Towards Real Time Facial Expression Recognition

N. K. Surendran, R. V. Raghavan, S. Q. Xie and K. C. Aw
Department of Mechanical Engineering
The University of Auckland
Auckland, New Zealand
nsur007@ec.auckland.ac.nz, rrag004@ec.auckland.ac.nz

Abstract

Facial expression recognition is a key aspect in the synthesis of adaptive human machine interfaces. If advanced expression recognition techniques are developed, machines can tailor their response to the feelings of their users. In this paper, we describe an algorithm which integrates Optical flow analysis and Support Vector Machines (SVM) to classify eight facial expressions with accuracies up to 98.73%. Colour images fed into the system are pre-processed to accentuate the subject's eyes, eyebrows and mouth. An Optical flow analysis on two successive frames from a video sequence enables us to identify the dominant feature on the face, for a given expression. Further, feature extraction is carried out on the selected feature and the data from this is input into the trained SVM classifiers. The results obtained from the tests carried out, suggest that the system is robust in dealing with subjects of a variety of races and both genders. In the near future, the proposed technique will be incorporated in a real time system.

1 Introduction

The human face is like a canvas that is constantly repainted to express a myriad of emotions. "A picture paints a thousand words"- the suggestive actions of humans in terms of facial expressions and gestures are capable of revealing far more in an instant than what is verbally possible. Mehrabian points out that 55% of the effect of communication is obtained from facial expressions, 38% from voice intonation and merely 7% from the spoken words [Mehrabian & Ferris.,1967]. The advent of an era boasting humanoids and automated systems aimed at aiding humans, has consequently needed the development of schemes to perceive human emotion. As a result Facial Expression Recognition has been a brimming area of research over the past few years. A good facial expression recogniser can significantly enhance human-computer interaction by allowing the system to adapt its behaviour to the user's emotions.

A detailed study of the existing techniques and approaches taken towards classifying human expression, has aided us in the development of an algorithm capable of classifying expressions of fear, anger, sorrow, digust, surprise, happiness (smile with the mouth closed and open) and neutral. It is the movement of the muscles associated with the facial features such as the mouth and eyes that enable us to express our emotions. Every individual smiles and frowns in a different way, some people are more expressive with their mouth and others with their eyebrows depending on the emotion they express. Our research presents an approach whereby the detection of the most dominant feature for a given expression is used to classify the emotion. It has been proven by [Heisele *et al*., 2001] that component based approach, i.e. classifying using just the features yields in a higher accuracy and a simpler recognition task than using the entire face as an input feature.

1.2 Literature Review

The three basic steps in recognising facial expressions from static images are face detection, feature extraction and facial expression classification [Pantic & Rothkrantz., 2000]. A common method of detecting faces as shown in [Essa & Pentland., 1995] is to apply spatial and temporal filtering to detect motion blobs from a sequence of images. These motion blobs are further evaluated using Principal Component Analysis (PCA) for detecting faces. Another widely used technique is to search for pixels in the image which match the human skin's hue and saturation value.

Following the detection of the face, a variety of feature extraction and associated classification procedures have been developed. Optical flow analysis has been used along with other spatial analysis techniques such as PCA and Gabor wavelet decomposition for extraction of features, as seen in [Donato *et al.,* 1999]. It was shown that at best an accuracy of 95.5% was achieved using Gabor wavelet representation and a worst performance of 53.1% using optical flow [Donato *et al.,* 1999]. Features extracted using static geometrical measurements from landmarks like the contours of the eye region and mouth, attain an accuracy of 91% [Pantic & Rothkrantz, 1999].

Research pioneered by [Ekman & Friesen.,1978] to classify 6 facial expressions requires the division of the face into 44 discrete sections known as Action Units (AUs) [Lien *et al*.,1998]. However, AUs still pose ambiguities because they are purely local spatial patterns and real facial motion is almost never completely localised so detecting a unique set of AUs for a specific facial expression is not gauranteed [Essa & Pentland., 1995]. This method of classification is called Facial Action Coding System (FACS) and has been used in conjunction with a single layer perceptron neural network to yield an accuracy of up to 93.3% [Tian *et al.,* 2001]. Along with neural networks, Support Vector Machines (SVM), a supervised learning algorithm has been used recently for face detection as well as facial expresssion recognition.

Real time facial expression recognition systems have been developed using SVM. One such system developed by [Michel & Kaliouby., 2003] requires the user to manually mark points on the face in the first frame of every video sequence. The geometrical displacements between these points are extracted in real-time and input to the SVM to classify the expressions. This technique achieves an average accuracy of 87.9% with Cohn-Kanade database. The advantage of this method is that there is no need of pre-processing involved. However, the system is not fully automated since specific points have to be marked manually for the classification to work. Our proposed technique of classifying the most dominant feature is not only autonomous but also yields higher accuracy. The processing time is also reduced because the extracted features reduce the size of input to SVM.

