International Conference on Automatic Face and Gesture Recognition.*

Williams, M. A., Mcglone, F., Abbott, D. F. & Mattingley, J. B. (2008) Stimulus-driven and strategic neural responses to fearful and happy facial expressions in humans. *European Journal of Neuroscience,* 27**,** 3074-3082.

Wimmer, M., Zucker, U. & Radig, B. (2007) Human Capabilities on Vieo-based Facial Expression Recognition. Waseda University, Tokyo and University of Technology, Munich.

Wiskott, L., Fellous, J., Kruger, N. & Malsburg, C. (1999) Face Recognition by Elastic Bunch Graph Matching. *Intelligent Biometric Techniques in Fingerprint and Face Recognition,* 11**,** 355-396.

Wu, H. Y., Y., Y. & Shioyama, T. (2002) Optimal Gabor filters for high speed face identification. *International Conference on Pattern Recognition* 107-110.

Yacoob, Y. & Davis, L. S. (1994) Recognizing human facial expression. *In Proceedings of Workshop on Visual form*

Yang, E., Zald, D. H. & Blake, R. (2007) Fearful Expressions Gain Preferential Access to Awareness During Continuous Flash Suppression. *Emotion,* 7**,** 882-886.

Yin, L., Wei, X., Sun, Y., Wang, J. & Rosato, M. J. (2006) A 3D Facial Expression Database For Facial Behavior Research. *International Conference on Automatic Face and Gesture Recognition (FGR06).*

Young, A. W., Hellawell, D. & Hay, D. C. (1987) Configural information in face perception. *Perception,* 16**,** 747-759.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A. & Perrett, D. I. (1997) Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition,* 63**,** 271-313.

Zhang, L. & Cottrell, G. W. (2005) Holistic Processing Develops Because it is Good. *Proceedings of the COGSCI.*

Zheng, D., Zhao, Y. & Wang, J. (2004a) Features Extraction using A Gabor Filter Family. *International conference on Signal and Image processing.* Hawaii, USA.

Zheng, E., Ping, L. & Song, Z. (2004b) Performance Analysis and Comparison of Neural Networks and Support Vector Machines Classifier. *World Congress on Intelligent Control and Automation.* Hangzhou, P.R. China.

Zheng, Z., Pantic , M., Roisman, G. I. & Huang, T. S. (2009) A Survey of Affect recognition methods: Audi, Visual and Spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 31**,** 39-58.

Zucker, U., Radig, B. & Wimmer, M. (2007) Facial Expression Recognition - A comparison between humans and algorithms. Technical Institute of Munich.

Appendix A

A.1 Steps involved in obtaining the Principal components by PCA

1. Consider a dataset which has N number of examples, each with D dimension.

2. It has a D × N matrix. The D dimensions form the number of rows and N examples from the number of columns of the matrix X.

3. Find the mean of each column (of the corresponding examples).

4. Subtract the mean from the every column to form a matrix which has zero mean.

5. $\bar{X} = \{x_i - m\}$ where $m = \frac{1}{N} \sum_{i=1}^{N} x_i$ is the mean.

6. Calculate the covariance matrix which is given by $C_X = \frac{1}{N} \bar{X} \bar{X}^T$ where T denotes transpose.

7. Calculate the eigenvectors and eigenvalues of the covariance matrix. The diagonal elements of this symmetric covariance matrix are the variances of the ith variable which varies from 1 to N.

8. Then, once eigenvectors are found from the covariance matrix, the next step is to order them by eigenvalue, highest to lowest. By retaining only the first p eigenvectors which attain 95% variance of the input, dimensionality reduction is achieved. Note that there can be no more than N Eigenvectors. The Important point here is that this method enables finding the eigenvectors even for large matrices.

The steps involved in finding the PCA projection and then reconstruction is as follows:

- For a dataset of face images X= $\{x_1, x_2, \ldots, x_N\}$ with N samples and D dimensions, the mean face m is found.

- The average face is subtracted from each image: $\bar{x_i} = x_i - m$ and $\bar{X} = \{\bar{x}_1, \ldots., \bar{x}_N\}$.

- Calculate the covariance matrix which is given by $C_X = \frac{1}{N} \bar{X} \bar{X}^T$ where T denotes transpose. As $\bar{X}$ has a large dimension (N*D × N*D), finding the eigenvectors is difficult. However, it is easy to find the eigenvectors of the $\bar{X}^T\bar{X}$ of dimension N × N as N<<D.

- If we take V_i as the eigenvector of $\bar{X}^T\bar{X}$ and λ_i as the eigenvalue, then $(\bar{X}^T\bar{X})V_i = \lambda_i V_i$.

- Therefore, by multiplying $\bar{X}$ on the left hand side of equation above, $\bar{X}(\bar{X}^T\bar{X}V_i) = \bar{X}(\lambda_i V_i)$ and hence $(\bar{X}\bar{X}^T) = \lambda_i(\bar{X}V_i)$ which implies that $\bar{X}V_i$ is the eigenvector solution of the matrix $\bar{X}^T\bar{X}$ with the same λ_i as the eigenvalue.

- Thus, $\bar{X}V_1 = U_1, \bar{X}V_2 = U_2, \ldots\ldots., \bar{X}V_n = U_n$ are the eigenfaces. Here n=N-1 considers only the first non zero eigenvalues.

- The PCA projection would be then to produce $L_k = {U_k}^T\bar{X}^T$ (where U^T is of size k × D) of size k × N. Hence the reconstruction would be obtained by $R_k = U_k L_k$ or $R_i = \sum_{i=1}^{k} L_i U_i$.

Appendix B

Table B.1: Significant components for all expressions

Expression	First highest component	Second highest component
Angry	26	3
Happy	7	6
Fear	7	14
Sad	26	14
Surprise	3	2
Disgust	26	13

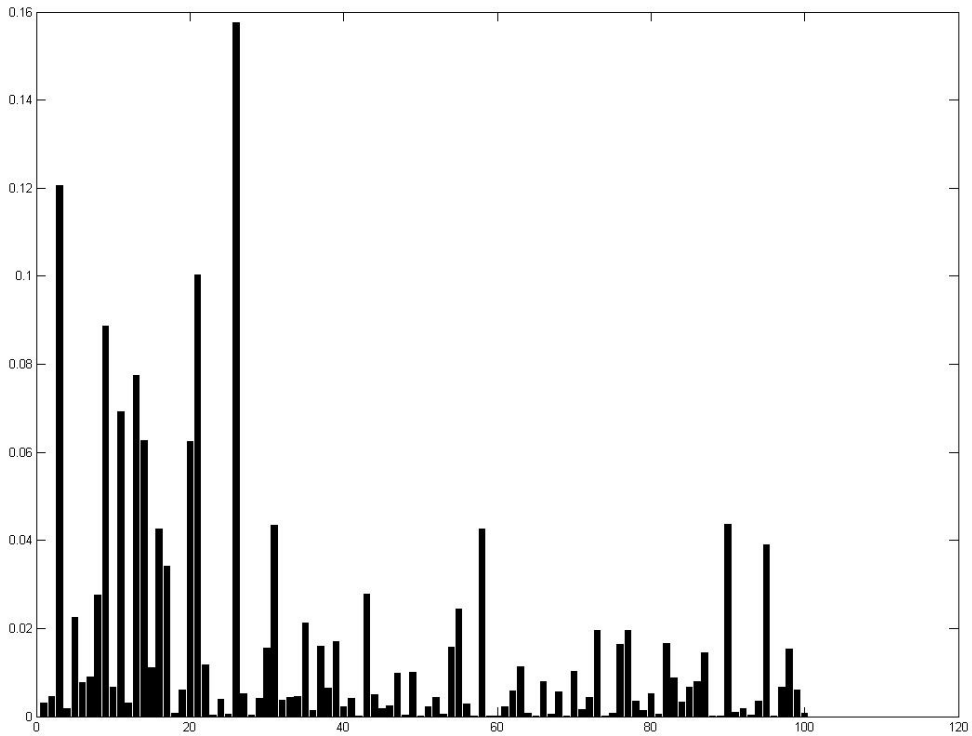


Figure B.1: Angry encoding power - 26th component has the highest anger encoding power and 3rd component has the second highest encoding power

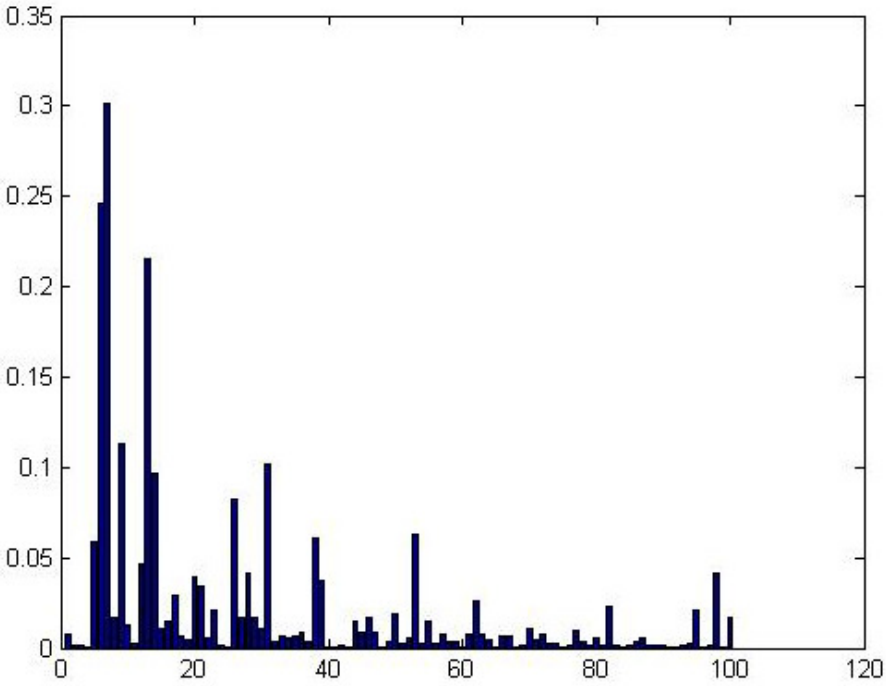


Figure B.2: Happy encoding power - 7th component has the highest happy encoding power and 6th component has the second highest encoding power

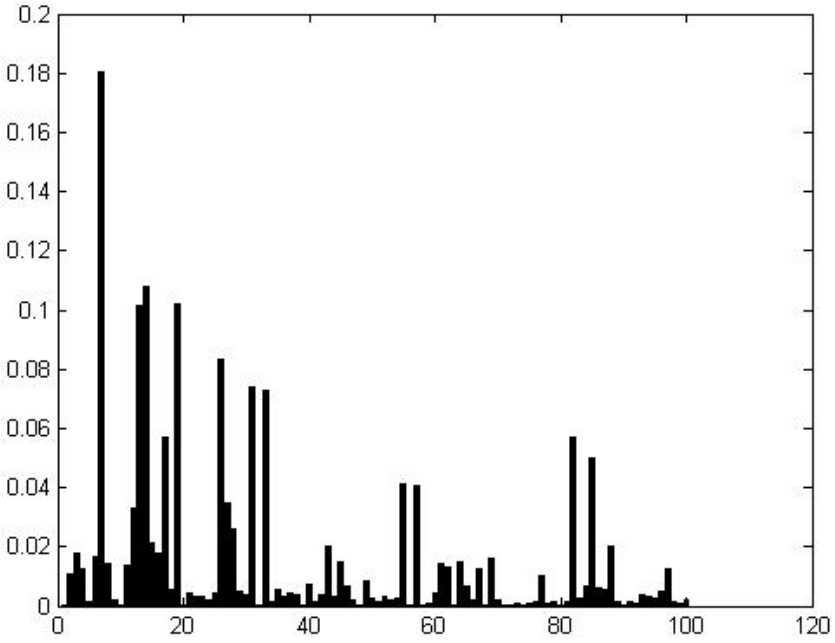


Figure B.3: Fear encoding power - 7th component has the highest fear encoding power and 14th component has the second highest encoding power

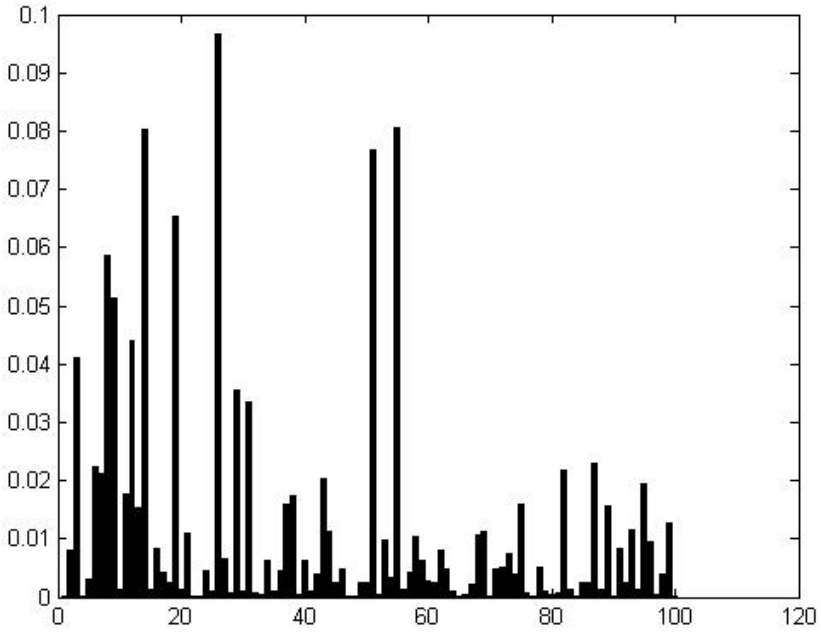


Figure B.4: Sad encoding power - 26th component has the highest sad encoding power and 14th component has the second highest encoding power

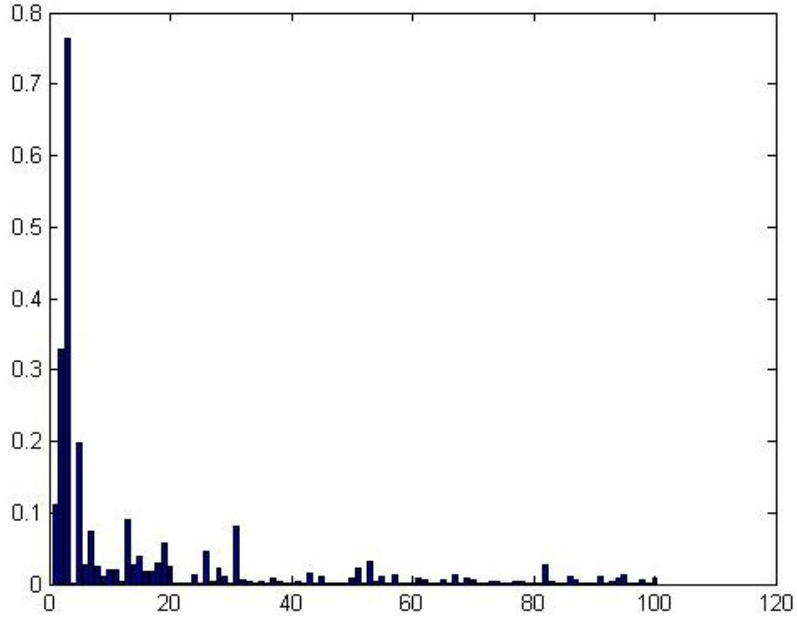


Figure B.5: Surprise encoding power – 3rd component has the highest surprise encoding power and 2rd component has the second highest encoding power

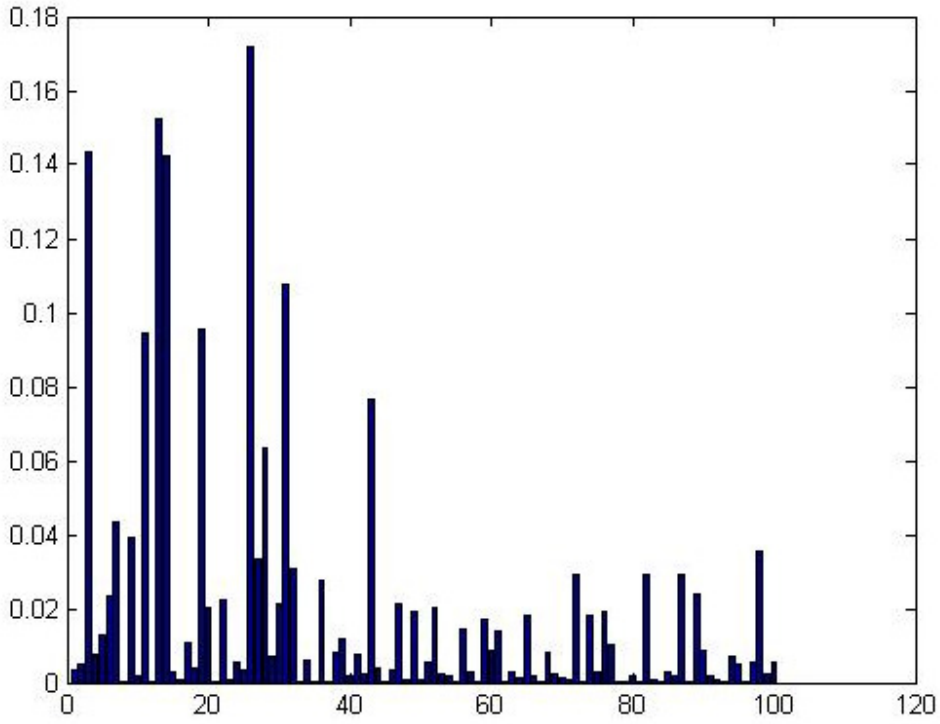


Figure B.6: Disgust encoding power - 26th component has the highest disgust encoding power and 13th component has the second highest encoding power

Appendix C

Table C.1: Classification Accuracy of the PCA + LDA processed data by measuring Euclidean distance

% ACCURACY	TEST SET 4	TEST SET 3	TEST SET 2	TEST SET 1	AVERAGE
ANGRY	61.364	65.909	56.818	52.273	59.091
HAPPY	56.818	54.545	63.636	84.091	64.7725
FEAR	54.545	56.818	63.636	65.909	60.227
SAD	59.091	52.273	59.091	68.182	59.65925
SURPRISE	63.636	75	79.545	59.091	69.318
DISGUST	65.909	59.091	68.182	61.364	63.6365

Table C.2: Classification Accuracy of LDA + PCA processed data with the SVM classifier

% ACCURACY	TEST SET 4	TEST SET 3	TEST SET 2	TEST SET 1	AVERAGE
ANGRY	72.73% (32/44)	84.0909% (37/44)	79.5455% (35/44)	88.6364% (39/44)	81.25% (143/176)
HAPPY	100% (44/44)	100% (44/44)	90.9091% (40/44)	100% (44/44)	97.7272% (172/176)
FEAR	84.0909% (37/44)	81.8182% (36/44)	84.0909% (37/44)	86.3636% (38/44)	84.0909% (148/176)
SAD	84.0909% (37/44)	79.5455% (35/44)	72.7273% (32/44)	84.0909% (37/44)	80.1136% (141/176)
SURPRISE	90.9091% (40/44)	95.4545% (42/44)	95.4545% (42/44)	100% (44/44)	95.4545% (168/176)
DISGUST	90.9091% (40/44)	88.6364% (39/44)	84.0909% (37/44)	90.9091% (40/44)	88.6363% (156/176)