2 Proposed Technique

The proposed algorithm described in this paper, utilises an Optical flow analysis to detect areas of maximum change in the face, thus indicating the dominant feature for that expression. This information is used to segment out the required features from the face and perform feature extraction, which is then input into the SVM for classification. Both Optical Flow and SVM have been used successfully for real time applications in face detection and facial expression recognition. The integration of Optical flow analysis and SVM is an optimised approach, and to the best of our knowledge, has not yet been researched for real time facial expression recognition systems. The system overview shown in Figure 1 outlines the processing techniques implemented for facial expression recogntion.

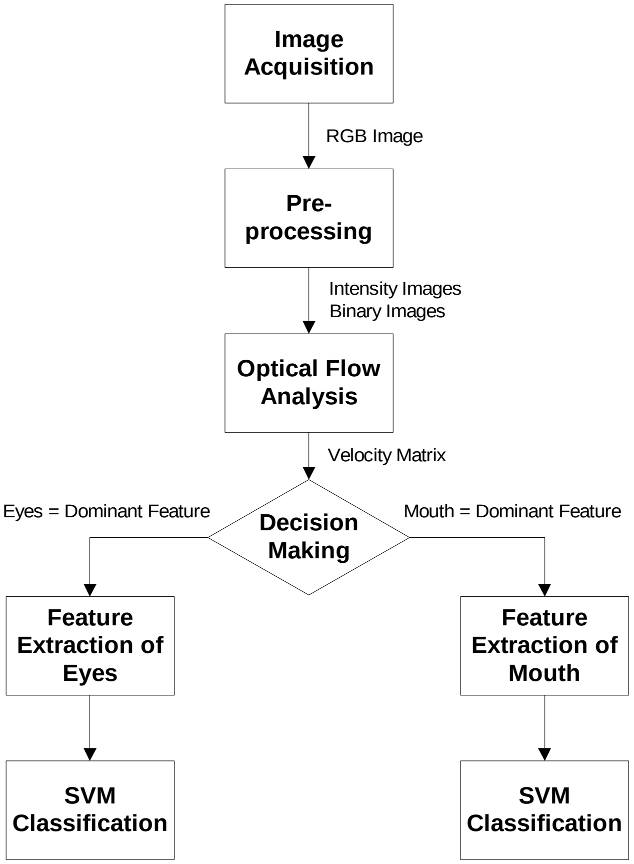

Figure 1: System overview

Two frames are captured from a video sequence and then subsequently compared using Optical flow analysis. Comparing the velocity vectors of the mouth and eye quivver plots, we are able to decide on which expression has moved the most. If the eyebrow region is the dominant feature, then the feature extraction process involves transforming the image into a density matrix. In the case of the mouth being the dominant feature, a vector is obtained from the binary image of the mouth to represents the number of white pixels in each column. This vector defines the width and the height of the mouth accross the image. These extracted features are then input to the trained SVM models for classifying into expressions.

3 Pre-processing

To facilitate our expression classification technique, the eyebrows, eyes and mouth need to be accentuated. The first step in this process is to locate and normalise the face contained within the image. Several algorithms exist for detecting a human face in an image sequence. The advantage of using a colour based face detection scheme is that it is faster than facial feature based techniques [Singh *et al*., 2003]. Also colour based skin detection is less invariant to the orientation and pose of the head.

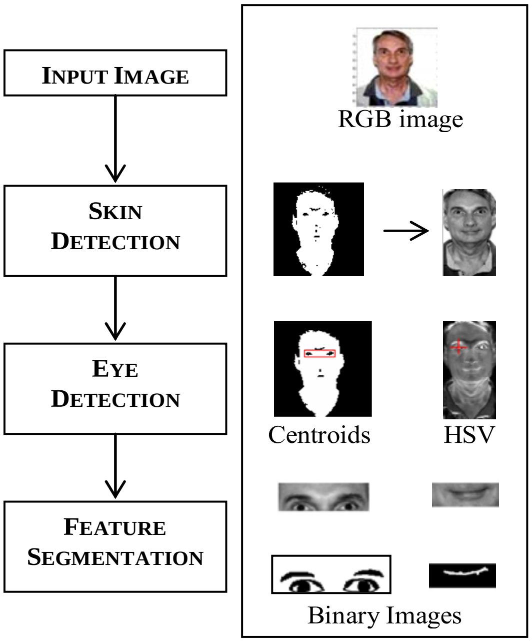

Figure 2: Stages of Preprocessing

The colour images acquired are transformed into the YCbCr spectrum for analysis. Using finely tuned Cb and Cr parameters, the presence of regions with pixel values similar to that of human skin are segmented. In comparison to a Hue Saturation Value spectrum, the YCbCr spectrum provides a better search space for human skin pixels as the effect of luminance on the image can be decoupled from itscolour properties hence making skin detection more reliable [Garcia & Tziritas., 1999]. The Cb and Cr parameters chosen have been determined in order to obtain optimum skin detection results after testing with 600 subjects of varying skin tones and complexions. Skin detection allows for a bounding box to be placed around the face in the image, thus allowing us to extract the face from the image.