Table C.3: Cross validation results for angry expression by the SVM classifier

Angry	**TEST SET 4**	**TEST SET 3**	**TEST SET2**	**TEST SET 1**	**Average**
RAW	79.54% (35/44)	93.18% (41/44)	79.54% (35/44)	84.09% (37/44)	84.09% (148/176)
RAWPCA97	68.18% (30/44)	77.27% (34/44)	70.45% (31/44)	65.91% (29/44)	70.45%
RAWCCA5	68.18% (30/44)	59.09% (26/44)	63.64% (28/44)	63.64% (28/44)	63.64%
GAB	68.18% (30/44)	79.55% (35/44)	72.73% (32/44)	81.82% (36/44)	75.57% (133/176)
GABPCA22	61.36% (27/44)	79.55% (35/44)	75% (33/44)	72.73% (32/44)	72.16%
GABCCA6	63.64% (28/44)	70.45% (31/44)	68.18% (30/44)	63.64% (28/44)	66.48%

Table C.4: Cross validation results for happy expression by the SVM classifier

HAPPY	**TEST SET 4**	**TEST SET 3**	**TEST SET2**	**TEST SET 1**	**Average**
RAW	100%	100%	97.73% (43/44)	100%	99.43% (175/176)
RAWPCA100	88.6364% (39/44)	86.36% (38/44)	93.18% (41/44)	88.64% (39/44)	89%
RAWCCA6	93.18% (41/44)	79.55% (35/44)	84.09% (37/44)	93.18% (41/44)	87.50%
GAB	90.91% (40/44)	90.91% (40/44)	86.36% (38/44)	90.91% (40/44)	89.77% (158/176)
GABPCA23	95.45% (42/44)	81.82% (36/44)	81.82% (36/44)	88.64% (39/44)	86.93%
GABCCA5	68.18% (30/44)	61.36% (27/44)	59.09% (26/44)	56.82% (25/44)	61.36%

Table C.5: Cross validation results for fear expression by the SVM classifier

FEAR	**TEST SET 4**	**TEST SET 3**	**TEST SET2**	**TEST SET 1**	**Average**
RAW	86.36%	86.36%	79.54%	81.82%	
	(38/44)	(38/44)	(35/44)	(36/44)	83.52%
RAWPCA99	77.27%	93.18%	79.55%	79.55%	
	(34/44)	(41/44)	(35/44)	(35/44)	82.39%
RAWCCA6	75%	75%	72.73%	70.45%	
	(33/44)	(33/44)	(32/44)	(31/44)	73%
GAB	72.73%	63.64%	84.09%	79.55%	
	(32/44)	(28/44)	(37/44)	(35/44)	75.00%
GABPCA23	77.27%	75%	84.09%	81.82%	
	(34/44)	(33/44)	(37/44)	(36/44)	79.55%
GABCCA5	50%	54.55%	63.64%	52.27%	
	(22/44)	(24/44)	(28/44)	(23/44)	55%

Table C.6: Cross validation results for sad expression by the SVM classifier

SAD	**TEST SET 4**	**TEST SET 3**	**TEST SET2**	**TEST SET 1**	**Average**
RAW	84.09%	79.55%	75%	70.45%	
	(37/44)	(35/44)	(33/44)	(31/44)	77.27%
RAWPCA96	68.18%	79.55%	79.55%	70.45%	
	(30/44)	(35/44)	(35/44)	(31/44)	74.43%
RAWCCA7	63.64%	56.82%	65.91%	63.64%	
	(28/44)	(25/44)	(29/44)	(28/44)	62.50%
GAB	68.18%	68.18%	75%	70.45%	
	(30/44)	(30/44)	(33/44)	(31/44)	70.45%
GABPCA22	68.18%	77.27%	70.45%	68.18%	
	(30/44)	(34/44)	(31/44)	(30/44)	71.02%
GABCCA5	54.55%	56.82%	61.36%	61.36%	
	(24/44)	(25/44)	(27/44)	(27/44)	58.52%

Table C.7: Cross validation results for surprise expression by the SVM classifier

SURPRISE	**TEST SET 4**	**TEST SET 3**	**TEST SET2**	**TEST SET 1**	**Average**
RAW	93.18%	95.45%	95.45%	95.45%	
	(41/44)	(42/44)	(42/44)	(42/44)	94.89%
RAWPCA103	93.18%	84.09%	93.18%	86.36%	
	(41/44)	(37/44)	(41/44)	(38/44)	89.20%
RAWCCA6	95.45%	95.45%	97.73%	86.36%	
	(42/44)	(42/44)	(43/44)	(38/44)	93.75%
GAB	95.45%	97.73%	95.45%	93.18%	
	(42/44)	(43/44)	(42/44)	(41/44)	95.45%
GABPCA23	88.64%	93.18%	93.18%	86.36%	
	(39/44)	(41/44)	(41/44)	(38/44)	90.34%
GABCCA5	81.82%	84.09%	86.36%	84.09%	
	(36/44)	(37/44)	(38/44)	(37/44)	84.09%

Table C.8: Cross validation results for disgust expression by the SVM classifier

DISGUST	**TEST SET 4**	**TEST SET 3**	**TEST SET2**	**TEST SET 1**	**Average**
RAW	90.91%	90.91%	90.91%	88.64%	
	(40/44)	(40/44)	(40/44)	(39/44)	90.34%
RAWPCA101	75%	81.82%	81.82%	81.82%	
	(33/44)	(36/44)	(36/44)	(36/44)	80%
RAWCCA5	70.46%	68.18%	75%	65.91%	
	(31/44)	(30/44)	(33/44)	(29/44)	69.89%
GAB	72.73%	70.45%	68.18%	81.82%	
	(32/44)	(31/44)	(30/44)	(36/44)	73.30%
GABPCA23	72.73%	72.73%	81.82%	79.46%	
	(32/44)	(32/44)	(36/44)	(35/44)	76.68%
GABCCA5	56.82%	59.09%	61.36%	65.91%	
	(25/44)	(26/44)	(27/44)	(29/44)	60.80%

Appendix D

Table D.3: Results of Bi-Variate correlation between average RT of human subjects and the distance measure of the hyper-plane for the SVM classifier used with all computational models for incorrect responses. The numbers in red font indicate significant levels and their corresponding correlation values.

Missclass-fications	Raw		RAWPCA		RAWCCA		GAB		GABPCA		GABCCA	
	S	C	S	C	S	C	S	C	S	C	S	C
Angry	**0.943 N=27**	**-0.015**	**0.642 N=49**	**-0.068**	**0.183 N=60**	**-0.174**	**0.391 N=40**	**0.139**	**0.414 N=48**	**0.121**	**0.422 N=58**	**0.108**
Happy	**-NA-**	**-NA-**	**0.696 N=19**	**-0.096**	**0.927 N=22**	**-0.021**	**0.685 N=17**	**-0.106**	**0.635 N=22**	**-0.107**	**0.378 N=65**	**-0.111**
Fear	**0.566 N=28**	**0.113**	**0.914 N=30**	**0.021**	**0.942 N=46**	**-0.011**	**0.143 N=44**	**0.224**	**0.031 N=36**	**0.360**	**0.959 N=76**	**0.006**
Surprise	**0.412 N=9**	**0.313**	**0.178 N=19**	**0.323**	**0.777 N=11**	**-0.097**	**0.630 N=8**	**-0.203**	**0.928 N=17**	**-0.024**	**0.540 N=28**	**0.121**
Sad	**0.174 N=39**	**-0.222**	**0.100 N=42**	**0.257**	**0.278 N=64**	**0.138**	**0.567 N=50**	**+0.083**	**0.314 N=49**	**0.147**	**0.595 N=70**	**0.065**
Disgust	**0.953 N=16**	**0.016**	**0.366 N=34**	**0.160**	**0.478 N=52**	**-0.101**	**0.436 N=43**	**-0.122**	**0.911 N=40**	**-0.018**	**0.896 N=66**	**-0.016**

Appendix E

Conference Proceedings and Poster Abstracts

Recognizing emotions by analyzing facial expressions

Aruna Shenoy a.1.shenoy@herts.ac.uk
Tim MGale t.gale@herts.ac.uk
Ray Frank r.j.frank@herts.ac.uk
Neil Davey n.davey@herts.ac.uk

School of Computer Science
University of Hertfordshire
UK

Abstract

Recognizing expressions is a key part of human social interaction, and processing of facial expression information is largely automatic. However, it is a non-trivial task for a computational system. Our purpose of this work is to develop Computational models capable of differentiating between ranges of human Facial expressions. The Gabor feature is effective for facial image representation. The Gabor feature dimensionality is so high that a dimensionality reduction technique such as PCA must be applied. Classification of various classes of expressions can be achieved by training and then testing with a Support Vector Machine (SVM).

1 Introduction

According to Ekman and Friesen (Ekman et al., 1971) there are six easily discernible facial expressions: anger, happiness, fear, surprise, disgust and sadness. Moreover these are readily and consistently recognized across different cultures (Batty et al., 2003). In the work reported here we show how a computational model can identify facial expressions from simple facial images. In particular we show how smiling faces and neutral faces can be differentiated.

We first pre-process the images using Gabor Filters (Jain et al., 1991; Movellan 2002). The features of the face (or any object for that matter) can be aligned at any angle. Using a suitable Gabor filter at the required orientation, certain features can be given high importance and other features less importance. Usually, a bank of such filters is used with different parameters and later the resultant image is a *L2 max* (at every pixel the maximum of feature vector obtained from the filter bank) superposition or average of the outputs from the filter bank. Gabor filters are interesting because simple cells in the visual cortex are known to be selective for the following four parameters: the x, y location in visual space, the preferred orientation, and the preferred spatial frequency (Daugman, 1985).

Recent work on these suggests that the various 2D receptive field profiles encountered in populations of simple cells are well described by the family of 2D Gabor filters (Daugman, 1985).

Data presentation plays an important role in any type of recognition. High dimensional data, such as the output of the Gabor filters of the face images, must be reduced to a manageable low dimensional data set by using a technique such as Principal Component Analysis (PCA). The Intrinsic Dimension (ID) (Grassberger et al., 1983), which is the true dimension of the data, is often much less than the original dimension of the data.

2 Background

We begin with a simple experiment to classify two expressions: neutral and smiling. The image is pre-processed by using a bank of Gabor filters. A Support Vector Machine (SVM) (Chih Chung et al., 2001) based classification technique is used.

2.1 Gabor Filters

A Gabor filter can be applied to images to extract features aligned at particular angles (orientations). Gabor filters possess the optimal localization properties in both spatial and frequency domains, and they have been successfully used in many applications (Zheng et al., 2004a). A Gabor filter is a function obtained by modulating a sinusoidal with a gaussian function. The useful parameters of a Gabor filter are orientation and frequency. It is used to enhance certain features that share an orientation and/or frequency and thereby enables useful pre-processing required for facial expressions, recognition and analysis to be carried out. The Gabor filter is thought to mimic the simple cells in the visual cortex. The various 2D receptive-field profiles encountered in populations of simple cells in the

visual cortex are well described by an optimal family of 2D filters (Daugman, 1985). In our case a Gabor filter bank is implemented on face images with 8 different orientations and 5 different frequencies.

A Gabor filter can be one or two dimensional (2D). A 2D Gabor filter is expressed as a Gaussian modulated sinusoid in the spatial domain and as shifted Gaussian in the frequency domain. Recent studies on modeling of visual cortical cells (Kulikowski, 1982) suggest a tuned band pass filter bank structure. These filters are found to have Gaussian transfer functions in the frequency domain. Thus, taking the Inverse Fourier Transform of this transfer function gives characteristics closely resembling Gabor filters.

A well designed Gabor filter bank can capture the relevant frequency spectrum in all directions. Phase can be taken as a feature because it contains information about the edge locations and other such details in the image; amplitude at every pixel can be taken as a feature as it contains some oriented frequency spectrum at every point of the image. We can extract many meaningful features using the Gabor filter family. Experimental results in texture analysis and character analysis demonstrate these features in the capture of local information with the different frequencies and orientations in the image (Zheng et al., 2004a).

The Gabor filter is a Gaussian (with variances S_x and S_y along x and y-axes respectively) modulated by a complex sinusoid (with centre frequencies U and V along x and y-axes respectively) described by the following equation:-

$$g(x,y)=\frac{1}{2\pi S_x S_y}\exp\left[-\frac{1}{2}\left\{\left(\frac{x}{S_x}\right)^2+\left(\frac{y}{S_y}\right)^2\right\}+2\pi j(Ux+Vy)\right] \quad (1)$$

The variance terms S_x and S_y dictate the spread of the band pass filter centered at the frequencies U and V in the frequency domain. This filter is complex and the plot of the Real and Imaginary parts of $g(x,y)$ is shown in Figure 1 :-

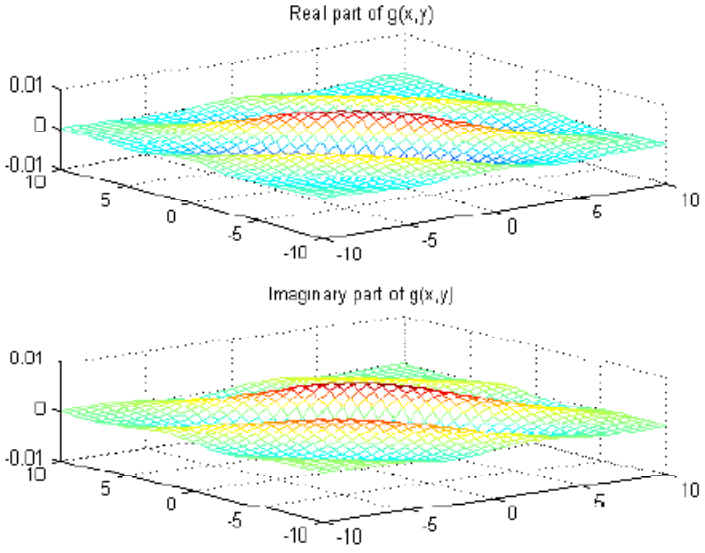


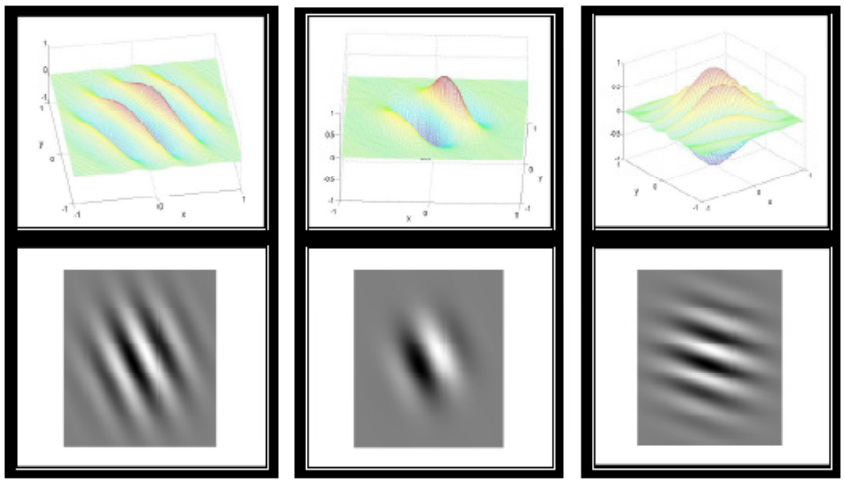
Figure 1: Plot of Real and Imaginary part of Gabor filter

It is found that as the 2D Gabor filter is applied to the images, the edges are smoothened out in all directions due to the presence of the Gaussian term. Each filter can be designed to pick out particular image features in orientation and the required frequency.

A Gabor filter can be best described by the following parameters:

1. The S_x and S_y of the gaussian explain the shape of the base (circle or ellipse).
2. The frequency (f) of the sinusoid.
3. The orientation (Θ) of the applied sinusoid.

Figures 2 and Figure 3 show examples of various Gabor filters



(a) (b) (c)

Figure 2: Figures (a), (b), (c) are examples of Gabor filter with different frequencies and orientations. Top row shows their 3D plots and the bottom row, the intensity plots of their amplitude along the image plane.

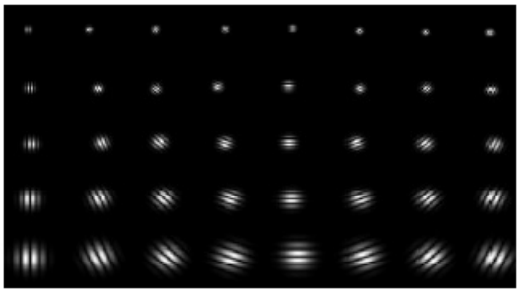


Figure 3: Gabor filters: Real part of the Gabor kernels at five scales and eight orientations

Figure 5 and Figure 6 show the effect of applying a particular Gabor filter on Figure 4 which is an image with lines at various angles. The highlighted lines in Figure 5 and Figure 6 shows the way the Gabor filter exaggerates lines at particular orientations.

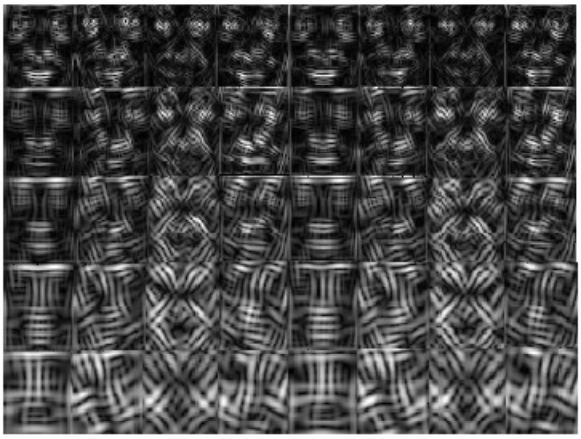
Figure 4: Image with lines at various angles

Figure 5: Frequency, $f = 12.5$ and orientation, $\theta = 135$ degrees

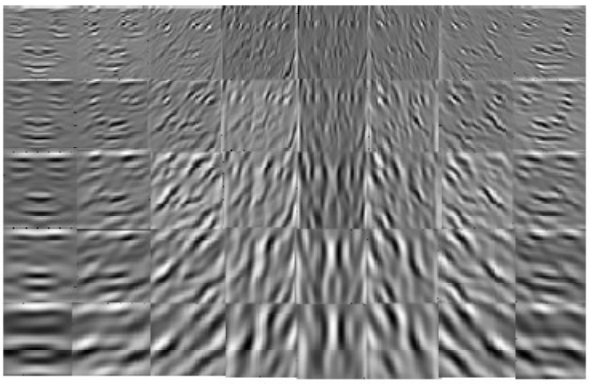
Figure 6: Frequency, f=25 and orientation, θ=0 degrees

Figure 8 shows the effect of applying variety of Gabor filters to the image shown in Figure 7. Note how the features at particular orientations are exaggerated.