Following this, the location of the subject’s right eye is determined by using the image’s saturation values, as shown in the Hue Saturation Value (HSV) image in Figure 2. An area of low intensity represents the subject’s white portion of the eye or the sclera and an adjoining bright patch identifies the pupil, hence allowing us to narrow down on the exact location of the pupil. A secondary measure taken to validate the accuracy of the eye position is a more statistical one. It is based on anthropometric data about the Inter-Pupillary Distance (IPD) [Dodgeson., 2004]. The image is thresholded and the centroids of probable eye regions are compared. If a centroid pair fits the criteria of being within IPD (45 to 80mm) apart, it is confirmed to be the centroid of an eye.

Some other eye detection techniques that exist assume that the iris of the eye is the darkest region in the image, as shown by [Fan *et al*., 2005]. This is often not the case if the person's eyes are blue in colour and in this case the eyebrows or eye lashes are detected as the darkest region in the image. Canny edge detection of the eyeballs and the distance between the eyeballs are used by [Huang & Mariani., 2000] to determine the position of eyes. Both of these techniques were tested and it was found that HSV based detection along with anthropometric data was more reliable when detecting eyes of people with varying iris colours and shapes.

Using the location of the eyes as referenece, a grey image of size 60 x 180 pixels is extracted from the subject's face in a manner that horizontally centralises the position of the eybrows and eyes within the image and positions the pupils 10 pixels above the bottom of the image. A histogram equalisation operation on this image allows for the regions of the eyebrows and eyes to be darkened in comparison to the forehead and upper cheeks, i.e. contrasting the image. Thresholding this image results in a binary image which represents the shape and position of the eyes and eyebrows alone.

A similar grey image of size 60 x 145 pixels is cut ensuring that the mouth is centralised within the image. Anthropometric data has been used to extract this image. This image is then subjected to Contrast Limited Adaptive Histogram Equalisation (CLAHE) to enhance the contrast between the mouth and subject's skin. This adaptive technique with a tile size of 15 x 15 pixels and a cliplimit parameter of 0.005, has proven to be effective in the pre-processing images with a wide range of skin tones and complexion. CLAHE operation is done instead of a normal histogram equalisation because the colour of the lips vary greatly across different people and therefore a simple histogram equalisation does not prove to be reliable. To detect the contours of the lips, the CLAHE image is filtered using the Laplacian of a Gaussian filter. The resulting intensity image revealing the contours of the subject's mouth is then thresholded and filled to provide a binary image as shown in Figure 2.

4 Motion Detection and Feature Selection

The two sets of cropped greyscale images obtained from the pre-processing phase are analysed with Optical Flow analysis. Comparing the two mouth and eye images from successive frames will allow us to determine if either of the features has been subjected to change. An advantage of using Optical Flow Analysis for this purpose is that the direction data provided by the motion vectors can further be used for motion tracking of individual features.

The Lucas-Kanade method is a simple yet powerful algorithm which has been proven to be more efficient in terms of computational time in comparison to the Horn Schunck, Uras and Anandan methods making it a better candidate for real-time applications [Liu *et al.,*1998]. It has also been observed to be adept at handling different types of motion sequences [Liu *et al*. 1998]. The reasons stated above made it the most suitable candidate for our particular application and has hence been incorporated in our system. The basic assumption with which this algorithm functions is that the pixel intensities remain unchanged along the motion trajectory. Hence, it deduces motion vectors based on the theory that any change in a pixel's colour value (brightness value for a grey scale image) is due to the movement of the respective pixel from one point to another [Tagliasacchi., 2006]. This assumption leads to the equation (1) which describes the constant brightness constraint under which the Lucas Kanade algorithm functions.

$$\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} \bullet \begin{bmatrix} u \\ v \end{bmatrix} = -\begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix} \tag{1}$$

In equation 1, I_x, I_y and I_t represent the intensity gradients in the horizontal vertical and temporal planes whilst u and v represent the velocity of a pixel in the x and y direction respectively. Each of these gradient values can be found by calculating the first order partial derivatives of the images under analysis. Subsequently using the equation above, u and v, the velocity vectors for each pixel in the x and y direction are worked out.

To do this two successive greyscale images of either the mouth or eye region are downsampled and then smoothed with a Gaussian filter. The Gaussian filter helps blur the images and in the process corrects small areas of pixels which are observed to be moving in a different direction to their neighbours. Thus, reducing the noise in optical flow. Two 6 x 6 Gaussian kernels, which are transposes of each other are used to convolute the two images and calculate their partial derivatives, I_x and $I_{y.}$ The temporal derivative, I_t is calculated by subtracting the two smoothed input images. The calculations for u and v for a pixel, are performed on successive pixels, taking into account each of the 8 around the one under analysis. The u and v matrix obtained, reveal the velocity of each pixel's motion. Analysing the resultant velocity matrices for both the eyes and the mouth enable in selecting the dominant feature to be used in the classification process.

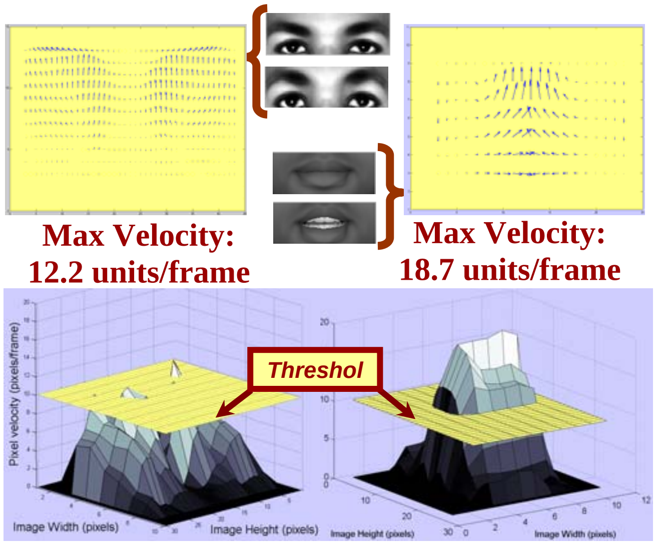

Figure 3: Optical flow analysis of eyebrow and mouth regions

Figure 3, graphically represents the optical flow results for the eyes and mouth when the subject's expression changes from neutral to surprise.

Testing with a variety of subjects and their facial expressions led to the conclusion that the motion of a particular feature is significant to recognise the expression only if it's optical flow velocity vectors are greater than 10 pixels per frame. This velocity threshold of 10 pixels per frame, represented by the yellow plane in Figure 3, is the minimum velocity needed for a feature, either the eyebrows or the mouth to show a recognisable change. The two graphs represent the magnitudes of the Optical Flow velocity vectors against the spatial coordinates of the image. As shown, only the velocity magnitudes of the mouth are observed to be well past this threshold signifying that it is the more expressive feature.

If no significant feature motion is detected in either of the two features, it is concluded that there is no need for classification as the subject's expression has not changed within the two frames. In the case that both the features are detected to have velocities above the specified threshold, classification is based on the feature which shows greater signs of motion.

5 Feature Extraction

In order to optimise the classification process with respect to processing time and classification accuracy, the relevant features that represent the expression will have to be extracted from either the mouth or eyes region. Features can be extracted relatively easily with our approach since the mouth and eyes regions have already been segmented out in the pre-processing stage. Since the SVM classifiers are trained using only features extracted data, the extraction of accurate and descriptive data from the eyebrows and mouth is paramount. Minimising the size of the extracted data is vital as it ensure quicker classification by the SVM and reduced training time.

5.1 Eyebrow Region Feature Extraction

If the eyebrowsa are selected as the dominant feature, the binary eyebrow image of size 60 x 180 pixels is further reduced to a 30 x 90 pixel image using density based feature extraction. Therefore, the input dimensionality has reduced by a factor of four. To reduce the dimensionality whilst maintining detail, the number of black pixels in every successive 2 x 2 grid of the 60 x 180 pixel image is calculated and stored to produce a 30 x 90 element data matrix. Therefore, each individual value in the density matrix stores information about four pixels in the binary image. In Figure 4, it can be seen that the binary image is assumed to be a 4 x 9 pixel image for demonstrating the density extraction technique, this image is converted to a 2 x 4 density matrix. The values in the density image represents the number of pixels in the 2 x 2 grid (coloured as yellow and green) that have a black pixel. The right hand side image in Figure 4 is the actual extracted density image from the binary image for the person's eyes. As you can see that the density image still resembles the eyes region thus the information is now represented in a more compressed form. The 30 x 90 matrix is converted to a row vector of size 2700 and then scaled to be between 0 and 1 by dividing each element in the matrix by 4. Scaling avoids attributes in greater numeric ranges from dominating the ones in smaller ranges while classifying using the SVM [Hsu et al.,2003].

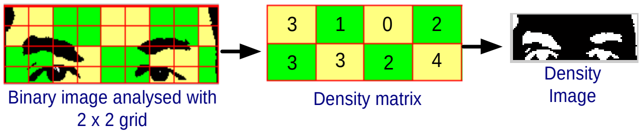

Figure 4: Converting binary image to density image for feature extraction of eyes

5.2 Mouth Region Feature Extraction

A bounding box is placed around the filled binary image of the mouth and the sum of the white pixels in each column is calculated. This sum for each column is stored in a feature vector of length 125 and is centralised for mouths with different sizes. Figure 5 shows disgust and smile expressions for the same person to illustrate how the feature extraction works. It can be seen from the figure that the disgust mouth is smaller in length so zeros are appended to centralise the mouth within the vector. The mouth feature vector contains information about the width of the mouth and its height across the mouth. The mouth feature vector is scaled so that the values are between 0 and 1.