Figure 7: Sample Image of size 64 × 64



(a)



(b)

Figure 8: Convolution outputs of a sample image shown in Figure 7 and the Gabor kernels (Fig. 3). (a) Magnitude part of the convolution outputs. (b) Real part of the convolution outputs.

Analytical methods make use of Gabor jets at specific points on the face which are vital feature points (fiducial points). There are different methods for identifying or locating these feature points. For elastic graph based analytic methods, a graph is first placed at an initial location and deformed using jets to optimize its similarity with a model graph. Non-graph based methods locate feature points manually or by color or edge etc. Once the location process is completed, recognition can then be performed using Gabor jets extracted from those feature points (Shen 2004).

Holistic methods on the other hand normally extract features from the whole face image. An augmented Gabor feature vector is thus created which produces a very large data for the image. Every pixel is then represented by a vector of size 40 and demands dimensionality reduction before further processing. So a 64 × 64 image is transformed to size 64 × 64 × 5 × 8. So, the feature vector consists of all useful information extracted from different frequencies, orientations and from all locations, and hence is very useful for expression recognition. Once the feature vector is obtained, it can be handled in various ways. We have performed the following operations and any one of them can be used for the feature extraction:

a) The final image can be of the average of the magnitudes of the Gabor filter coefficients at each location in the filter bank output.

b) The pixel value in the final image would be the *L2 max* norm value of the feature vector obtained from the Gabor filter bank

The *L2 max* norm Superposition principle is used on the outputs of the filter bank and the figure 10 shows the output for the original image of figure 9. Similarly the outputs of the 40 filter banks can also be averaged or summed to give an output as in figure 11 shown below.

Figure 9: Original Image used for the Filter bank

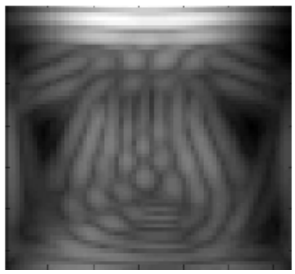


Figure 10: Superposition output (*L2 max* norm)

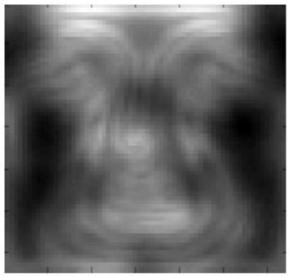


Figure 11: Average Output

2.2 Classification using Support Vector Machines

A number of classifiers can be used in the final stage for classification. We have concentrated on the Support Vector Machine. Support Vector Machines (SVM) are a set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classifiers. The SVM finds the optimal separating hyper-plane that has the maximal margin of separation between the classes, while having minimum classification errors. This means the SVM classifier tries to find the plane which separates the two different classes such that it is equidistant from the members of either class which are nearest to the plane. SVM's are used extensively for a lot of classification tasks such as: handwritten digit recognition (Cortes et al., 1995) or Object Recognition (Blanz et al., 1996). SVM's can be slow in test phase, although they have a good generalization performance. In total the SVM theory says that the best generalization performance can be achieved with the right balance between the accuracy attained on the training data and the ability to learn any training set without errors, for the given amount of training data. The SVM shows better classification accuracy than Neural Networks (NNs) if the data set is small. Also, the time taken for training and predicting the test data is much smaller for a SVM system than for a NN (Zheng et al., 2004b).

In short, they can be explained as a classifier which finds the optimum plane that performs the classification task by constructing a hyperplane in a multidimensional space that separate cases of different class labels.

In this example, the objects belong either to class GREEN or RED. The separating line defines a boundary on the right side of which all objects are GREEN and to the left of which all objects are RED. Any new object falling to the right is labeled, i.e., classified, as GREEN or classified as RED if it falls to the left of the separating line.

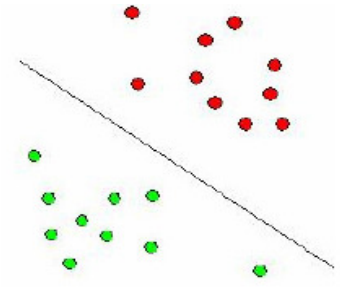


Figure 12: A Linear Classifier

Most classifications are not this simple, and a more complicated example is shown in Figure 13. In this example, it needs a curve rather than a straight line to separate the two classes.

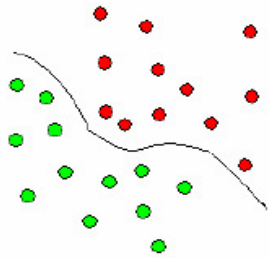


Figure 13: A non Linear Classifier.

A SVM rearranges the original objects (data points) according to a mathematical function (kernels) and transforms it into a feature space which allows the classification to be accomplished more easily, and is illustrated in Figure 14.

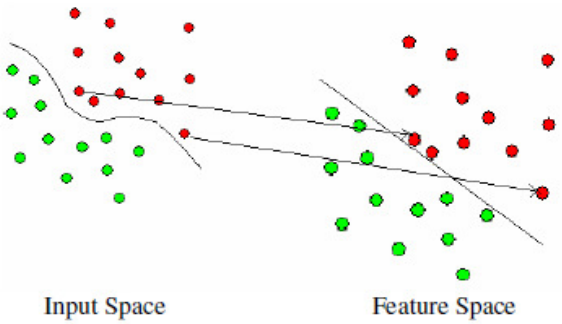


Figure 14: Transformation from input space to Feature space by the Support Vector Machine

We have used the LIBSVM tool (Chih-Chih, 2001) for SVM classification.

2.3 Principal Component Analysis

Principal Component Analysis (PCA) transforms higher dimensional datasets into lower dimensional uncorrelated outputs by capturing linear correlations among the data, and preserving as much information as possible in the data. PCA transforms data from the original coordinate system to the principal axes coordinate system such that the principal axis passes through the maximum possible variance in the data. The second principal axis passes through the next largest possible variance and this is orthogonal to the first axis. This is repeated for the next largest possible variances and so on. All these axes are orthogonal to each other. On performing the PCA on the high dimensional data, eigenvalues or principal components are thus obtained (Smith, 2002). The required dimensionality reduction is obtained by retaining only the first few principal components.

The PCA is used to project a D-dimensional dataset X onto an uncorrelated d- dimensional dataset Y, where $d \leq D$, by capturing the linear correlation between the data and preserving as much information as possible. In other words, the aim is to find a set of d orthogonal vectors in the data space that account for as much as possible of the variance of the data. Projecting the data from their original D-dimensional space onto the d-dimensional subspace spanned by these vectors then performs a dimensionality reduction that often retains most of the intrinsic information in the data. The variances measured on these orthogonal axes are the eigenvalues of the Principal components (Smith, 2002).

The Principal Components have the following properties: They can be ranked by decreasing order of "importance". The first few most "important" Principal Components account for most of the information in the data. In other words, one may then discard the original data set, and replace it with a new data set with the same observations, but fewer variables, without throwing away too much information.

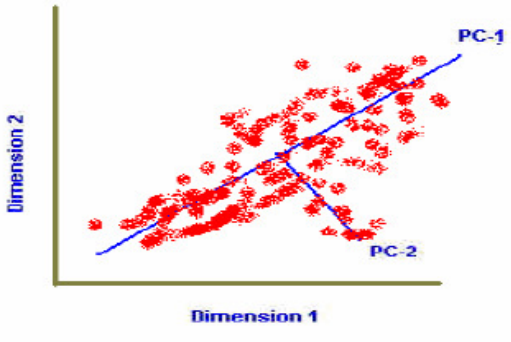


Figure 15: Figure shows the first two consecutive principal components.

The principal components are:
1. Orthogonal (at right angles) to each other.
2. They are uncorrelated.

3 Experiments and Results

We experimented on 120 faces (60 male and 60 female) each with two classes, namely, Neutral and smiling (60 faces for each expression). The images are from The FERET dataset (Philips et al., 2003).

Figure 16: Example FERET images used in our experiments and then cropped to the size of 128 × 128 to extract the facial region.

The training set was 80 faces (with 40 female, 40 male and equal numbers of them with neutral and smiling). Two test sets were created. In both test sets the number of each type of face is balanced. For example, there were 5 smiling male faces, 5 smiling female faces and 5 neutral male faces. For the purpose of comparison, the SVM classification was performed on the raw face images (150 × 130). With all faces aligned based on their eye location, a 128 × 128 image was cropped from the original (150 × 130). The resolution of these faces is reduced to 64 × 64. The classification was then performed on these images after reducing the dimensionality of the images by PCA. Later classification was performed on the Gabor pre processed image. The results are shown in Table 1. PCA was used to reduce dimensionality of the image of size 64 × 64 dimensions (4096 pixels) to 350 i.e. taking the first 350 PCA components. The first set has images which are easily discernible smiling faces. The second test set has smiling and neutral faces, but the smiling faces are not easily discernible.

Figures 17 and 18 show the first 5 eigenfaces from the neutral faces (top row), smiling faces (bottom row) and the complete training set.

Figure 17: The top row shows the first 5 Eigen faces of all the neutral faces of the data set. The bottom row shows the first 5 Eigen faces of the smiling faces of the data set.

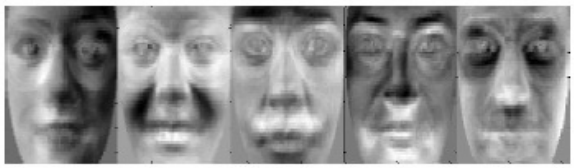


Figure 18: The first 5 Eigen faces of the whole set of faces (male and female with equal number of smile and neutral faces).

The SVM was trained in the following way:

1. Transforming the data to a format required for using the SVM software package - LIBSVM -2.83 (Chih-Chung, 2001).

2. Perform simple scaling on the data so that all the features or attributes are in the range [-1, +1].

3. Choose a kernel. We have used RBF $k(x, y) = e^{-\gamma|x-y|^2}$ kernel.

4. Perform five fold cross validation with the specified kernel to find the best values of the parameter C and γ.

5. Use the best parameter value of C and γ to train the whole training set.

6. Finally Test.

The results of the classification are as in Table 1:

% accuracy	Test set1	Test set2
SVM on Raw faces	100	80
SVM after PCA	80	80
SVM after Gabor	95	80

Table 1: SVM Classification accuracy with faces without any pre-processing and with PCA dimensionality reduction.

It is notable and surprising that the classification using the raw images produces good generalisation on the two test sets in all cases outperforming data sets using pre-processed PCA. Some examples of misclassifications are shown in Figure 20. Whilst, some of the misclassifications are explainable some are more puzzling. The relatively poor performance of the PCA suggests that a dimensionality reduction more tuned to identifying relevant features is needed. This motivates our investigation of Gabor filter pre-processing. The images are reduced to size 64 $\times$ 64 and then Gabor processing is performed. The SVM classification results are extremely good with the fact that the images being reasonably reduced in size has not reduced the accuracy in classification. The Gabor filters have managed to pick up the relevant features from the images of half the resolution and is indicative of the power of the Gabor filters.

Figure 20: Examples of the misclassified set of faces Top row shows neutral faces wrongly classified as smiling. Bottom row shows smiling faces wrongly classified as neutral.

4 Conclusions

Identifying facial expressions is a challenging and interesting task. Our experiment shows that identification from raw images can be performed very well. However, with larger data sets, it is computationally intractable. PCA does not appear to be sufficiently tunable to identify features that are relevant for facial expression characterization. However, on performing Gabor preprocessing on the images which are reduced in size, the features are well extracted and support accurate classification.

References

Batty, B., & Taylor, M.J (2003) Early processing of the six basic facial emotional expressions, Cognitive Brain Research, 17.

Blanz, V., Sch¨olkopf, B., B¨ulthoff, H., Burges, C., Vapnik, V. & Vetter, T., "Comparison of view-based object recognition algorithms using realistic 3D models", Proc. Int. Conf. on Artificial Neural Networks 1996, 251-256.

Chih-Chung Chang & Chih-Jen Lin, LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Cortes, C. & Vapnik, V.(1995) Support Vector Networks, Machine Learning, 20, 273-297.

Daugman, J.G, "Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two dimensional visual cortical filters", Journal of Optical.Society of .America .A, Vol 2, No.7.July 1985.

Ekman, P., & Friesen, W. V. (1971) Constants across cultures in the face of the emotion, Journal of Personality and Social Psychology, 17.

Grassberger, P. & Proccacia, I. (1983). Measuring the strangeness of strange attractors. *Physica D, 9.*

Jain, A.K & Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. Pattern Recognition 24(12).

Kulikowski (1982). "Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex." Biological Cybernetics 43(3): 187-198.

Movellan, J.R. (2002). Tutorial on Gabor Filters. http://mplab.ucsd.edu/tutorials/pdfs/gabor.pdf

Philips, P.J., Grother,P., Micheals,R.J., Blackburn, D., Tabassi, E.&Bone, M.(2003). Face recognition vendor test 2002: overview and summary. Available at www.frvt.org, 2003.

Shen et.al., (2006) Review on Gabor wavelets for face recognition Pattern Anal. Applic. (2006) 9:273–292

Smith, L.I., (2002) Tutorial on Principal Component Analysis.

Zheng, D., Zhao,Y., Wang, J.,(2004a) Features Extraction Using A Gabor Filter Family, Proceedings of the sixth LASTED International conference, Signal and Image processing, Hawaii, USA.

Zheng.E., Ping, L., Song, Z.(2004b) Performance Analysis and Comparison of Neural Networks and Support Vector Machines Classifier, Proceedings of the 5th World Congress on Intelligent Control and Automation, June 15-19. 2004, Hangzhou, P.R. China.

On the Recognition of Emotion from Facial Expressions

Aruna Shenoy a.1.shenoy@herts.ac.uk
Tim M Gale t.gale@herts.ac.uk
Ray Frank r.j.frank@herts.ac.uk
Bruce Christianson b.christianson@herts.ac.uk
Neil Davey n.davey@herts.ac.uk

School of Computer Science
University of Hertfordshire
UK AL 10 9AB
Tel: 0044 1707 284321

ABSTRACT

Recognizing expressions is a key part of human social interaction, and processing of facial expression information is largely automatic. However, it is a non-trivial task for a computational system. The purpose of this work is to develop computational models capable of differentiating between a range of human facial expressions. Raw face images are examples of high dimensional data, so here we use two dimensionality reduction techniques: Principal Component Analysis and Curvilinear Component Analysis. We also preprocess the images with a bank of Gabor filters, so that important features in the face images are identified. Subsequently the faces are classified using a Support Vector Machine. We show that it is possible to differentiate faces with a neutral expression from those with a smiling expression with high accuracy. Moreover we can achieve this with data that has been massively reduced in size: in the best case the original images are reduced to just 11 dimensions.

General Terms
Algorithms, Measurement, Performance, Experimentation.

Keywords
Facial Expressions, Image Analysis, Classification, Dimensionality Reduction.

1. INTRODUCTION

According to Ekman and Friesen [7] there are six easily discernible facial expressions: anger, happiness, fear, surprise, disgust and sadness. Moreover these are readily and consistently recognized across different cultures [1]. In the work reported here we show how a computational model can identify facial expressions from simple facial images. In particular we show how smiling faces and neutral faces can be differentiated.

We first pre-process the images using Gabor Filters [9, 11]. The features of the face (or any object for that matter) can be aligned at any angle. Using a suitable Gabor filter at the required orientation, certain features can be given high importance and other features less importance. Usually, a bank of such filters is used with different parameters and later the resultant image is a *L2 max* (at every pixel the maximum of feature vector obtained from the filter bank) superposition or average of the outputs from the filter bank. Gabor filters are interesting because recent work on these suggests that the various 2D receptive field profiles encountered in populations of simple cells are well described by the family of 2D Gabor filters [5].

Data presentation plays an important role in any type of recognition. High dimensional data, such as the output of the Gabor filters of the face images, is normally reduced to a manageable low dimensional data set by using a technique such as Principal Component Analysis (PCA). PCA is a linear projection technique and it may be more appropriate to use a non linear Curvilinear Component Analysis [6]. The Intrinsic Dimension (ID) [8], which is the true dimension of the data, is often much less than the original dimension of the data. To use this efficiently, the actual dimension of the data must be estimated. We use Correlation Dimension to estimate Intrinsic Dimension.

2. BACKGROUND

We begin with a simple experiment to classify two expressions: neutral and smiling. The image is pre-processed by using a bank of Gabor filters. A Support Vector Machine (SVM) [3] based classification technique is used.

2.1 Gabor Filters

A Gabor filter can be applied to images to extract features aligned at particular angles (orientations). Gabor filters possess the optimal localization properties in both spatial and frequency domains, and they have been successfully used in many applications [15]. A Gabor filter is a function obtained by

modulating a sinusoidal with a Gaussian function. The useful parameters of a Gabor filter are orientation and frequency. It is used to enhance certain features that share an orientation and/or frequency and thereby enables useful pre-processing required for facial expressions, recognition and analysis to be carried out. The Gabor filter is thought to mimic the simple cells in the visual cortex. The various 2D receptive field profiles encountered in populations of simple cells in the visual cortex are well described by an optimal family of 2D filters [5]. In our case a Gabor filter bank is implemented on face images with 8 different orientations and 5 different frequencies.

A Gabor filter can be one or two dimensional (2D). A 2D Gabor filter is expressed as a Gaussian modulated sinusoid in the spatial domain and as shifted Gaussian in the frequency domain. Recent studies on modeling of visual cortical cells [10] suggest a tuned band pass filter bank structure. These filters are found to have Gaussian transfer functions in the frequency domain. Thus, taking the Inverse Fourier Transform of this transfer function gives characteristics closely resembling Gabor filters.

A well designed Gabor filter bank can capture the relevant frequency spectrum in all directions. Phase can be taken as a feature because it contains information about the edge locations and other such details in the image; amplitude at every pixel can be taken as a feature as it contains some oriented frequency spectrum at every point of the image. We can extract many meaningful features using the Gabor filter family. Experimental results in texture analysis and character analysis demonstrate these features in the capture of local information with the different frequencies and orientations in the image [15].

The Gabor filter is a Gaussian (with variances S_x and S_y along x and y axes respectively) modulated by a complex sinusoid (with centre frequencies U and V along x and y - axes respectively) described by the following equation:-

$$g(x,y) = \frac{1}{2\pi S_x S_y} \exp\left[-\frac{1}{2}\left\{\left(\frac{x}{S_x}\right)^2 + \left(\frac{y}{S_y}\right)^2\right\} + 2\pi j(Ux + Vy)\right] \quad (1)$$

The variance terms S_x and S_y dictates the spread of the band pass filter centered at the frequencies U and V in the frequency domain. This filter is complex and the plot of the Real and Imaginary parts [15] of $g(x, y)$ is shown in Figure 1:-

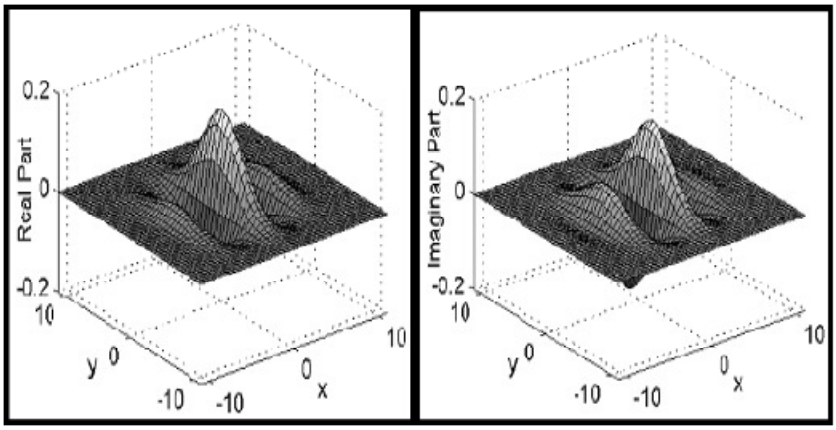


Figure 1: Plot of Real and Imaginary part of Gabor filter

It is found that as the 2D Gabor filter is applied to the images, the edges are smoothened out in all directions due to the presence of the Gaussian term. Each filter can be designed to pick out particular image features in orientation and the required frequency.

A Gabor filter can be best described by the following parameters:

1. The S_x and S_y of the Gaussian explain the shape of the base (circle or ellipse).

2. The frequency (f) of the sinusoid.

3. The orientation (Θ) of the applied sinusoid

Figures 2 and Figure 3 show examples of various Gabor filters

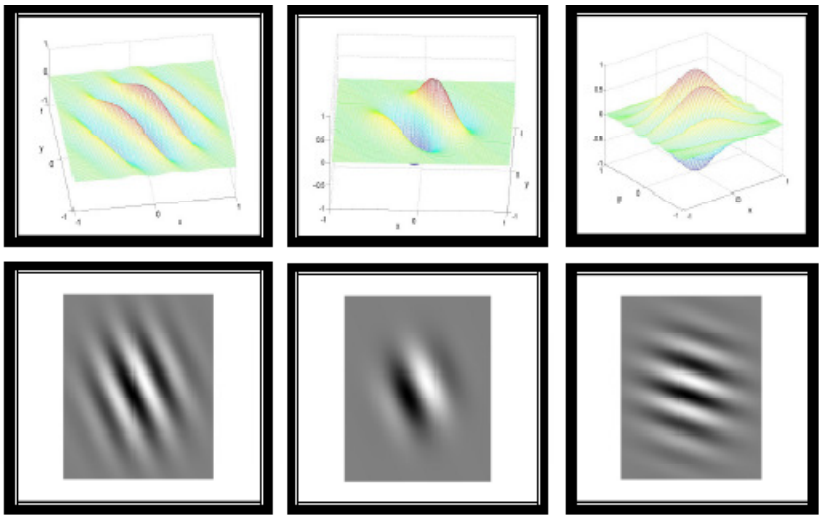


Figure 2: Figures (a), (b), (c) are examples of Gabor filter with different frequencies and orientations. Top row shows their 3D plots and the bottom row, the intensity plots of their amplitude along the image plane.

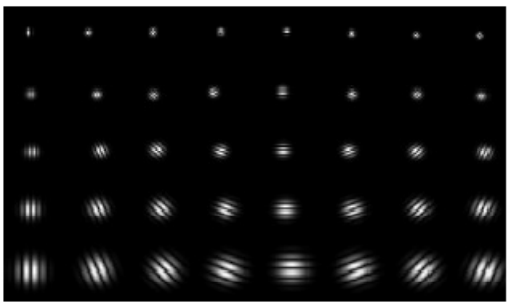


Figure 3: Gabor filters: Real part of the Gabor kernels at five scales and eight orientations

Figure 5 and Figure 6 show the effect of applying a particular Gabor filter on Figure 4 which is an image with lines at various angles. The highlighted lines in Figure 5 and Figure 6 shows the way the Gabor filter exaggerates lines at particular orientations.

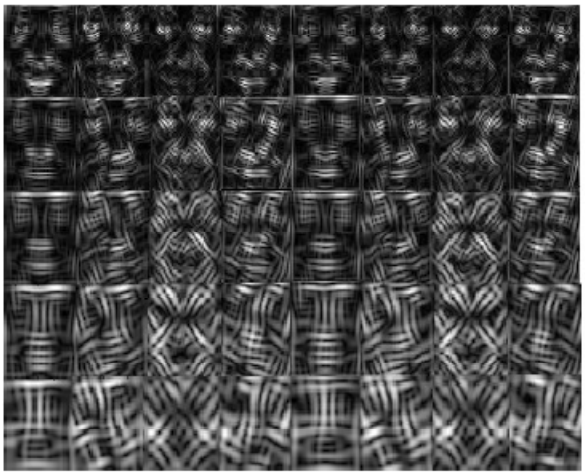
Figure 4: Image with lines at various angles

Figure 5: Frequency, f = 12.5 and orientation, θ= 135 degrees

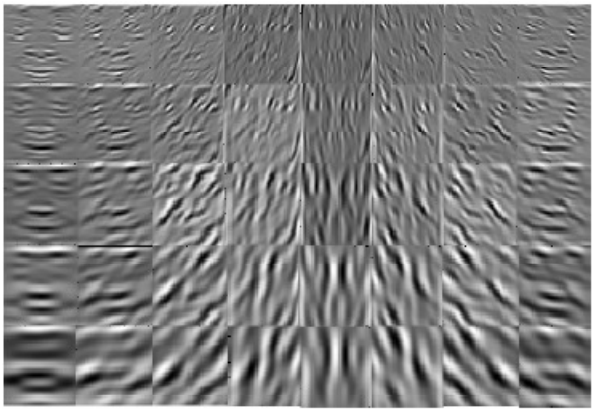
Figure 6: Frequency, f =25 and orientation, θ=0 degrees

Figure 8 shows the effect of applying a variety of Gabor filters to the image shown in Figure 7. Note how the features at particular orientations are exaggerated.

Figure 7: Sample Image of size 64 × 64



(a)



(b)

Figure 8: Convolution outputs of a sample image shown in Figure 7 and the Gabor kernels (Fig. 3). (a) Magnitude part of the convolution outputs. (b) Real part of the convolution outputs.

Analytical methods make use of Gabor jets at specific points on the face which are vital feature points (fiducial points). There are different methods for identifying or locating these feature points. For elastic graph based analytic methods, a graph is first placed at an initial location and deformed using jets to optimize its similarity with a model graph. Non-graph based methods locate feature points manually or by color or edge etc. Once the location process is completed, recognition can then be performed using Gabor jets extracted from those feature points [13].

Holistic methods on the other hand normally extract features from the whole face image. An augmented Gabor feature vector is thus created of a size far greater than the original data for the image. Every pixel is then represented by a vector of size 40 and demands dimensionality reduction before further processing. So a 64 × 64 image is transformed to size 64 × 64 × 5 × 8. Thus, the feature vector consists of all useful information extracted from different frequencies, orientations and from all locations, and hence is very useful for expression recognition. Once the feature vector is obtained, it can be handled in various ways. We have performed the following operations and any one of them can be used for the feature extraction:

a) The final image can be of the average of the magnitudes of the Gabor filter coefficients at each location in the filter bank output.

b) The pixel value in the final image would be the *L2 max* norm value of the feature vector obtained from the Gabor filter bank

The *L2 max* norm Superposition principle is used on the outputs of the filter bank and the figure 10 shows the output for the original image of figure 9. Similarly the outputs of the 40 filter banks can also be averaged or summed to give an output as in figure 11 shown below.

Figure 9: Original Image used for the Filter bank

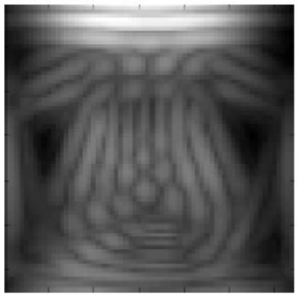


Figure 10: Superposition output (*L2 max* norm)

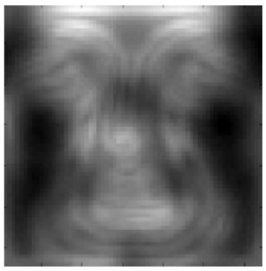


Figure 11: Average Output

2.2 Classification using Support Vector Machines

A number of classifiers can be used in the final stage for classification. We have concentrated on the Support Vector Machine. Support Vector Machines (**SVM**) are a set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classifiers. The SVM classifier tries to find the plane which separates two different classes such that it is equidistant from the members of either class which are nearest to the plane. SVM's are used extensively for a lot of classification tasks such as: handwritten digit recognition [4] or Object Recognition [2].

A SVM implicitly transforms the data into a higher dimensional data space (determined by the kernel) which allows the classification to be accomplished more easily. We have used the LIBSVM tool (Chih-Chih, 2001) for SVM classification.

2.3 Principal Component Analysis

Principal Component Analysis (PCA) transforms higher dimensional datasets into lower dimensional uncorrelated outputs by capturing linear correlations among the data, and preserving as much information as possible in the data. PCA transforms data from the original coordinate system to the principal axes coordinate system such that the principal axis passes through the maximum possible variance in the data. The second principal axis passes through the next largest possible variance and this is orthogonal to the first axis. This is repeated for the next largest possible variances and so on. All these axes are orthogonal to each other. On performing this PCA on the high dimensional data, Eigen values or principal components are thus obtained [14]. The required dimensionality reduction is obtained by retaining only the first few principal components.

The Principal Components have the following properties: They can be ranked by decreasing order of "importance". The first few most "important" Principal Components account for most of the information in the data. In other words, one may then discard the original data set, and replace it with a new data set with the same observations, but fewer variables, without throwing away too much information.

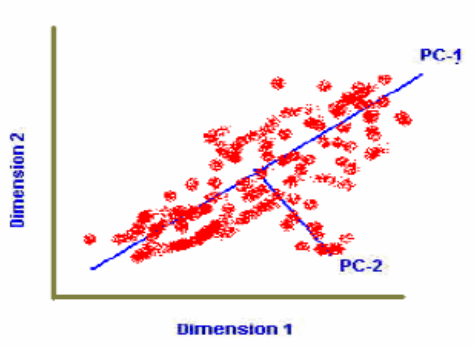


Figure 12: The first two consecutive principal components are shown.

2.4 Curvilinear Component Analysis

Curvilinear Component Analysis (CCA) is a non-linear projection method that preserves distance relationships in both input and output spaces. CCA is a useful method for redundant and non linear data structure representation and can be used in dimensionality reduction. CCA is useful with highly non-linear data, where PCA or any other linear method fails to give suitable information [6].

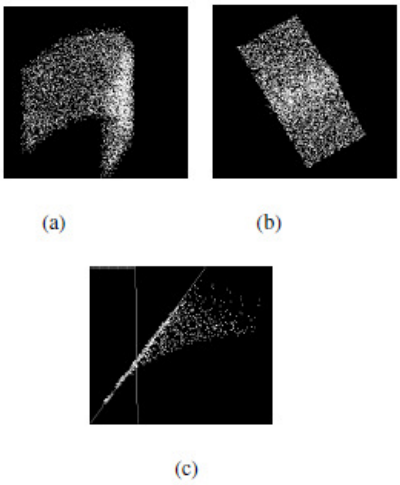
The D –dimensional input X should be mapped onto the output p-dimensional space Y. Their d -dimensional output vectors { y_i } should reflect the topology of the inputs { x_i }. In order to do that, Euclidean distances between the x_i's are considered. Corresponding distances in the output space y_i's is calculated such that the distance relationship between the data points is maintained.

CCA puts more emphasis on maintaining the short distances than the longer ones. Formally, this reasoning leads to the following error function:

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(d_{i,j}^{X} - d_{i,j}^{X}\right)^{2} F_{\lambda}\left(d_{i,j}^{Y}\right) \quad \forall \quad j \neq i \qquad (2)$$

where $d_{i,j}^{X}$ and $d_{i,j}^{Y}$ are the Euclidean distances between the points i and j in the input space *X* and the projected output space *Y* respectively and *N* is the number of data points. $F\left(d_{i,j}^{Y}\right)$ is the neighbourhood function, a monotonically decreasing function of distance. In order to check that the relationship is maintained a plot of the distances in the input space and the output space ($dy - dx$ plot) is produced. For a well maintained topology, dy should be proportional to the value of dx at least for small values of $dy's$.

Figure 13 shows CCA projections for the 3D data taken initially. The $dy - dx$ plot shown is good in the sense that the smaller distances are very well matched [6].



(a) (b)

(c)

Figure 13: (a) 3D horse shoe dataset (b) The 2D CCA projection of the horse shoe dataset (c) $dy - dx$ plot of the projection showing that small distances are maintained, although it is not possible to maintain the larger distances.

2.5 Intrinsic Dimension

One problem with CCA is deciding how many dimensions the projected space should occupy and one way of obtaining this is to use intrinsic dimension of the data manifold. The Intrinsic Dimension (ID) can be defined as the minimum number of free variables required to define data without any significant information loss. Due to the possibility of correlations among the data, both linear and nonlinear, a D-dimensional dataset may actually lie on a d-dimensional manifold ($D \geq d$). The ID of such data is then said to be d. There are various methods of calculating the ID; here we use the correlation Dimension [8] to calculate the ID of face image dataset.

3. EXPERIMENTS AND RESULTS

We experimented on 120 faces (60 male and 60 female) each with two classes, namely, Neutral and smiling (60 faces for each expression). The images are from The FERET dataset [12].

Figure 14: Example FERET images used in our experiments which are cropped to the size of 128 × 128 to extract the facial region and reduced to 64 × 64 for all experiments.

The training set was 80 faces (with 40 female, 40 male and equal numbers of them with neutral and smiling). Two test sets were created. In both test sets the number of each type of face is balanced. For example, there were 5 smiling male faces, 5 smiling female faces and 5 neutral male faces. With all faces aligned based on their eye location, a 128 × 128 image was cropped from the original (150 × 130). The resolution of these faces is reduced to 64 × 64. For the purpose of comparison, the SVM classification was performed on the raw face images (64 × 64), raw faces reduced in dimensionality with PCA, raw faces reduced in dimensionality by CCA, Gabor pre-processed images, Gabor pre-processed images reduced by PCA and Gabor pre-processed images reduced by CCA. For PCA reduction we use the first few principal components which account for 95% of the total variance of the data, and project the data onto these principal components. This resulted in using 66 components of the raw dataset and 35 components in the Gabor pre-processed dataset. As CCA is highly non-linear dimensionality reduction technique, we use the intrinsic dimensionality technique and reduce the components to its Intrinsic Dimension. The Intrinsic dimension of the raw faces was 14 and that of Gabor pre-processed images was 11. The classification results are shown in Table 1. The first set has images which are easily discernible smiling faces. The second test set has smiling and neutral faces, but the smiling faces are not easily discernible.

Figures 15 and 16 show the first 5 eigenfaces from the neutral faces (top row), smiling faces (bottom row) and the complete training set.

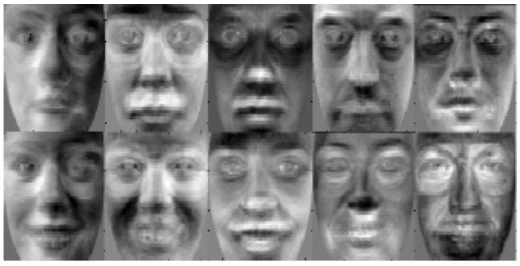


Figure 15: The top row shows the first 5 Eigenfaces of all the neutral faces of the data set. The bottom row shows the first 5 Eigenfaces of the smiling faces of the data set.

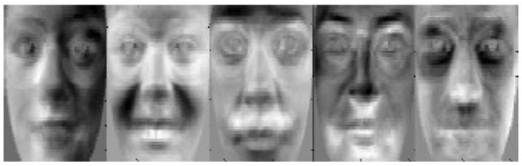


Figure 16: The first 5 Eigenfaces of the whole set of faces (male and female with equal number of smile and neutral faces).

The SVM was trained in the following way:

1. Transforming the data to a format required for using the SVM software package - LIBSVM -2.83 (Chih-Chung, 2001).

2. Perform simple scaling on the data so that all the features or attributes are in the range [-1, +1].

3. Choose a kernel. We have used a RBF kernel, $k(x, y) = e^{-\gamma|x-y|^2}$.

4. Perform five fold cross validation with the specified kernel to find the best values of the cost parameter C and γ.

5. Use the best parameter value of C and γ to train the whole training set.

6. Finally test the trained classifier using the test sets.

The results of the classification are as in Table 1:

Table 1: SVM Classification accuracy of raw faces and Gabor pre-processed images with PCA and CCA dimensionality reduction techniques.

SVM% accuracy	Test set1	Test set2
Raw faces	95	80
Raw with PCA66	90	75
Raw with CCA14	90	80
Gabor pre-processed faces	95	80
Gabor with PCA35	70	60
Gabor with CCA11	95	80

The PCA, being a linear dimensionality reduction technique, did not do quite as well when compared to the CCA. With CCA there was good generalization, but the key point to be noted here is the number of components used for the classification. The CCA makes use of just 14 components with raw faces and just 11 components with the Gabor pre-processed images to get good classification results. The Gabor filters have picked up the required features very well to help the more non linear dimensionality reduction technique such as the CCA to perform better. Some examples of misclassifications are shown in Figure 17.

Figure 17: Examples of the misclassified set of faces. Top row shows smiling faces wrongly classified as neutral. Bottom row shows neutral faces wrongly classified as smiling.

The reason for this misclassification probably is due to the relatively small size of training set. For example, the mustachioed face in the middle of the bottom row is misclassified as smiling. The only mustachioed face in the training set is of the same man smiling.

4. CONCLUSIONS

Identifying facial expressions is a challenging and interesting task. Our experiment shows that identification from raw images can be performed very well. However, with a larger data set, it may be computationally intractable to use the raw images. It is therefore important to reduce the dimensionality of the data. A linear method such as PCA does not appear to be sufficiently tunable to identify features that are relevant for facial expression characterization. However, on performing Gabor preprocessing on the images and following it with the CCA, there was good generalization in spite of the massive reduction in dimensionality. The most remarkable finding in this study is that the facial expression can be identified with just 11 components found by CCA.

5. REFERENCES

[1] Early Batty, B., & Taylor, M.J (2003) Early processing of the six basic facial emotional expressions, Cognitive Brain Research, 17.

[2] Blanz, V., Sch¨olkopf, B., B¨ulthoff, H., Burges, C., Vapnik, V. & Vetter, T., "Comparison of view-based object recognition algorithms using realistic 3D models", Proc. Int. Conf. on Artificial Neural Networks 1996, 251-256.

[3] Chih-Chung Chang & Chih-Jen Lin, LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[4] Cortes, C. & Vapnik, V.(1995) Support Vector Networks, Machine Learning, 20, 273-297.