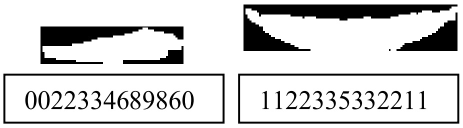

Figure 5: Disgust feature vector (left) and smile feature vector (right)

6 Feature Classification - SVM

Supervised machine learning algorithms such as SVM belong to a class of maximum margin classifiers that are based on the results of statistical learning theory introduced by Vapnik in 1979 [Vapnik., 1995]. They have been proven to be better in their generalisation capability when compared to artificial neural networks [Burges., 1998] and have been used recently for real-time facial expression recognition systems. Hybrid approaches to combine SVM with Gabor wavelet decomposition [Qin & He., 2005], Independent Component Analysis (ICA) [Qi et al., 2002] and FACS [Schulze., 2002] have yielded in higher recognition accuracy than using it as a standalone classifier. Although the training of the SVM classifiers is time consuming, it has been chosen among other learning methods because the classification process is instantaneous, thus making it ideal for real time applications. Unlike Artifical Neural Networks, SVMs do not assume the data to be gaussian in nature. SVM can be implemented efficiently because they use sequential minimal optimisation to train support vectors [Platt., 1998]. An integrated package called LibSVM has been used for the implementation of SVM classification in our application [Chang & Lin., 2001].

6.1 Basic Theory of SVM

For a given set of points belonging to two classes, an SVM tries to find a decision function for an optimal separating hyperplane (OSH) which maximises the margin between the two sets of data points. The solutions to such optimisation problems are derived by training the SVM with sets of data similar to what it may encounter during its application.

Consider a binary classification task with a training set of points xi ϵ Rn, i = 1,2,…, N where each point belongs to a corresponding class label $y_i \in \{-1,1\}$ and let the decision function be f(x) = sign(**w**.**x** + *b*), where w denotes the weights and *b* the bias of the decision function. Therefore a point lying directly on the hyperplane satisfies the condition, **w**.**x** + *b* = 0 and the points lying on the right and left of the hyperplane must satisfy the following conditions:

$$x_i \cdot w + b > 0 \qquad \forall i \tag{2}$$

$$x_i \cdot w + b < 0 \qquad \forall i \tag{3}$$

$$y_i (x_i \cdot w + b \geq 0) \qquad \forall i \tag{4}$$

The above equations (2) and (3) can be implicitly formulated as in (4) and hold true for all input points if the classification is correct. Figure 6 shows the optimal and arbitrary hyperplanes for the same input training set of points. Assuming the data is linearly separable, there is only one optimal hyperplane that exists which maximises the distance between the support vectors. For a maximum margin, the distance from the hyperplane to the support vectors on either side must be equidistant. Let us denote H- as the hyperplane which satisfies x_i***w** + **b** = -1 and H+ as the hyperplane which satisfies x_i***w** + **b** = +1. Hence the maximum margin between hyperplanes becomes 2/||**w**||. Thus, the hyperplane that optimally separates the data is the one that minimises $||\mathbf{w}||^2$= 0.5***wTw** = 0.5*($\mathbf{w}_1^2$ + $\mathbf{w}_1^2$ +…. +$\mathbf{w}_1^2$), subject to constraints in (4). (Multiplication of wTw by 0.5 is for numerical convenience and does not change the solution).

This is a classic quadratic optimisation problem with inequality constraints and is solved by the saddle point of the Lagrange functional (Lagrangian),

$$L_p \equiv \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^{l} \alpha_i \tag{5}$$

where the α_i are Lagrange multipliers. The optimal saddle point ($\mathbf{w}_o$, b_o, α_o) must be found by minimising L_p with respect to **w** and *b* and has to be maximised with respect to non negative α_i. Taking the derivatives of L_p with respect to b and w and re-substituting it back in to (5) gives:

$$L_D = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \tag{6}$$

$$\alpha_i \geq 0 \quad i = 1 ... l \quad \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{7}$$

Each of the training points is associated with a Lagrange multiplier α_i and training points with $\alpha_i > 0$ are the support vectors which lie closest to the hyperplane and hence influence the shape of the decision boundary. Training is achieved by maximising L_D with respect to the constraints in (7) to obtain the classification for a new data point **x**,

$$f(x) = sign\left(\overline{w} \cdot x + \overline{b}\right) \tag{8}$$

where the weights and bias are,

$$\overline{w} = \sum_{i=1}^{l} \overline{\alpha}_i y_i x_i \qquad b = -\frac{1}{2} w \cdot [x_r + x_s] \tag{9}$$

where x_r and x_s are support vectors satisfying

$$\alpha_r, \alpha_s > 0, \quad y_r = 1, \quad y_s = -1 \tag{10}$$

Until now, it has been assumed that the input data is linearly separable, however, real world data - in our case the extracted facial features are non linear in nature. To separate non linear data, SVMs map the input data nonlinearly into a higher dimensional space called feature space, so that the data can be linearly separated by an optimal hyperplane. Therefore, each point x in the input space is mapped to a point z = Φ(x), and the mapping is subjected to the dot product of two points in the feature space, expressed as a kernel function K (x_i, x_j) = Φ (x_i) .Φ (x_j). We use the Radial Basis Function (RBF) kernel because it works well with nonlinearities and is less numerically difficult, for further explanation – refer to [Hsu *et al.,* 2003]. The RBF kernel is expressed as:

$$K(x_i, x_j) = e^{\left(-\chi \|x_i - x_j\|^2, \chi > 0\right)} \tag{11}$$

The presence of noise in most real world data sets mean that data points cannot be separated by an optimal hyperplane, therefore leading to poor generalisation. To overcome this problem, the slack variable, ζ is introduced for each training sample, and the new optimisation problem thus becomes maximisation of the margin between two sets and the minimisation of misclassifications caused by points lying in the range of ζ ≥ 0. The error of misclassifications is weighted with a penalty parameter C, therefore a high C value assigned to the classification errors leads to better separation and a low C value results in a soft margin – causes the separation fuzzy. Therefore the optimisation problem now becomes:

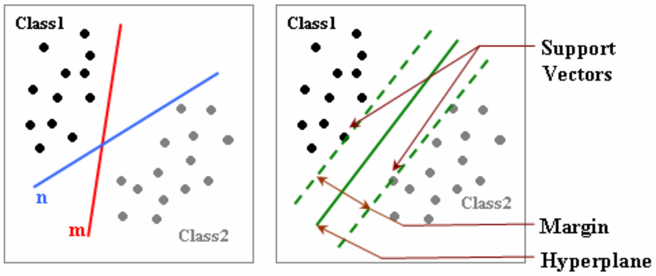

Figure 6: Classification using arbitrary hyperplanes (left) and using optimal hyperplane (right) with largest margin identified, passing the two support vectors.

Minimise w, b, ζ: $$\tfrac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \tag{12}$$

subject to $$y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \tag{13}$$

6.2 SVM Models for classification

The task of classifying eight expressions is a multi-class classification problem unlike the simple two class recognition problem described in the previous section. There are two approaches for the construction of multi-class SVMs – the first is a 'one against all' approach to classify one class against all the remaining ones; the second is a 'one against one' approach to classify between each pair of classes. For the latter approach, if there are k classes then it is necessary to train k*(k-1)/2 classifiers and each classifier is trained with data from two classes.

Our technique, of classifying expressions using only dominant feature requires us to have 16 classifiers i.e. each expression has a mouth and a eye classifier, this is with the 'one against all approach'. For example, if the mouth is detected as the dominant feature, then its feature vector will be passed to the 8 mouth classifiers, one for each expression. Although, we will be training each expression to be classified with both mouth and eyes, the classification of the expression will only be determined with the dominant feature. If a 'one against one' approach had been chosen, it would require 56 separeate SVM classifiers for the eyes and the mouth features (8 classes). The training of the SVM classifiers is a time consuming process since it involves determining the optimum C and γ parameters for each individual SVM classifier and then subsequently training them all with large volumes of data. Therefore the 'one against all' design has been incorporated to classify the facial expressions.

Currently, the system is able to successfully classify expressions from static images using either the eight SVMs for the mouth or the eight SVMs for the eyebrows. Since we are working with static images, the optical flow is calculated by comparing two successive images to determine the feature with the highest amount of change. In order to develop a training set, a total of 2400 images, (300 subjects protraying each of the 8 expression) have been used. The training of the classifiers was done using a 5 fold cross validation wherein we first divide the training set into 5 subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining four subsets. The cross validation procedure prevents the overfitting problem, therefore obtaining good generalisation i.e. the learning algorithm does not memorise the patterns.

In order to yield the best accuracy for classification, we have determined the optimum penalty parameter, C and kernel parameter, γ for each of the 16 classifiers. The calculation of the C and γ parameters were based on the training data of 2400 images alone.

7 Results and Discussion

The trained SVM classifiers were tested with 80 unseen faces each with eight expressions hence resulting in 640 images. The test faces have been chosen so as to cover the spectrum of skin tones and complexion. Table 1 shows the results of testing using the eyebrows as the dominant feature and Table 2 with the mouth. The tables also show the number of support vectors from the training stage, i.e. images classified on the margin of the hyperplane. The results obtained are in accord with our proposed approach of facial expression recognition. For example, the eyebrows are the dominant feature when a person is sad, thus resulting in a higher classification accuracy than using the mouth feature.

It can be seen clearly that the accuracies obtained using the mouth region are in general higher than the ones obtained from the eyebrow region. An average accuracy of 95.5 % and 92.8 % is achieved for classifying 640 unknown expressions using mouth and eyebrow classifier respectively. It is inferred from this observation that some expressions are classified better with mouth than with eyebrows. For example, when a person is smiling there is hardly any movement in the eyebrows and the expression is completely conveyed through the mouth movement. Hence the results in Table 1 show 87.66 % classification accuracy with the eyebrows in comparison to the 97.94 % (refer to Table 2) with the mouth. It is hard to classify the expression using the eyebrows between a smile (with a closed mouth and open mouth) and neutral expression because of the similarity in the shape of the eyes and eyebrows for these expressions. This is validated by the large high number of support vectors obtained during the training stage for these three classifiers.