[5] Daugman, J.G, "Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two dimensional visual cortical filters", Journal of Optical.Society of .America .A, Vol 2, No.7.July 1985.

[6] Demartines, P. and J. Hérault (1997). "Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets." IEEE Transactions on Neural Networks 8(1): 148-154.

[7] Ekman, P., & Friesen, W. V. (1971) Constants across cultures in the face of the emotion, Journal of Personality and Social Psychology, 17.

[8] Grassberger, P. & Proccacia, I. (1983). Measuring the strangeness of strange attractors. *Physica D, 9.*

[9] Jain, A.K & Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. Pattern Recognition 24(12).

[10] Kulikowski (1982). "Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex." Biological Cybernetics 43(3): 187-198.

[11] Movellan, J.R. (2002). Tutorial on Gabor Filters. http://mplab.ucsd.edu/tutorials/pdfs/gabor.pdf

[12] Philips, P.J., Grother,P., Micheals,R.J., Blackburn, D., Tabassi, E.&Bone, M.(2003). Face recognition vendor test 2002: overview and summary. Available at www.frvt.org, 2003.

[13] Shen et.al., (2006) Review on Gabor wavelets for face recognition Pattern Anal. Applic. (2006) 9:273–292

[14] Smith, L.I., (2002) Tutorial on Principal Component Analysis.Available at: http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

[15] Zheng, D., Zhao,Y., Wang, J.,(2004a) Features Extraction Using A Gabor Filter Family, Proceedings of the sixth lasted International conference, Signal and Image processing, Hawaii, USA.

Recognizing Facial Expressions: A comparison of Computational approaches

Aruna Shenoy, Tim M Gale [1,2*], Neil Davey [1*],
Bruce Christiansen[1*], and Ray Frank [1*]

1* School of Computer Science, University of Hertfordshire,
United Kingdom, AL10 9AB
{a.1.shenoy, t.gale, r.j.frank, n.davey}@herts.ac.uk
2*Department of Psychiatry, Queen Elizabeth II Hospital,Welwyn Garden City
Herts,AL7 4HQ UK

Abstract. Recognizing facial expressions are a key part of human social interaction,and processing of facial expression information is largely automatic, but it is a non-trivial task for a computational system. The purpose of this work is to develop computational models capable of differentiating between a range of human facial expressions. Raw face images are examples of high dimensional data, so here we use some dimensionality reduction techniques: Linear Discriminant Analysis, Principal Component Analysis and Curvilinear Component Analysis. We also preprocess the images with a bank of Gabor filters, so that important features in the face images are identified. Subsequently the faces are classified using a Support Vector Machine. We show that it is possible to differentiate faces with a neutral expression from those with a smiling expression with high accuracy. Moreover we can achieve this with data that has been massively reduced in size: in the best case the original images are reduced to just 11 dimensions.
Keywords: Facial Expressions, Image Analysis, Classification, Dimensionality Reduction.

1 Introduction

According to Ekman and Friesen [1] there are six easily discernible facial expressions: anger, happiness(smile), fear, surprise, disgust and sadness, apart from neutral. Moreover these are readily and consistently recognized across different cultures [2]. In the work reported here we show how a computational model can identify facial expressions from simple facial images. In particular we show how smiling faces and neutral faces can be differentiated. Data presentation plays an important role in any type of recognition. High dimensional data is normally reduced to a manageable low dimensional data set. We perform dimensionality reduction and classification using Linear Discriminant Analysis and also dimensionality reduction using Principal Component Analysis (PCA) and Curvilinear Component Analysis (CCA). PCA is a linear projection technique and it may be more appropriate to use a non linear Curvilinear Component Analysis (CCA) [3]. The Intrinsic Dimension (ID) [4], which is the true dimension of the data, is often much less than the original dimension of the data. To use this efficiently, the actual dimension of the data must be estimated. We use the Correlation Dimension to estimate

the Intrinsic Dimension. We compare the classification results of these methods with raw face images and of Gabor Pre-processed images [5],[6]. The features of the face (or any object for that matter) may be aligned at any angle. Using a suitable Gabor filter at the required orientation, certain features can be given high importance and other features less importance. Usually, a bank of such filters is used with different parameters and later the resultant image is a L2 max (at every pixel the maximum of feature vector obtained from the filter bank) superposition of the outputs from the filter bank.

2 Background

We begin with a simple experiment to classify two expressions: neutral and smiling. We use Linear Discriminant Analysis (LDA) for dimensionality reduction and classification. We also use a variety of other dimensionality reduction techniques, a Support Vector Machine (SVM) [7] based classification technique and these are described below.

2.1 Linear Discriminant Analysis (LDA)

For a two class problem, LDA is commonly known as Fisher Linear discriminant analysis after Fisher [8] who used it in his taxonomy based experiments. Belhuemer was the first to use the LDA on faces and used it for dimensionality reduction [9] and it can be used as a classifier. LDA attempts to find the linear projection of the data that produces maximum between class separation and minimum within class scatter. In the simple example shown in Figure 1, a projection on to the vertical axis separates the two classes whilst minimizing the within class scatter. Conversely, a projection onto horizontal axis does not separate the classes. Formally the algorithm can be described as follows. The between class scatter covariance matrix is given by:

$$\mathbf{S_B} = (\mathbf{m_2} - \mathbf{m_1})(\mathbf{m_2} - \mathbf{m_1})^{\mathbf{T}} \tag{1}$$

The within class covariance matrix is given by:

$$\mathbf{S_W} = \sum_{i=1}^{C_j} \sum_{n \epsilon C_k} (\mathbf{X^n} - \mathbf{m_i})(\mathbf{X^n} - \mathbf{m_i})^{\mathbf{T}} \tag{2}$$

where m_1 and m_2 are the means of the datasets of the class 1 and 2 respectively. C is the number of classes and C_k is the k_{th} class. The eigenvector solution of $\mathbf{S}_w^{-1}\mathbf{S}_B$ gives the projection vector which in the context of face image classification is known as the Fisher face.

2.2 Gabor Filters

A Gabor filter can be applied to images to extract features aligned at particular orientations. Gabor filters possess the optimal localization properties in both spatial and frequency domains, and they have been successfully used in many applications [10]. A

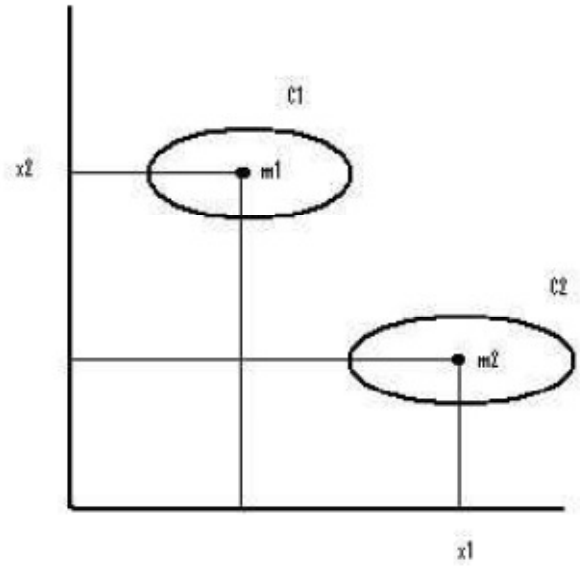


Fig. 1. The figure shows the classes which are overlapping along the direction of x1. However, they can be projected on to direction x2 where there will be no overlap at all.

Gabor filter is a function obtained by modulating a sinusoidal with a Gaussian function. The useful parameters of a Gabor filter are orientation and frequency. The Gabor filter is thought to mimic the simple cells in the visual cortex. The various 2D receptive field profiles encountered in populations of simple cells in the visual cortex are well described by an optimal family of 2D filters [11]. In our case a Gabor filter bank is implemented on face images with 8 different orientations and 5 different frequencies. Recent studies on modeling of visual cortical cells [12] suggest a tuned band pass filter bank structure. Formally, the Gabor filter is a Gaussian (with variances S_x and S_y along x and y-axes respectively) modulated by a complex sinusoid (with centre frequencies U and V along x and y-axes respectively) and is described by the following equation 3

$$g(x,y) = \frac{\exp\left[-\frac{1}{2}\left[\left(\frac{x}{S_x}\right)^2 + \left(\frac{y}{S_y}\right)^2\right] + 2\delta j(Ux + Vy)\right]}{2\delta S_x S_y} \tag{3}$$

The variance terms and dictates the spread of the band pass filter centered at the frequencies U and V in the frequency domain. This filter is complex in nature.

A Gabor filter can be described by the following parameters: The S_x and S_y of the Gaussian explain the shape of the base (circle or ellipse), frequency (f) of the sinusoid, orientation (θ) of the applied sinusoid Figure 2 shows examples of various Gabor filters. Figure 3b) shows the effect of applying a variety of Gabor filters shown in Figure 2 to the sample image shown in Figure 3a). Note how the features at particular orientations are exaggerated.

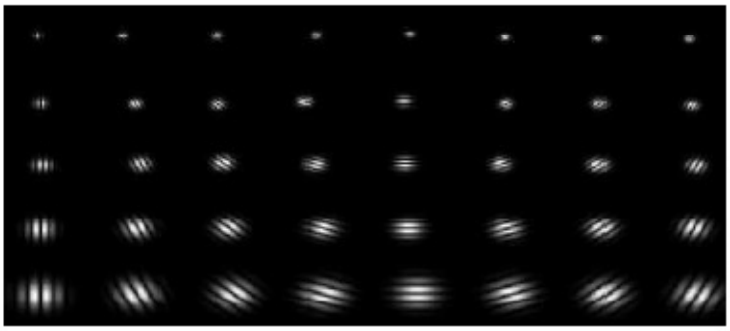
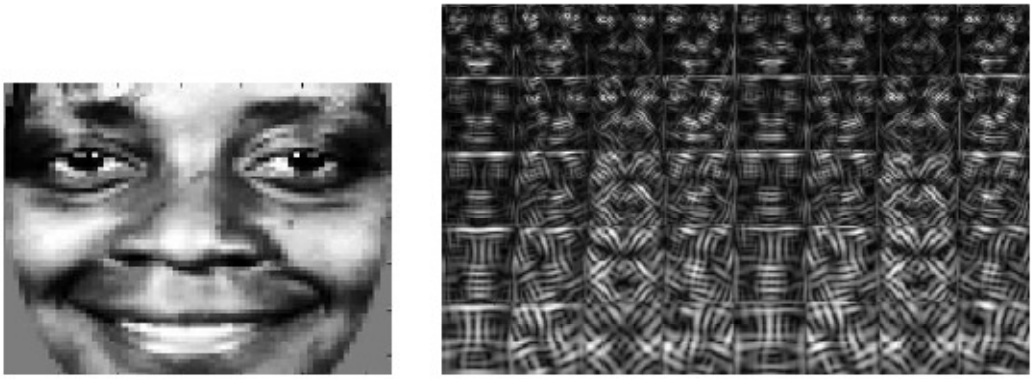


Fig. 2. Gabor filters: Real part of the Gabor kernels at five scales and eight orientations



(a) (b)

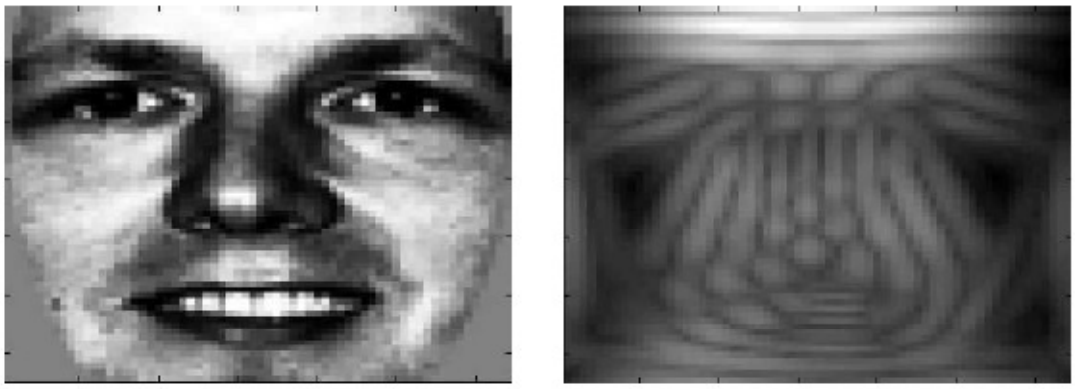
Fig. 3. *(a)* Original face image,*(b)* Forty Convolution outputs of Gabor

An augmented Gabor feature vector is created of a size far greater than the original data for the image. Every pixel is then represented by a vector of size 40 and demands dimensionality reduction before further processing. So a 64 x 64 image is transformed to size 64 x 64 x 5 x 8. Thus, the feature vector consists of all useful information extracted from different frequencies, orientations and from all locations, and hence is very useful for expression recognition.

Once the feature vector is obtained, it can be handled in various ways. We simply take the L2 max norm for each pixel in the feature vector. So that the final value of a pixel is the maximum value found by any of the filters for that pixel. The L2 max norm Superposition principle is used on the outputs of the filter bank and the Figure 4b) shows the output for the original image of Figure 4 a).

2.3 Curvilinear Component Analysis

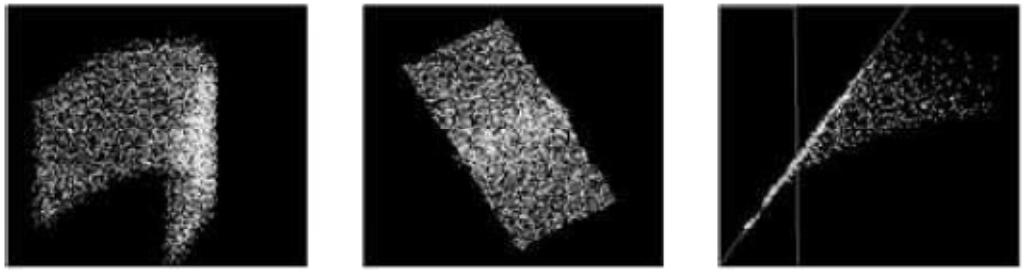
Curvilinear Component Analysis (CCA) is a non-linear projection method that preserves distance relationships in both input and output spaces. CCA is a useful method



(a) (b)

Fig. 4. a) Original Image used for the Filter bank b)Supposition Output(L2 max norm)

for redundant and non linear data structure representation and can be used in dimensionality reduction. CCA is useful with highly non-linear data, where PCA or any other linear method fails to give suitable information [3]. The D-dimensional input X should be mapped onto the output p-dimensional space Y. Their d-dimensional output vectors y_i should reflect the topology of the inputs x_i. In order to do that, Euclidean distances between the x_i 's are considered. Corresponding distances in the output space y_i's is calculated such that the distance relationship between the data points is maintained.



(a) b) (c)

Fig. 5. (a) 3D horse shoe dataset (b) 2D CCA projection (c) plot.

CCA puts more emphasis on maintaining the short distances than the longer ones. Formally, this reasoning leads to the following error function:

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[d_{i,j}^{X} - d_{i,j}^{Y}\right]^2 F_\lambda\left[d_{i,j}^{Y}\right] \;\forall j \neq i \tag{4}$$

where $d_{i,j}$ and $d^Y_{i,j}$ are the Euclidean distances between the points i and j in the input space X and the projected output space Y respectively and N is the number of data points. $F_\lambda\left[d^Y_{i,j}\right]$ is the neighbourhood function, a monotonically decreasing function of distance. In order to check that the relationship is maintained a plot of the distances in the input space and the output space $(d_y - d_x)$plot is produced. For a well maintained topology,d_y should be proportional to the value of d_x at least for small values of d_y's. Figure 5 shows CCA projections for the 3D data horse shoe data. The $(d_y - d_x)$ plot shown is good in the sense that the smaller distances are very well matched [5].

2.4 Intrinsic Dimension

One problem with CCA is deciding how many dimensions the projected space should occupy, and one way of obtaining this is to use the intrinsic dimension of the data manifold. The Intrinsic Dimension (ID) can be defined as the minimum number of free variables required to define data without any significant information loss. Due to the possibility of correlations among the data, both linear and nonlinear, a D-dimensional dataset may actually lie on a d-dimensional manifold (D $\geq$ d). The ID of such data is then said to be d. There are various methods of calculating the ID; here we use the correlation Dimension [8] to calculate the ID of face image dataset.

3 Classification using Support Vector Machines

A number of classifiers can be used in the final stage for classification. We have concentrated on the Support Vector Machine. Support Vector Machines (SVM) are a set of related supervised learning methods used for classification and regression. SVM's are used extensively for many classification tasks such as: handwritten digit recognition [14] or Object Recognition [15]. A SVM implicitly transforms the data into a higher dimensional data space (determined by the kernel) which allows the classification to be accomplished more easily. We have used the LIBSVM tool [7] for SVM classification.

4 Experiments and Results

We experimented on 120 faces (60 male and 60 female) each with two classes, namely: neutral and smiling (60 faces for each expression). The images are from The FERET dataset [16] and some examples are shown in Figure 76. The training set was 80 faces (with 40 female, 40 male and equal numbers of them with neutral and smiling). Two test sets were created. In both test sets the number of each type of face is balanced. For example, there were 5 smiling male faces and 5 smiling female faces. The first set has images which are easily discernible smiling faces. The second test set has smiling and neutral faces, but the smiling faces are not easily discernible. With all faces aligned based on their eye location, a 128 x 128 image was cropped from the original (150 x 130). The resolution of these faces is then reduced to 64 x 64.

A LDA projection was made onto the Fisher face shown in Figure 7. The two test sets were then classified by using the nearest neighbor in the test set in the projection

Fig. 6. Example FERET images used in our experiments which are cropped to the size of 128 x 128 to extract the facial region and reduced to 64 x 64 for all experiments.

Table 1. Classification accuracy of raw faces using LDA.

Accuracy %	Test Set 1	Test Set 2
LDA	95	75

space. The results are as in Table 1. Figure 7 shows the Fisher face obtained by performing the LDA on the training data set. For PCA reduction we use the first few principal components which account for 95% of the total variance of the data, and project the data onto these principal components. This resulted in using 66 components of the raw dataset and 35 components in the Gabor pre-processed dataset. As CCA is a highly non-linear dimensionality reduction technique, we use the intrinsic dimensionality technique and reduce the components to its Intrinsic Dimension. The Intrinsic Dimension of the raw faces was approximated as 14 and that of Gabor pre- processed images was 11. The classification results are shown in Table 2. Figure 8 shows the Eigenfaces obtained by the PCA technique.

After dimensionality reduction a standard SVM (with Gaussian kernel) was used to classify the images. The parameters of the SVM were optimized using 5-fold validation.

The results of the classification are as in Table 2. The PCA, being a linear dimensionality reduction technique, did not do quite as well as CCA. With CCA there was good generalization, but the key point to be noted here is the number of components used for the classification. The CCA makes use of just 14 components with raw faces and just 11 components with the Gabor pre-processed images to get good classification results.