Table 1: Results of classification using eyebrow feature vector

Expression	**Classification Accuracy**	**Support Vectors** (max 2400)
Anger	96.20%	455
Disgust	96.94%	355
Fear	96.20%	397
Neutral	87.50%	1920
Smile (closed mouth)	87.66%	1890
Smile (open mouth)	87.66%	1910
Surprise	94.62%	334
Sad	96.20%	424

Table 2: Results of classification using mouth feature vector

Expression	**Classification Accuracy**	**Support Vectors** (max 2400)
Anger	97.63%	455
Disgust	98.73%	291
Fear	97.63%	555
Neutral	91.14%	1152
Smile (closed mouth)	90.03%	953
Smile (open mouth)	97.94%	258
Surprise	98.73%	349
Sad	92.10%	590

However, a large number of support vectors is also obtained for the smile with a closed mouth and neutral expression, using the mouth as the dominant feature, as shown in Table 2. This is because the features extracted from these two expressions are not unique using the current method which involves calculating the height vector. It can be seen from Figure 7 that the height vectors are the same for the two expressions. There is no information about the distribution of the white pixels within the black background i.e. the gradient of the mouth is not taken into account.

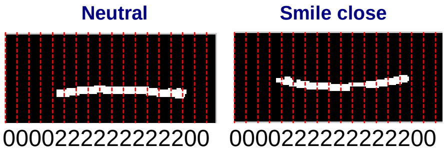

Figure 7: Neutral and smile close expressions having similar height vectors

8 Future Work

The proposed technique requires a full frontal view of the subject's face and is also sensitive to the angular rotation of the subject's head. These conditions need to be taken into account for classifying expressions because along with the movement of facial muscles, the head is also subjected to small movements. The currently devoloped algorithms will need to be integrated in real time using frames from video sequences instead of static images. The high accuracies achieved through the classification phase using static images are a testament to efficient pre-processing and feature extraction tecniques.

To attain even higher classification accuracies, alternate feature extraction techiniques, which calculate the distance of the eyebrows from the eyes, will be investigated. The advantage of having optical flow integrated into this approach is that its results can be used for two purposes. Firstly, it has been used to identify the dominant feature for a given expression, thus reducing the need for verifying the expression with all 16 SVM classifiers (8 for the mouth and 8 for the eyes). Secondly, the algorithm can be integrated to function with a pan and tilt enabled camera. Pixel velocities along the edges of the subject's face can be used to direct the motion of the camera so as to maintain the subject's face in the centre of the frame. This will allow for facial tracking along with facial expression recognition, thus making the system more robust when implemented.

Most vision groups around the world, use LibSVM package for classifying facial expression using SVM. As part of the future work, a comprehensive testing procedure will be developed to compare the LibSVM package and Kecman's Iterative Single Data Algorithm (ISDA) for classifying facial expresions [Kecman *et al.,* 2005]. The search for optimum C and γ parameters for training the SVM classifiers is a time consuming process and it is imperative to obtain the optimised parameters for higher classification accuracy. Therefore, alternate techniques will be investigated to optimise this search using ISDA.

9 Conclusions

The proposed approach towards facial expression recognition has been evaluated to achieve accuracies of up to 98.73%. The pre-processing technique adopted has been tailored to cater to both genders and a variety of races. The nature of human facial expression is such that they are not set in stone i.e. there are different degrees associated for the same expression. For example, a person being slightly angry could express using his/her eyebrows but a person expressing intense anger could use their mouth to show their teeth as well. Therefore, the integrated approach presented in this paper is able to classify irrespective of the degrees of expression because the system is only concerned with identifying the dominant feature.

To reduce processing time, Lukas Kanade optical flow analysis has been incorporated. This allows us to identify features that show signs of motion, hence allowing us to skip the classification phase if the subject's expression has not changed. The extracted features could be the mouth region or the eyes region depending on the optical flow analysis. The decision making process allows the dominant feature to be used for feature extraction. The extracted features are classified using 16 SVM classifiers (8 for the mouth region and 8 for the eyes region).

Currently the proposed approach has been tested with static images to classify expressions. The results obtained from this approach to facial expression recognition are promising enough to pursue its implementation in real-time. Other real time systems using SVM have been implemented for facial expression recognition but they are still not fully automated and involve manual selection of distinctive geometric points on the face. The proposed approach does not require any manual input and also optimises the classification of expressions using just the dominant feature. Thus, reducing processing time and increasing the classification accuracy.

References

[Burges., 1998] Christopher J.C. Burges. A tutorial on support vector machines for pattern recognition. Knowledge Discovery and Data Mining, Vol. 2(2), 1998.

[Chang & Lin., 2001] LIBSVM : a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2001.

[Dodgeson., 2004]. Dodgeson A. Neil. Variation and Extrema of Human Interpupillary Distance. *Proceedings of Stereoscopic Displays and Virtual Reality Systems,* Vol. 5291, pp. 36-46, 2004.

[Donato *et al.,* 1999] Donato Gianluca, Bartlett S. Marian, Hager C. Joseph, Ekman Paul and Sejnowski J. Terrence, "Classifying Facial Actions", *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 21, No. 10, pp. 974-989, 1999.

[Ekman & Friesen.,1978] Ekman Paul and Friesen V. Wallace. The Facial Action Coding System. *Consulting Psychologists Press*, Inc. San Francisco, CA, 1978.

[Essa & Pentland., 1995] Essa A. Irfan and Pentland Alex. "Facial Expression Recognition using a Dynamic Model and Motion Energy". *IEEE 3rd International Conference on Face & Gesture Recognition*, April 14-16, 1998.

[Fan *et al*., 2005] Chao Fan, Hossein Sarrafzadeh and Farhad Dadgosta. Facial Expression Analysis by Support Vector Regression. Retrieved, August 13, 2006 from http://pixel.otago.ac.nz/ipapers/55.pdf#search=%22Chao%20fan%20SVM%22

[Garcia & Tziritas., 1999] Christopher Garcia. and Georgios Tziritas. Face Detection Using Quantized Skin Colour Regions Merging and Wavelet Packet *Analysis. IEEE Transaction on Multimedia*, Vol. 1(3), September 1999.

[Heisele *et al*., 2001] Heisele Bernd, Purdy Ho and Poggio Tomaso. Face Recognition with Support Vector Machines: Global versus Component-based Approach. *Proceedings of the Eighth IEEE International Conference on Computer Vision,*2001.

[Hsu et al.,2003] Hsu C-W, Chang C-C & Lin C-J. A practical guide to Support Vector Classification. Technical Report by the Department of Computer Science and Information Engineering, National Taiwan University, 2003.

[Huang & Mariani., 2000] Weimin Huang and Robert Mariani. Face Detection and Precise Eyes Location. *Proceedings of the International Conference on Pattern Recognition*, pp. 4722, Vol. 4, 2000.

[Kecman *et al.,* 2005] Kecman Vojislav, Huang T.-M. and Vogt Michael. Iterative Single Data Algorithm for Training Kernel Machines from Huge Data Sets: Theory and Performance, Chapter in 'Support Vector Machines: Theory and Applications, Ed. Wang, L., Series: Studies in Fuzziness and Soft Computing, Springer Verlag, Vol. 177, pp.255-274, 2005.

[Lien et al.,1998] Lien J.James, Kanade Takeo, Cohn, F.

Jeffrey and Li Ching-Chung. "Automated Facial Expression Recognition Based on FACS Action Units", *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 390-395, 1998.

[Liu *et al.,*1998] Liu Hongche, Hong Tsai-Hong, Herman Martin and Chellapa Rama. Accuracy vs Efficiency Trade-offs in Optical Flow Algorithms. *Computer Vision and Image Understanding Journal*, Vol. 72, No. 3, pp. 271-286, 1998.

[Mehrabian & Ferris., 1967] Mehrabian, A and S.R. Ferris. Inference of Attitudes From Nonverbal Communication in Two Channels. Journal of Counseling Psychology Vol. 31 pp. 248-52, 1967.

[Michel & Kaliouby., 2003] Michel Philipp and Kaliouby E. Rana. Real Time Facial Expression Recognition in Video using Support Vector Machines. *Proceedings of the Fifth International Conference on Multimodal Interfaces,* pp. 258-264, 2003.

[Pantic & Rothkrantz., 1999] Pantic Maja and Rothkrantz J. M. Leon. An Expert System for Multiple Emotional Classification of Facial Expressions. *Proceedings of the Eleventh International Conference on Tools with Artificial Intelligence ,* pp. 113-120, 1999.

[Pantic & Rothkrantz., 2000] Pantic Maja and Rothkrantz J. M. Leon. Automatic Analysis of Facial Expressions: the State of the Art. *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 22, No.12, pp. 1424-1445, 2000.

[Platt., 1998] Platt C. John. Sequential minimal optimisation: A fast algorithm for training support vector machines. Technical report 98-14, Microsoft Research, Redmond, Washington. Available at http://www.research.microsoft.com/~jplatt/smo.html, April 1998.

[Singh et al., 2003], Sanjay Kr. Singh, Chauhan, D.S., Mayank Vasta and Richa Singh . A Robust Skin Color Based Face Detection Algorithm. *Tamkang Journal of Science and Engineering*, Vol. 6(4), pp 227-234, 2003.

[Tagliasacchi., 2006] Tagliasacchi Marco. Optical Flow Estimation using Genetic Algorithms. *Lecture notes in Computer Science,* Vol. 2955, pp. 309-316, Jan. 2006.

[Tian *et al.,* 2001] Tian Ying-Li., Kanade Takeo and Cohn F. Jeffrey. Recognising Action Units for Facial Expression Analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 23, No. 2, pp. 97-115, 2001.

[Vapnik., 1995] Vapnik N. Vladimir. The Nature of Statistical Learning Theory. Springer-Verlag New York, Inc., New York, NY, USA, 1995.