This suggests that the Gabor filters are highlighting salient information which can be encoded in a small number of dimensions using CCA. Some examples of misclassifications are shown in Figure 9. The reason for these misclassifications is probably due to the relatively small size of training set. For example, the mustachioed face in the

Table 2. SVM Classification accuracy of raw faces and Gabor pre-processed images with PCA and CCA dimensionality reduction techniques.

SVM Accuracy %	Test Set 1	Test Set 2
Raw Faces(64x64)	95	80
Raw with PCA66	90	75
Raw with CCA14	90	80
Gabor pre-processed Faces (64x64)	95	80
Gabor with PCA35	70	60
Gabor with CCA11	95	80

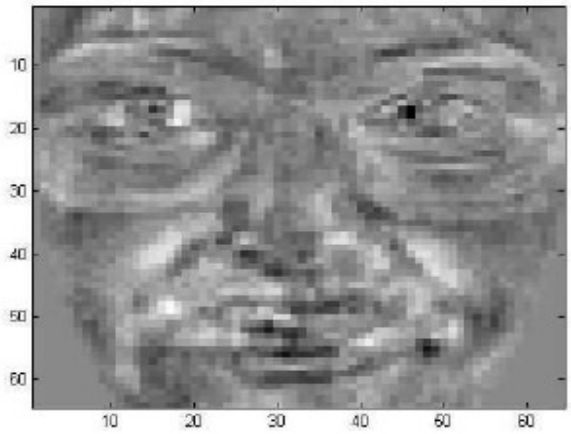


Fig. 7. Fisher face obtained for a dataset with two classes namely, Neutral and Smiling.

Fig. 8. The first 5 eigenfaces of the complete data set.

middle of the bottom row is misclassified as smiling. The only mustachioed face in the training set is of the same man smiling.

5 Conclusion

Identifying facial expressions is a challenging and interesting task. Our experiment shows that identification from raw images can be performed very well. However, with a larger data set, it may be computationally intractable to use the raw images. It is therefore important to reduce the dimensionality of the data. Performing classification using

Fig. 9. Examples of the misclassified set of faces. Top row shows smiling faces wrongly classified as neutral. Bottom row shows neutral faces wrongly classified as smiling.

LDA was a trivial task and the result was very impressive. It is interesting to see the effect size for each pixel in the image.

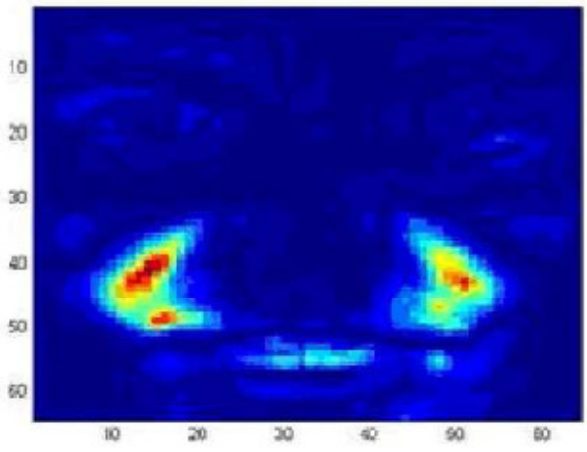


Fig. 10. Encoding face.

In other words which pixels discriminate most between smile and neutral faces can be seen and the result of this analysis is shown in Figure 10. The Creasing of the cheeks is diagnostic of smiling faces; teeth may also be an important indicator, though to a lesser extent. A linear method such as PCA does not appear to be sufficiently tunable to identify features that are relevant for facial expression characterization. Though the result of classification with LDA is impressive, for large datasets with face images, PCA

needs to be done prior to the LDA. However, on performing Gabor preprocessing on the images and following it with the CCA, there was good generalization in spite of the massive reduction in dimensionality. The most remarkable finding in this study is that the facial expression can be identified with just 11 components found by CCA. Future work will include extend the experiment to a larger data set and for other expressions.

References

1. Ekman, P. and W.V. Friesen, Constants across cultures in the face of the emotion. Journal of Personality and Social Psychology, 1971. 17.
2. Batty, B., M.J. Taylor, and, Early processing of the six basic facial emotional expressions. Cognitive Brain Research, 2003. 17.
3. Demartines, P. and D.J. Herault, Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. IEEE Transactions on Neural Networks, 1997. 8(1): p. 148-154.
4. Grassberger, P. and I. Proccacia, Measuring the strangeness of strange attractors. Physica D, 1983. 9.
5. Jain, A.K. and F. Farrokhnia, Unsupervised texture segmentation using Gabor filters. Pattern Recognition, 1991. 24(12).
6. Movellan, J.R., Tutorial on Gabor Filters. 2002.
7. Chang, C.-C. and Chih-Jen Lin, LIBSVM: a library for support vector machines. 2001.
8. Fisher, R.A., The use of mutliple measures in anatomical problems. Ann. Eugenics, 1936. 7: p. 179-188.
9. Belhumeur and Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on patternAnalysis and Machine Intelligence, 1997. 19(7): p. 711-720.
10. Zheng, D., Y. Zhao, and J. Wang, Features Extraction using A Gabor Filter Family. Proceedings of the sixth Lasted International conference, Signal and Image processing, Hawaii, 2004.
11. Daugman, J.G., Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two dimensional visual cortical filters. Journal of Optical.Society of.America.A, 1985. 2(7).
12. Kulikowski, Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. Biological Cybernetics, 1982. 43(3): p. 187-198.
13. Smith, L.I., Tutorial on Principal Component Analysis. 2002.
14. Cortes, C. and V. Vapnik, Support Vector Networks. Machine Learning, 1995. 20: p. 273-297.
15. Blanz, V., et al., Comparison of view-based object recognition algorithms using realistic 3D models. Proc. Int. Conf. on Artificial Neural Networks, 1996: p. 251-256.
16. Philips, P.J., et al., The FERET evaluation methodology for face recognition algorithms. Image and Vision Computing, 1998. 16(5): p. 295-306.

Discriminating Angry, Happy and Neutral facial Expression: A Comparison of computational models

Aruna Shenoy[1], Sue Anthony[2], Ray Frank[1], Neil Davey[1]
{a.1.shenoy, s.1.anthony, r.j.frank, n.davey}@herts.ac.uk

Abstract: Recognizing expressions are a key part of human social interaction, and processing of facial expression information is largely automatic for humans, but it is a non-trivial task for a computational system. The purpose of this work is to develop computational models capable of differentiating between a range of human facial expressions. Raw face images are examples of high dimensional data, so here we use two dimensionality reduction techniques: Principal Component Analysis and Curvilinear Component Analysis. We also preprocess the images with a bank of Gabor filters, so that important features in the face images are identified. Subsequently the faces are classified using a Support Vector Machine. We show that it is possible to differentiate faces with a neutral expression from those with a happy expression and neutral expression from those of angry expressions and neutral expression with high accuracy. Moreover we can achieve this with data that has been massively reduced in size: in the best case the original images are reduced to just 5 components with happy faces and 5 components with angry faces.

Keywords: Facial Expressions, Image Analysis, Classification, Dimensionality Reduction.

1. INTRODUCTION

According to Ekman and Friesen [1] there are six easily discernible facial expressions: anger, happiness, fear, surprise, disgust and sadness, apart from neutral. Moreover these are readily and consistently recognized across different cultures [2]. In the work reported here we show how a computational model can identify facial expressions from simple facial images. In particular, we show how happy faces with neutral faces and angry faces with neutral faces can be differentiated.

Data presentation plays an important role in any type of recognition. High dimensional data is normally reduced to a manageable low dimensional data set. We perform dimensionality reduction using Principal Component Analysis (PCA) and Curvilinear Component Analysis (CCA). PCA is a linear projection technique and it may be more appropriate to use a non linear Curvilinear Component Analysis (CCA) [3]. The Intrinsic Dimension (ID) [4], which is the true dimension of the data, is often much less than the original dimension of the data. To use this efficiently, the actual dimension of the data must be estimated. We use the Correlation Dimension to estimate the Intrinsic Dimension. We compare the classification results of these methods with raw face images and of Gabor Pre-processed images [5, 6]. The features of the face (or any object for that matter) may be aligned at any angle. Using a suitable Gabor filter at the required orientation, certain features can be given high importance and other features less importance. Usually, a bank of such filters is used

1. Department of Computer Science, University of Hertfordshire, Hatfield, AL10 9AB, UK.
2. Department of Psychology, University of Hertfordshire, Hatfield,AL10 9AB, UK.

with different parameters and later the resultant image is a *L2 max* (at every pixel the maximum of feature vector obtained from the filter bank) superposition of the outputs from the filter bank.

2 BACKGROUND

We perform feature extraction with Gabor filters and then use dimensionality reduction techniques such as Principal Component Analysis (PCA) and Curvilinear Component Analysis (CCA) followed by a Support Vector Machine (SVM) [7] based classification technique and these are described below.

2.1 Gabor Filters

A Gabor filter can be applied to images to extract features aligned at particular orientations. Gabor filters possess the optimal localization properties in both spatial and frequency domains, and they have been successfully used in many applications [8]. A Gabor filter is a function obtained by modulating a sinusoidal with a Gaussian function. The useful parameters of a Gabor filter are orientation and frequency. The Gabor filter is thought to mimic the simple cells in the visual cortex. The various 2D receptive field profiles encountered in populations of simple cells in the visual cortex are well described by an optimal family of 2D filters [9]. In our case a Gabor filter bank is implemented on face images with 8 different orientations and 5 different frequencies.

Recent studies on modeling of visual cortical cells [10] suggest a tuned band pass filter bank structure. Formally, the Gabor filter is a Gaussian (with variances S_x and S_y along x and y-axes respectively) modulated by a complex sinusoid (with centre frequencies U and V along x and y-axes respectively) and is described by the following equation:-

$$g(x,y) = \frac{\exp\left[-\frac{1}{2}\left\{\left(x/S_x\right)^2 + \left(y/S_y\right)^2\right\} + 2\pi j(Ux+Vy)\right]}{2\pi S_x S_y} \qquad (1)$$

The variance terms S_x and S_y dictates the spread of the band pass filter centered at the frequencies U and V in the frequency domain. This filter has real and imaginary part.

A Gabor filter can be described by the following parameters: The S_x and S_y of the Gaussian explain the shape of the base (circle or ellipse), frequency (f) of the

sinusoid, orientation (Θ) of the applied sinusoid. Figure 1 shows examples of various Gabor filters. Figure 2 b) shows the effect of applying a variety of Gabor filters shown in Figure 1 to the sample image shown in Figure 2 a). Note how the features at particular orientations are exaggerated.

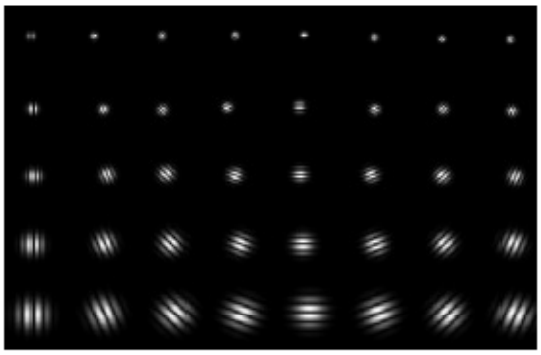
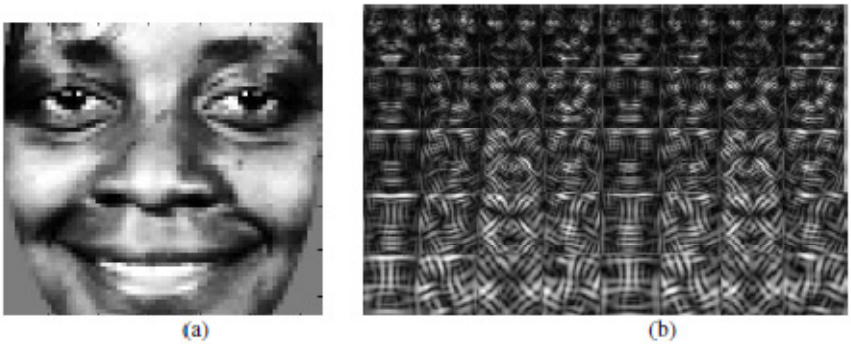


Figure 1: Gabor filters: Real part of the Gabor kernels at five scales and eight orientations

An augmented Gabor feature vector is created of a size far greater than the original data for the image. Every pixel is then represented by a vector of size 40 and demands dimensionality reduction before further processing. So a 63×63 image is transformed to size $63 \times 63 \times 5 \times 8$. Thus, the feature vector consists of all useful information extracted from different frequencies, orientations and from all locations, and hence is very useful for expression recognition.



(a) (b)

Figure 2: a) Original face image 63×63 (3969 dimensions). b) Forty Convolution outputs of Gabor filters.

Once the feature vector is obtained, it can be handled in various ways. We simply take the *L2 max* norm for each pixel in the feature vector. So that the final value of a pixel is the maximum value found by any of the filters for that pixel.

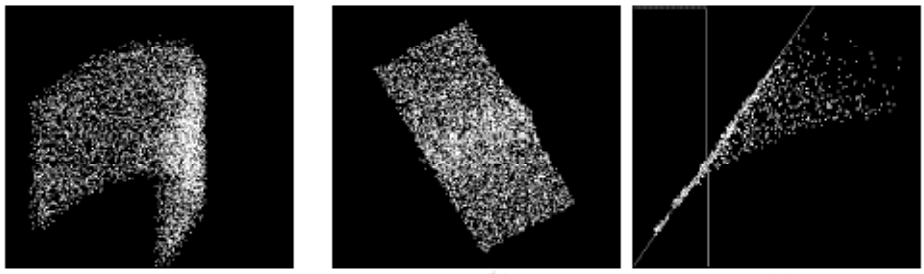
The *L2 max* norm Superposition principle is used on the outputs of the filter bank and the Figure 3 b) shows the output for the original image of Figure 3 a).

(a) (b)

Figure 3 a): Original Image used for the Filter bank b) Superposition output (*L2 max* norm)

2.2 Curvilinear Component Analysis

Curvilinear Component Analysis (CCA) is a non-linear projection method that preserves distance relationships in both input and output spaces. CCA is a useful method for redundant and non linear data structure representation and can be used in dimensionality reduction. CCA is useful with highly non-linear data, where PCA or any other linear method fails to give suitable information [3]. The D –dimensional input X should be mapped onto the output d-dimensional space Y. Their d-dimensional output vectors $\{ y_i \}$ should reflect the topology of the inputs $\{ x_i \}$. In order to do that, Euclidean distances between the x_i's are considered. Corresponding distances in the output space y_i's is calculated such that the distance relationship between the data points is maintained.



(a) (b) (c)

Figure 4: (a) 3D horse shoe dataset (b) 2D CCA projection (c) $dy - dx$ plot.

CCA puts more emphasis on maintaining the short distances than the longer ones. Formally, this reasoning leads to the following error function:

$$E = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(d_{i,j}^{X} - d_{i,j}^{Y} \right)^2 F_\lambda \left(d_{i,j}^{Y} \right) \forall \quad j \neq i \qquad (2)$$

where $d^X_{i,j}$ and $d^Y_{i,j}$ are the Euclidean distances between the points i and j in the input space X and the projected output space Y respectively and N is the number of data points. $F(d^Y_{i,j})$ is the neighbourhood function, a monotonically decreasing function of distance. In order to check that the relationship is maintained a plot of the distances in the input space and the output space ($dy - dx$ plot) is produced. For a well maintained topology, dy should be proportional to the value of dx at least for small values of $dy's$. Figure 4 shows CCA projections for the 3D data horse shoe data. The $dy - dx$ plot shown is good in the sense that the smaller distances are very well matched [3].

2.3 Intrinsic Dimension

One problem with CCA is deciding how many dimensions the projected space should occupy, and one way of obtaining this is to use the intrinsic dimension of the data manifold. The Intrinsic Dimension (ID) can be defined as the minimum number of free variables required to define data without any significant information loss. Due to the possibility of correlations among the data, both linear and nonlinear, a D-dimensional dataset may actually lie on a d-dimensional manifold ($D \geq d$). The ID of such data is then said to be d. There are various methods of calculating the ID; here we use the correlation Dimension [8] to calculate the ID of face image dataset.

2.4 Classification using Support Vector Machines

A number of classifiers can be used in the final stage for classification. We have concentrated on the Support Vector Machine. Support Vector Machines (SVM) are a set of related supervised learning methods used for classification and regression. SVM's are used extensively for many classification tasks such as: handwritten digit recognition [11] or Object Recognition [12]. A SVM implicitly transforms the data into a higher dimensional data space (determined by the kernel) which allows the classification to be accomplished more easily. We have used the LIBSVM tool [7] for SVM classification.

The SVM is trained in the following way:

1. Transform the data to a format required for using the SVM software package - LIBSVM -2.83 [7].

2. Perform simple scaling on the data so that all the features or attributes are in the range [-1, +1].

3. Choose a kernel. We used the RBF kernel,

$$k(x, y) = e^{-\gamma|x-y|^2}.$$

4. Perform fivefold cross validation with the specified kernel to find the best values of the cost parameter C and γ.

5. By using the best value of C and γ, train the model and finally evaluate the trained classifier using the test sets.

3. EXPERIMENTS AND RESULTS

We experimented on 264 faces (132 female and 132 male) each with three classes, namely: *Neutral* and *Happy* and *Angry* (88 faces for each expression). Neutral and Happy are used in one experiment and Neutral and Angry faces are used in another. The images are from the BINGHAMTON dataset [13] and some examples are shown in Figure 5. Two training and test sets are used. One training set had 132 faces (with 66 female, 66 male and equal numbers of them with neutral and happy expression) and another training set had 132 faces (with 66 female and 66 male and equal numbers of them with neutral and angry expression). The original 128×128 image was reduced to 63×63. As we have two training sets, we have two test sets. Each consists of 44 faces (11 female, 11 male and equally balanced number of expression: neutral with angry and neutral with happy).

For PCA reduction we always use the first principal components which account for 95% of the total variance of the data, and project the data onto these principal components - we call this is our standard PCA reduction. With Neutral and Happy faces, this resulted in using 100 components of the raw dataset and 23 components in the Gabor pre-processed dataset. With Neutral and Angry faces, this resulted in using 97 components of the raw dataset and 22 components in the Gabor pre-processed dataset. As CCA is a highly non-linear dimensionality reduction technique, we use the intrinsic dimensionality technique and reduce the components to its Intrinsic Dimension. The Intrinsic Dimension of the raw faces with Neutral and Happy was approximated as 6 and that of Gabor pre- processed images was 5. Likewise, the Intrinsic Dimension of the raw faces with Neutral and Angry was approximated as 5 and that of Gabor pre-processed images was 6. Figure 6 shows the Eigenfaces obtained by the PCA technique with raw faces (Happy with Neutral set and Angry with Neutral set).

Figure 5: Example BINGHAMTON images used in our experiments which are cropped to the size of 128 $\times$ 128 to extract the facial region and reduced to 63 $\times$ 63 for all experiments. The first row has examples of angry expression, middle row has happy expression and last row has images with neutral expression.

(a)

(b)

Figure 6: a) The first 5 eigenfaces of the neutral and happy data set. B) The first 5 eigenfaces of the neutral and angry data set.

The results of the SVM classification for Neutral and Happy are as in Table 1 and for Neutral and Angry are as in Table 2. The PCA, being a linear dimensionality reduction technique, did not do quite as well as CCA with happy and neutral data set; however, there has been no difference with the angry and neutral dataset. With CCA there was good generalization, but the key point to be noted here is the number of components used for the classification. The CCA makes use of just 6 components with raw faces get good classification result and 5 components with the Gabor pre-processed images with the neutral and happy dataset. With the angry and neutral dataset, the CCA makes use of 5 components with raw faces and 6 components with Gabor pre-processed faces with results comparable with the raw faces.

Table 1: SVM Classification accuracy of raw faces and Gabor pre-processed images with PCA and CCA dimensionality reduction techniques for Neutral and Happy dataset.

SVM% accuracy	Happy and Neutral (44 faces)
Raw faces	100 (44/44)
Raw with PCA100	88.64 (39/44)
Raw with CCA6	93.18 (41/44)
Gabor pre-processed faces	90.91(40/44)
Gabor with PCA23	95.45 (42/44)
Gabor with CCA5	68.18 (30/44)

Table 2: SVM Classification accuracy of raw faces and Gabor pre-processed images with PCA and CCA dimensionality reduction techniques for Neutral and Angry dataset.

SVM% accuracy	Angry and Neutral (44 faces)
Raw faces	79.54 (35/44)
Raw with PCA97	68.18 (30/44)
Raw with CCA5	68.18 (30/44)
Gabor pre-processed faces	68.18 (30/44)
Gabor with PCA22	61.36 (27/44)
Gabor with CCA6	63.64 (28/44)

The classification results for the Neutral and Happy face images shown in Table 1, indicates best classification using raw faces. The intrinsic dimensionality of the raw images is found to be just 6 and the CCA projection therefore reduces the images to 6 components. It should be noted that even with just these 6 components, the SVM gives very good classification. The standard PCA reduced raw images did not give good classification. However, with Gabor pre-processed faces followed by standard PCA reduction gave much better results. Interestingly, Gabor pre-processing does not help the non-linear CCA method.

The classification results for the Neutral and Angry face images shown in Table 2, indicates the overall classification accuracy is not as good as with the happy versus neutral dataset. Classifying angry faces is a difficult task for computation models and can be seen from these results. Nevertheless, the SVM performs well with 79.54% accuracy with raw faces. There is not much difference in the classification accuracy with raw faces reduced in dimensionality with PCA and CCA.

(a)

(b)

Figure 7: Examples of the most often misclassified set of faces. (a) Top row shows happy faces wrongly classified as neutral. Bottom row shows neutral faces wrongly classified as happy. (b) Top row shows angry faces wrongly classified as neutral. Bottom row shows neutral faces wrongly classified as angry.

The results with both sets of data suggest that the raw face images give the best classification results. From some of the examples of misclassifications shown in Figure 7, it is not clear which feature has caused misclassification. Hence, we are currently undertaking further experiments with human subjects. We are attempting to see if there are any associations between the computational model and human performance.

4 CONCLUSIONS

Identifying facial expressions is a challenging and interesting task. Our experiment shows that identification from raw images can be performed very well with happy faces and angry faces. However, with a larger data set, it may be computationally intractable to use the raw images. It is therefore important to reduce the dimensionality of the data. The dimensionality reduction methods do fairly well. A linear method such as PCA does not appear to be sufficiently tunable to identify features that are relevant for facial expression characterization. However, performing

Gabor pre-processing on the images increases the classification accuracy of the data after performing PCA in the case of happy and neutral face images. This, however, does not apply to images that are subjected to dimensionality reduction with CCA. Gabor pre-processed PCA data with just 23 components is capable of performing well in comparison to the raw images reduced with PCA. The Gabor pre-processed CCA images, however, with just 5 components does not yield such comparable results. With the second model, classifying angry with neutral faces, the raw faces manage to give just 35 out of 44 faces correct (79.54%) and indicates the difficulty of classifying angry faces. Though the results of the classification for PCA and CCA processed raw images are comparable, it can be noted that Gabor pre-processing has managed to provide good classification with PCA reduced data and with CCA with just 23 and 6 components respectively. Future work will include extending the experiment to other four expressions and comparing the performance of the computational model with performance by human subjects.

REFERENCES

1. Ekman, P. and W.V. Friesen, *Constants across cultures in the face of the emotion.* Journal of Personality and Social Psychology, 1971. **17**.
2. Batty, B., M.J. Taylor, and *Early processing of the six basic facial emotional expressions.* Cognitive Brain Research, 2003. **17**.
3. Demartines, P. and d.J. Hérault, *Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets* IEEE Transactions on Neural Networks, 1997. **8**(1): p. 148-154.
4. Grassberger, P. and I. Proccacia, *Measuring the strangeness of strange attractors.* Physica D, 1983. **9**.
5. Jain, A.K. and F. Farrokhnia, *Unsupervised texture segmentation using Gabor filters.* Pattern Recognition, 1991. **24**(12).
6. Movellan, J.R., *Tutorial on Gabor Filters.* 2002.
7. Chang, C.-C. and Chih-Jen Lin *LIBSVM: a library for support vector machines.* 2001.
8. Zheng, D., Y. Zhao, and J. Wang, *Features Extraction using A Gabor Filter Family.* Proceedings of the sixth Lasted International conference, Signal and Image processing, Hawaii, 2004.
9. Daugman, J.G., *Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two dimensional visual cortical filters.* Journal of Optical.Society of .America .A, 1985. **2**(7).
10. Kulikowski, *Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex.* Biological Cybernetics 1982. **43**(3): p. 187-198.
11. Cortes, C. and V. Vapnik, *Support Vector Networks.* Machine Learning, 1995. **20**: p. 273-297.
12. Blanz, V., et al., *Comparison of view-based object recognition algorithms using realistic 3D models.* Proc. Int. Conf. on Artificial Neural Networks 1996: p. 251-256.

13. Yin, L., Wei, X., Sun, Y., Wang, J. & Rosato, M. J. , *A 3D Facial Expression Database For Facial Behavior Research.* 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 2006

REPRESENTATION AND CLASSIFICATION OF FACIAL EXPRESSION IN A MODULAR COMPUTATIONAL MODEL

ARUNA SHENOY[1]
TIM GALE [1, 2]
RAY FRANK[1]
NEIL DAVEY[1]

[1]*Department of Computer Science, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK*

[2]*Department of Psychiatry, QEII Hospital, Welwyn Garden City, AL7 4HQ, UK*

Recognizing expressions is a key part of human social interaction; processing of facial expression information is largely automatic in humans, but it is a non-trivial task for a computational system. The purpose of this work is to develop computational models capable of differentiating between a range of human facial expressions. Here we use two sets of images, namely: Angry and Neutral. Raw face images are examples of high dimensional data, so here we use some dimensionality reduction techniques: Principal Component Analysis and Curvilinear Component Analysis. We preprocess the images with a bank of Gabor filters, so that important features in the face images are identified. Subsequently the faces are classified using a Support Vector Machine. We also find the effect size of the pixels for the Angry and Neutral faces. We show that it is possible to differentiate faces with a neutral expression from those with an angry expression with high accuracy. Moreover we can achieve this with data that has been massively reduced in size: in the best case the original images are reduced to just 6 dimensions.

1. Introduction

According to Ekman and Friesen [1] there are six easily discernible facial expressions: anger, happiness(smile), fear, surprise, disgust and sadness. Moreover these are readily and consistently recognized across different cultures [2]. In the work reported here we show how a computational model can identify facial expressions from simple facial images. Specifically we investigate the differentiation of angry from neutral faces. In particular we show how angry faces and neutral faces can be differentiated.

Data presentation plays an important role in any type of recognition. High dimensional data is normally reduced to a manageable low dimensional data set. We perform dimensionality reduction using Principal Component Analysis (PCA) and Curvilinear Component Analysis (CCA). PCA is a linear projection

technique and it may be more appropriate to use a non linear Curvilinear Component Analysis (CCA) [3]. The Intrinsic Dimension (ID) [4], which is the true dimension of the data, is often much less than the original dimension of the data. To use this efficiently, the actual (Intrinsic) dimension of the data must be estimated. We use the Correlation Dimension to estimate the Intrinsic Dimension and is explained in later section. We compare the classification results of these methods with raw face images and of Gabor Pre-processed images [5, 6]. The features of the face (or any object for that matter) may be aligned at any angle. Using a suitable Gabor filter at the required orientation, certain features can be given high importance and other features less importance. Usually, a bank of such filters is used with different parameters and later the resultant image is a *L2 max* norm(at every pixel the maximum of feature vector obtained from the filter bank) superposition of the outputs from the filter bank.

2. Background

We basically perform an experiment to classify two expressions: neutral and Angry. We do pre-processing by Gabor filters and dimensionality reduction by techniques, namely, Principal Component Analysis and Curvilinear Component Analysis followed by a Support Vector Machine (SVM) [7] based classification technique and these are described below.

2.1. *Gabor Filters*

A Gabor filter can be applied to images to extract features aligned at particular orientations. Gabor filters possess the optimal localization properties in both spatial and frequency domains, and they have been successfully used in many applications [8]. A Gabor filter is a function obtained by modulating a sinusoidal with a Gaussian function. The useful parameters of a Gabor filter are orientation and frequency. The Gabor filter is thought to mimic the simple cells in the visual cortex. The various 2D receptive field profiles encountered in populations of simple cells in the visual cortex are well described by an optimal family of 2D filters [9]. In our case a Gabor filter bank is implemented on face images with 8 different orientations and 5 different frequencies.

Recent studies on modeling of visual cortical cells [10] suggest a tuned band pass filter bank structure. Formally, the Gabor filter is a Gaussian (with variances S_x and S_y along x and y-axes respectively) modulated by a complex sinusoid (with centre frequencies U and V along x and y-axes respectively) and is described by the Equation 1:-

$$g(x,y) = \frac{\exp\left[-\frac{1}{2}\left\{\left(\frac{x}{S_x}\right)^2 + \left(\frac{y}{S_y}\right)^2\right\} + 2\pi j(Ux + Vy)\right]}{2\pi S_x S_y} \quad (1)$$

The variance terms S_x and S_y dictates the spread of the band pass filter centered at the frequencies U and V in the frequency domain. This filter is complex in nature.

A Gabor filter can be described by the following parameters: The S_x and S_y of the Gaussian explain the shape of the base (circle or ellipse), frequency (f) of the sinusoid, orientation (Θ) of the applied sinusoid. Figure 1 shows examples of various Gabor filters. Figure 2b) shows the effect of applying a variety of Gabor filters shown in Figure 1 to the sample image shown in Figure 2 a). Note how the features at particular orientations are exaggerated.

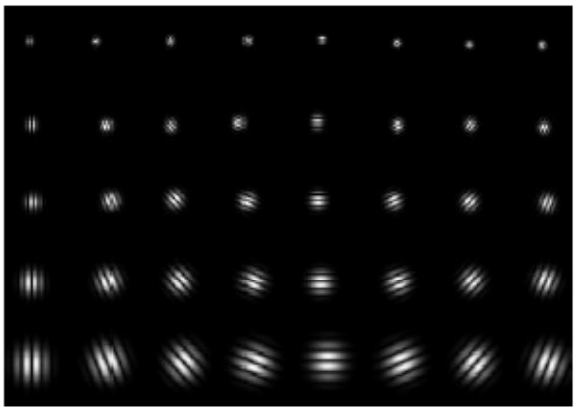
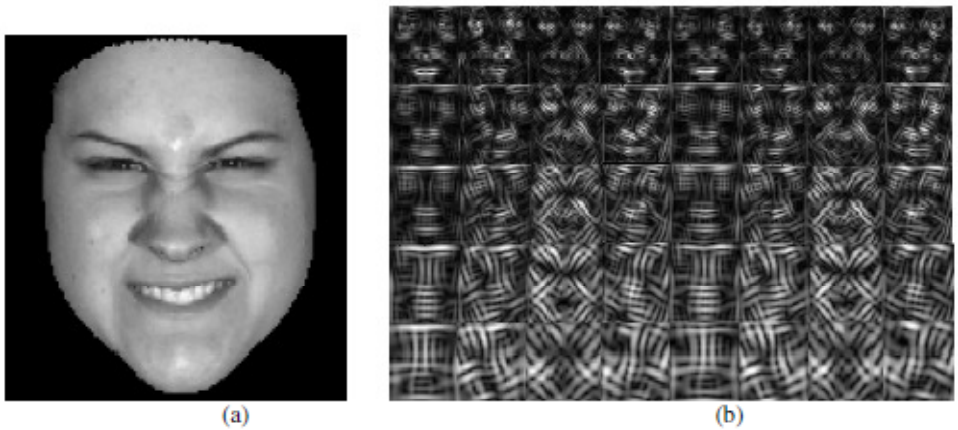


Figure 1: Gabor filters: Real part of the Gabor kernels at five scales and eight orientations

An augmented Gabor feature vector is created of a size far greater than the original data for the image. Every pixel is then represented by a vector of size 40 and demands dimensionality reduction before further processing. So a 63 × 63 image is transformed to size 63 × 63 × 5 × 8. Thus, the feature vector consists of all useful information extracted from different frequencies,

orientations and from all locations, and hence is very useful for expression recognition.



(a) (b)

Figure 2: a) Original face image. b) Forty Convolution outputs of Gabor: The rows correspond to decreasing frequency (from top to bottom) and columns represent various orientation.

Once the feature vector is obtained, it can be handled in various ways. We simply take the *L2 max* norm for each pixel in the feature vector. So that the final value of a pixel is the maximum value found by any of the filters for that pixel. The *L2 max* norm Superposition principle is used on the outputs of the filter bank and the Figure 3 b) shows the output for the original image of Figure 3 a).

(a) (b)

Figure 3 a): Original Image used for the Filter bank b) Superposition output (*L2 max* norm)

2.2. *Principal Component Analysis*

Principal Component Analysis (PCA) transforms higher dimensional datasets into lower dimensional uncorrelated outputs by capturing linear correlations among the data, and preserving as much information as possible in the data. PCA transforms data from the original coordinate system to the principal axes coordinate system such that the principal axis passes through the maximum possible variance in the data. The second principal axis passes through the next largest possible variance and this is orthogonal to the first axis. This is repeated for the next largest possible variances and so on. All these axes are orthogonal to each other. On performing this PCA on the high dimensional data, Eigenvalues or principal components are thus obtained [11]. The required dimensionality reduction is obtained by retaining only the first few principal components. Figure 4 shows the first two principal components.

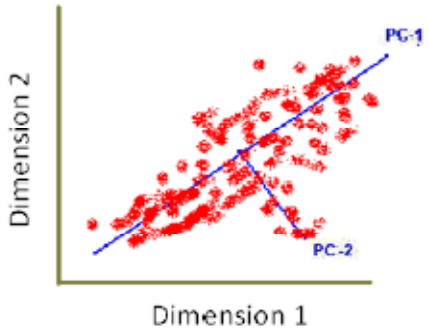


Figure 4: The first two consecutive principal components are shown.

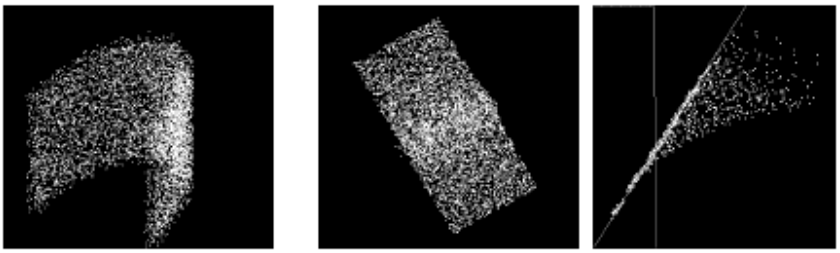
2.3. *Curvilinear Component Analysis*

Curvilinear Component Analysis (CCA) is a non-linear projection method that preserves distance relationships in both input and output spaces. CCA is a useful method for redundant and non linear data structure representation and can be used in dimensionality reduction. CCA is useful with highly non-linear data, where PCA or any other linear method fails to give suitable information [3]. The D –dimensional input X should be mapped onto the output P - dimensional space Y, where $P << D$. Their P - dimensional output vectors $\{ y_i \}$ should reflect the topology of the inputs $\{ x_i \}$. In order to do that, Euclidean distances between the x_i's are considered.

Corresponding distances in the output space y_i's is calculated such that the distance relationship between the data points is maintained. CCA puts more emphasis on maintaining the short distances than the longer ones. Formally, this reasoning leads to the following error function:

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(d_{i,j}^{X} - d_{i,j}^{Y}\right)^2 F_{\lambda}\left(d_{i,j}^{Y}\right) \qquad \forall\ j \neq i \qquad (2)$$

where $d_{i,j}^{x}$ and $d_{i,j}^{Y}$ are the Euclidean distances between the points i and j in the input space X and the projected output space Y respectively and N is the number of data points. $F\left(d_{i,j}^{Y}\right)$ is the neighbourhood function, a monotonically decreasing function of distance. In order to check that the relationship is maintained a plot of the distances in the input space and the output space ($dy - dx$ plot) is produced. For a well maintained topology, dy should be proportional to the value of dx at least for small values of $dy's$. Figure 5 shows CCA projections for the 3D data horse shoe data. The $dy - dx$ plot shown is good in the sense that the smaller distances are very well matched [3].



(a) (b) (c)

Figure 5: (a) 3D horse shoe dataset (b) 2D CCA projection (c) $dy - dx$ plot.

2.4. *Intrinsic Dimension*

One problem with CCA is deciding how many dimensions the projected space should occupy, and one way of obtaining this is to use the intrinsic dimension of the data manifold. The Intrinsic Dimension (ID) can be defined as the minimum number of free variables required to define data without any significant

information loss. Due to the possibility of correlations among the data, both linear and nonlinear, a D-dimensional dataset may actually lie on a P-dimensional manifold ($D \geq P$). The ID of such data is then said to be P. There are various methods of calculating the ID; here we use the correlation Dimension [8] to calculate the ID of face image dataset.

2.5. *Encoding Face*

'Effect Size' is a way of expressing the difference between two groups. Here two groups: Angry and Neutral are used. Cohen [12] defined d as the difference between the means, $M_1 - M_2$, divided by standard deviation, σ of either group.

$$d = \frac{M1 - M2}{\sigma} \tag{3}$$

M_1 and M_2 are the means of two groups and σ is the standard deviation and it is calculated by Equation 4.

$$\sigma = \sqrt{\frac{(\sigma 1^2 + \sigma 2^2)}{N}} \tag{4}$$

σ_1 and σ_2 are the standard deviation of the two classes, Angry and Neutral respectively and N is the total number of samples. 'Encoding face' is obtained by finding the Effect size of each pixel in an image. In other words it shows which pixels discriminate most between Angry and Neutral faces.

2.6. *Classification Using Support Vector Machines*

A number of classifiers can be used in the final stage for classification. We have concentrated on the Support Vector Machine. Support Vector Machine (SVM) is a set of related supervised learning methods used for classification and regression. SVM's are used extensively for many classification tasks such as: handwritten digit recognition [13] or Object Recognition [14]. A SVM implicitly transforms the data into a higher dimensional data space (determined by the kernel) which allows the classification to be accomplished more easily. We have used the LIBSVM tool [7] for SVM classification.

3. Experiments and Results

We experimented on 200 faces (112 female and 88 male) each with two classes, namely: *Neutral* and *Angry* (100 faces for each expression). The images are from the BINGHAMTON dataset [15] and some examples are shown in Figure 6. The training set had 160 faces (with 46 female, 34 male and equal numbers of them with neutral and angry expression). The original 128 $\times$ 128 image was reduced to 63 $\times$ 63. The test set consists of 40 faces (10 female, 10 male and equally balanced number of expression).

Figure 6: Examples of BINGHAMTON images used in our experiments was converted to gray scale and then reduced to size 63 $\times$ 63 for all experiments.

For PCA reduction we use the first few principal components which account for 95% of the total variance of the data, and project the data onto these principal components. This resulted in using 105 components of the raw dataset and 22 components in the Gabor pre-processed dataset. As CCA is a highly non-linear dimensionality reduction technique, we use the intrinsic dimensionality technique and reduce the components to its Intrinsic Dimension. The Intrinsic Dimension of the raw faces was approximated as 10 and that of Gabor pre- processed images was 6. The SVM classification results are shown in Table 1. Figure 7 shows the Eigen faces obtained by performing the PCA on the data set. Figure 8 shows the *dy-dx* plot of the CCA projection for the data set.

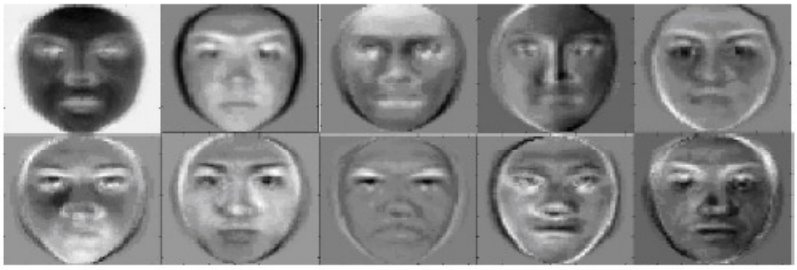


Figure 7: First ten Eigen faces of the dataset with two classes namely, Neutral and Angry.

The SVM was trained in the following way:

1) Transform the data to a format required for using the SVM software package – LIBSVM -2.86 [7].
2) Perform simple scaling on the data so that all the features or attributes are in the range [−1, +1].
3) Choose a kernel. We used the RBF kernel, $k(x, y) = e^{-\gamma|x-y|^2}$.
4) Perform fivefold cross validation with the specified kernel to find the best values of the cost parameter C and γ.
5) Using the best value of C and γ, train the model and finally evaluate the trained classifier using the test sets.

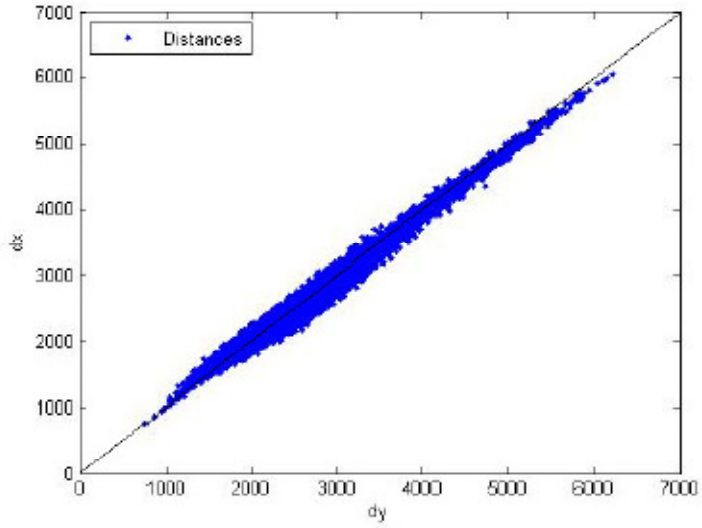


Figure 8: The ***dy-dx*** plot of the CCA projection for the raw data set. If there is a good matching between input and output spaces and the data is linear, then all the distances would be on the line ***dy = dx*** line. Here the original 3969 dimensions have been reduced to just 10 components by CCA.

Table 1. SVM Classification accuracy of raw faces and Gabor pre-processed images with PCA and CCA dimensionality reduction techniques.

% SVM Accuracy	Testset (40 images)
Raw faces	37 (92.5%)
Raw with PCA105	27 (67.5%)
Raw with CCA10	31 (77.5%)
Gabor pre-processed faces	29 (72.5%)
Gabor with PCA22	30 (75%)
Gabor with CCA6	28(70%)

The Encoding Angry face, the image where pixels which discriminate most between Angry and Neutral faces, is shown in Figure 9. The eyebrows are pulled together and down to form vertical wrinkles between the eyebrows in the forehead which is diagnostic of angry faces and can be seen clearly in the image. The glaring stare which is caused by the tightening of the muscles around the

eyelids can also be somewhat seen [16]. The flaring of the nostrils and the clenching of jaws [17] may also be an important indicator, though to a lesser extent.

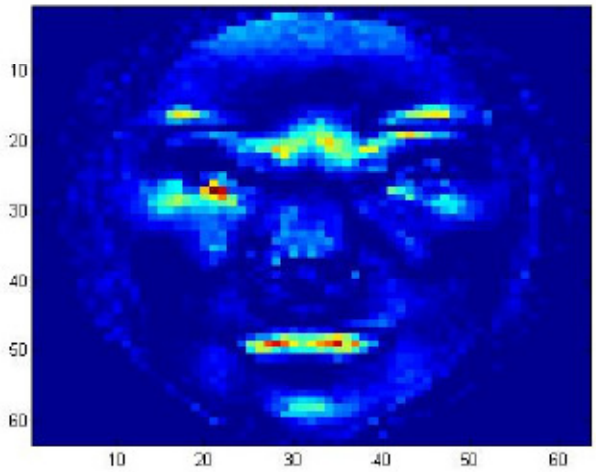


Figure 9: Encoding face: for Angry and Neutral

4. Conclusions

Identifying facial expressions is a challenging and interesting task. Our experiment shows that identification from raw images can be performed very well. However, with a larger data set, it may be computationally intractable to use the raw images. It is therefore important to reduce the dimensionality of the data. The experiments so far have shown that Gabor pre-processed images, with dimensionality reduced by CCA to just 6 components, offer a promising approach for investigation. In order to examine the consistency of the different models, further experiments need to be run with larger datasets and with other expression categories. The Similarities and differences in these results may be useful and informative in developing a better computational model and may contribute to our understanding of human processing of face expressions. Also, the performance of the computational model will have to be compared with the performance accuracy of human subjects with respect to a range of expressions.

References

1. Ekman, P. and W.V. Friesen, *Constants across cultures in the face of the emotion.* Journal of Personality and Social Psychology, 1971. **17**.
2. Batty, B., M.J. Taylor, and *Early processing of the six basic facial emotional expressions.* Cognitive Brain Research, 2003. **17**.
3. Demartines, **P**. and d.J. Hérault, *Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets* IEEE Transactions on Neural Networks, 1997. **8**(1): p. 148-154.
4. Grassberger, P. and I. Proccacia, *Measuring the strangeness of strange attractors.* Physica D, 1983. **9**.
5. Jain, A.K. and F. Farrokhnia, *Unsupervised texture segmentation using Gabor filters.* Pattern Recognition, 1991. **24**(12).
6. Movellan, J.R., *Tutorial on Gabor Filters.* 2002.
7. Chang, C.-C. and Chih-Jen Lin *LIBSVM: a library for support vector machines.* 2001.
8. Zheng, D., Y. Zhao, and J. Wang, *Features Extraction using A Gabor Filter Family.* Proceedings of the sixth Lasted International conference, Signal and Image processing, Hawaii, 2004.
9. Daugman, J.G., *Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two dimensional visual cortical filters.* Journal of Optical.Society of .America .A, 1985. **2**(7).
10. Kulikowski, *Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex.* Biological Cybernetics 1982. **43**(3): p. 187-198.
11. Smith, L.I., *Tutorial on Principal Component Analysis.* 2002.
12. Cohen, J., ed. *Statistical power analysis for the behavioural sciences.* 1988, Lawrence Earlbaum Associates.: Hillsdale, New Jersey, .
13. Cortes, C. and V. Vapnik, *Support Vector Networks.* Machine Learning, 1995. **20**: p. 273-297.
14. Blanz, V., et al., *Comparison of view-based object recognition algorithms using realistic 3D models.* Proc. Int. Conf. on Artificial Neural Networks 1996: p. 251-256.
15. Yin, L., Wei, X., Sun, Y., Wang, J. & Rosato, M. J. , *A 3D Facial Expression Database For Facial Behavior Research.* 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 2006
16. Hager, J.C. *Data Face.* 2003 [cited; Available from: http://www.face-and-emotion.com/dataface/expression/interpretations.html.
17. Novaco, R.W., *Anger.* Encyclopedia of Psychology. 2000: Oxford University Press.

A computational model of facial expressions: Classification and representation

A. Shenoy [1](a.1.shenoy@herts.ac.uk), N. Davey[1], R.J. Frank[1], T.M. Gale[1, 2]

1. Department of Computer Science, University of Hertfordshire, Hatfield, AL10 9AB, UK.
2. Department of Psychiatry, QEII Hospital, Welwyn Garden City, AL7 4HQ, UK.

According to Ekman and Friesan (1971), there are six discernible facial expressions (anger, happiness, fear, surprise, disgust, and sadness) that are expressed in a consistent way by all humans. These expressions are recognized across different cultures (Batty and Taylor, 2003) suggesting an innate, rather than learned, tendency. Although recognition of facial expressions seems to be an almost trivial task for human observers, it is difficult to describe the precise criteria and category boundaries for facial expressions. Not surprisingly then, recognition of facial expressions is a non-trivial task for computational models.

In this paper we will describe the development of a multi-module computational model of facial expression analysis. The model initially carries out feature detection, then dimensionality reduction and, finally, classification of facial expressions

In the initial stage, we use Gabor Filters (Movellan, 2002). A Gabor filter is a linear filter comprising a sinusoid function multiplied by a Gaussian function. The features of a face (or any object for that matter) can be aligned at any angle, and by using a suitable Gabor filter at the required orientation, certain features can be given importance or enhanced. Usually, a bank of such filters is used, with different parameters, and the resultant image is a convolution of the outputs from the filter bank. Simple cells in the mammalian visual cortex can be compared with 2D Gabor filters, to understand their receptive field profiles and the relationship between selectivity for orientation and frequency (Daugman, 1985).

High dimensional data, such as face images, includes a great deal of redundant information and may be more economically represented in a smaller number of dimensions. The Intrinsic Dimension (ID), which is the true dimension of the data, is often much smaller than the original dimensionality. In our model we experiment with different dimensionality reduction methods including Linear Discriminant Analysis (LDA) and Fisher LineaR Discriminant Analysis in particular, Principal Component Analysis (PCA) and Curvilinear Component Analysis (CCA). We examine the impact of these methods on the ability of our model to represent different facial expression categories.

Finally, we use a Support Vector Machine (SVM) to classify the reduced-dimensionality representations. We will present data on the model's classification performance as parameters within the different modules are systematically varied. Initially we will use two of the facial expression categories but our aim is to expand the dataset such that the model will be testable with all six expression categories proposed by Ekman and Friesen.

References:

Ekman, P., & Friesen, W. V. (1971) Constants across cultures in the face of the emotion, Journal of Personality and Social Psychology, 17.

Batty, B., & Taylor, M.J(2003) Early processing of the six basic facial emotional expressions, Cognitive Brain Research, 17.

Movellan, J.R.(2002). Tutorial on Gabor Filters. http://mplab.ucsd.edu/tutorials/pdfs/gabor.pdf

Daugman, J.G, "Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two dimensional visual cortical filters", J.Opt.Soc.Am.A,Vol 2, No.7.July 1985.

A Comparison of the performance of humans and computational models in the classification of facial expression

A.Shenoy[1] (a.1.shenoy@herts.ac.uk)
Sue Anthony[2] (s.1.anthony@herts.ac.uk),
R.J.Frank[1] (r.j.frank@herts.ac.uk),
Neil Davey [1] (n.davey@herts.ac.uk)
[1] Department of Computer Science, [2] Department of Psychology,
University of Hertfordshire, College Lane, Hatfield
AL10 9AB

Keywords: Facial Expressions, Image Analysis, Classification, Reaction Time.

Abstract

Recognizing expressions are a key part of human social interaction, and processing of facial expression information is largely automatic for humans, but it is a non-trivial task for a computational system. In the first part of the experiment, we develop computational models capable of differentiating between two human facial expressions. We perform pre-processing by Gabor filters and dimensionality reduction using the methods: Principal Component Analysis, and Curvilinear Component Analysis. Subsequently the faces are classified using a Support Vector Machines. We also asked human subjects to classify these images and then we compared the performance of the humans and the computational models. The main result is that for the Gabor pre-processed model, the probability that an individual face was classified in the given class by the computational model is inversely proportional to the reaction time for the human subjects.

Introduction

In this work we compare the performance of human subjects classifying facial expressions, with the performance of a variety of computational models. We use a set of 176 face images, half of which express anger and the other half have a neutral expression. The images are from the BINGHAMTON BU-3DFE database (Yin, Wei et al. 2006) and some examples are shown in Figure 1.

Pre-Processing Methods and Classification

This section describes how the computational model classifies angry faces and neutral faces. High dimensional data such as face images are often reduced to a more manageable low dimensional data set. We perform dimensionality reduction using both Principal Component Analysis (PCA) and Curvilinear Component Analysis (CCA). PCA is a linear projection technique but it may be more appropriate to use a non linear Curvilinear Component Analysis (CCA) (Demartines and Hérault 1997). Gabor filters are also often used for extracting features of images, and they are thought to mimic some aspects of human visual processing (Daugman 1985). Classification is performed using a Support Vector Machines (SVM). An SVM performs classification by finding the maximum margin hyper-plane in a feature space. The relative distance of an instance from this hyper-plane can be interpreted as its probability of belonging to the appropriate class. We have used the LIBSVM-2.86 tool (Chang and Lin 2001).

Experiment

Two sets of experiments were performed. Part A - Computational models. Part B - Classification performed by human subjects.

Part A- Computational Models

The data was divided into four subsets, and training/testing took place with a leave one out strategy, so that results are averages over four independent runs. Once a training set had been selected the two parameters of the SVM were optimized by cross-validation. Six variations of data processing are tested as detailed in Table 1.

(a) (b)

Figure 1: Example face images. a) Angry b) Neutral

Computational Model Results

For PCA, the first 97 components of the raw dataset and 22 components in the Gabor pre-processed dataset account for 95% of the total variance. For CCA, we reduce the data to its Intrinsic Dimension. The intrinsic dimension of the raw faces was approximated as 5 and that of the Gabor pre-processed images was 6.

The results in Table 2 indicate the overall classification accuracy is not very good; however, classifying angry faces is a difficult task for computation models (Susskind 2007) and can be seen from the results. Nevertheless, the SVM performs well with an average of 84.09% accuracy with raw face images

Table 1: Types of Computational Models

Name model	Type of Input	Dimensionality Reduction
Model 1	Raw faces	None
Model 2	Raw faces	PCA
Model 3	Raw faces	CCA
Model 4	Gabor pre processed	None
Model 5	Gabor pre-processed	PCA
Model 6	Gabor pre-processed	CCA

Table 2: SVM classification Results

Accuracy	TEST SET 4	TEST SET 3	TEST SET 2	TEST SET 1	Average
Model 1	79.54% (35/44)	93.18% (41/44)	79.54% (35/44)	84.09% (37/44)	84.09%
Model 2 (PCA97)	68.18% (30/44)	77.27% (34/44)	70.45% (31/44)	65.91% (29/44)	70.45%
Model 3 (CCA5)	68.18% (30/44)	59.09% (26/44)	63.64% (28/44)	63.64% (28/44)	63.64%
Model 4	68.18% (30/44)	79.55% (35/44)	72.73% (32/44)	81.82% (36/44)	75.57%
Model 5 (PCA22)	61.36% (27/44)	79.55% (35/44)	75% (33/44)	72.73% (32/44)	72.16%
Model 6 (CCA6)	63.64% (28/44)	70.45% (31/44)	68.18% (30/44)	63.64% (28/44)	66.48%

Part B - Human subjects

The 184 raw images were used in this experiment. Twenty individuals took part in the study.

Method

A total of 16 images were used in the pre-view block and the remaining 168 images were divided into 6 balanced blocks of 28 images each. We used a tool called as TESTBED (Taylor 2003) which is a response test generator program to record the classification and the Response Time (RT) of individuals.

Human Subject Results

Humans correctly classified the target expression with a mean of 82.86% (SD = 0.174) and the average RT was 1.132 seconds (SD = 0.714). The average RT ranges between a maximum value of 1.792sec and a minimum value of 0.714sec.

Discussion

We use the Bi-Variate Correlation to find any correlation between the average RT for human subjects and the class membership probability for the computational models. The results are considered to be significant at the level of 0.05, or below. The results of comparison are shown in correlation matrix of Table 3.

Table 3: The Bi-Variate Correlation Results

Model	Correlation value	Significance value
Model 1	-0.005	0.391
Model 2	+0.002	0.645
Model 3	0.022	0.126
Model 4	-0.045	0.016
Model 5	-0.028	0.065
Model 6	-0.003	0.597

Interestingly all but one of the correlations are negative, but only for Model 4 (Gabor filtered images with no dimensionality reduction) was this correlation significant, with the probability of the null hypothesis being 0.016. The correlation is negative with value -0.045. This negative correlation indicates large average RT (which presumably indicates that the subjects found it hard to classify), correlates with smaller class membership probability for the model. The results are interesting and encouraging (suggestive of Gabor filtering is similar to human face processing) and our next step is extending these experiments to other expressions.

References

Chang, C.-C. and C.-J. Lin (2001). "LIBSVM: a library for support vector machines."

Daugman, J. G. (1985). "Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two dimensional visual cortical filters." *Journal of Optical Society of America* A 2(7).

Demartines, P. and D. J. Hérault (1997). "Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets " *IEEE Transactions on Neural Networks* 8(1): 148-154.

Susskind, J. M., G. Littlewort, M.S. Bartlett, J. Movellan,A.K. Anderson((2007). "Human and computer recognition of facial expressions of emotion." *Neuropsychologia* 45(1).

Taylor, N. (2003). Developing with Authorware- Test bed. *ATSiP Conference at the University of Hertfordshire*

Yin, L., X. Wei, et al. (2006). A 3D Facial Expression Database For Facial Behavior Research. *7th International Conference on Automatic Face and Gesture Recognition* (FGR06)