

# Towards Portable Facial Expression Recognition by Machine Learning

*Siu-Yeung Cho, Teik-Toe Teoh and Yok-Yen Nguwi*  
*Centre for Computational Intelligence, School of Computer Engineering*  
*Nanyang Technological University, Nanyang Avenue, Singapore 639798*  
*E-mail: [assycho@ntu.edu.sg](mailto:assycho@ntu.edu.sg)*

## **Abstract**

Facial expression recognition is a challenging task. A facial expression is formed by contracting or relaxing different facial muscles on human face which results in temporally deformed facial features like wide open mouth, raising eyebrows or etc. The challenges of such system have to address with some issues. For instances, lighting condition is a very difficult problem to constraint and regulate. On the other hand, real-time processing is also a challenging problem since there are so many facial features to be extracted and processed and sometime conventional classifiers are not even effective to handle those features and then produce good classification performance. This chapter discusses the issues on how the advanced feature selection techniques together with good classifiers can play a vital important role of real-time facial expression recognition. Several feature selection methods and classifiers are discussed and their evaluations for real-time facial expression recognition are presented in this chapter. The content of this chapter is a way to open-up a discussion about building a real-time system to read and respond to the emotions of people from facial expressions.

## **1. Introduction**

Given the significant role of the face in our emotional and social lives, it is not surprising that the potential benefits from efforts to automate the analysis of facial signals, in particular rapid facial signals, are varied and numerous (Ekman et al., 1993), especially when it comes to

computer science and technologies brought to bear on these issues (Pantic, 2006). As far as natural interfaces between humans and computers are concerned, facial expressions provide a way to communicate basic information about needs and demands to the machine. In fact, automatic analysis of facial signals seem to have a natural place in various vision sub-systems, including automated tools for tracking gaze and focus of attention, lip reading, bimodal speech processing, face/visual speech synthesis, and face-based command issuing.

Facial Expression Analysis is a challenging task. A facial expression is formed by contracting or relaxing different facial muscles on human face which results in temporally deformed facial features like wide open mouth, raising eyebrows or etc. The challenges of such system have to address the following issues:

- a. Lighting conditions is a very difficult problem to constraint and regulate. The strength of the light depends on the light source (see Figure 1).
- b. The direction of the subjects face is not always ideal which may pose difficulties when the system is implemented live that captures moving subjects' facial expression (see Figure 2).
- c. Another difficulty is the way image is acquired by the image acquisition system. The characteristics of the image acquisition system can affect the quality of the images or videos captured.
- d. Occlusion of subject face may tumble the hit rate of many established approaches. The experiments being carried out by most researchers do not take occlusion into account (see Figure 3).

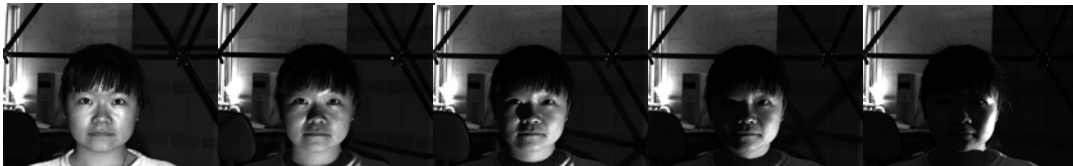


Figure 1: Light variations problem: face images are taken from different illumination conditions (source: Yale Face Database B <http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html>)

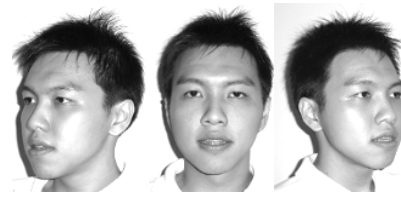


Figure 2: Pose variations problem: face images are taken from different postures of the object (source NTU Asian Emotion Database <http://www3.ntu.edu.sg/SCE/labs/forse/Asian%20Emotion%20Database.htm>)



Figure 3: Occlusion problems: facial components are occluded by some artifact objects. (source NTU Asian Emotion Database <http://www3.ntu.edu.sg/SCE/labs/forse/Asian%20Emotion%20Database.htm>)

Because of the above challenges, this chapter is going to introduce about recent advances in feature selection and classification methodologies for facial expression analysis. It first describes the background of different techniques used for facial expression analysis. Then it introduces the ideas of an automatic facial expression recognition system proposed by the authors which is included feature extraction, feature selection and classification methods. Finally, some of the future trends in terms of scientific and engineering challenges are discussed and recommendations for achieving a better facial expression technology are outlined.

## 2. Background

The first known facial expression analysis was presented by Darwin in 1872 (Darwin, 1872). He presented the universality of human face expressions and the continuity in man and animals. He pointed out that there are specific inborn emotions, which originated in serviceable associated habits. After about a century, Ekman and Friesen (1971) postulated six primary

emotions that possess each a distinctive content together with a unique facial expression. These prototypic emotional displays are also referred to as basic emotions in many of the later literature. They seem to be universal across human cultures and are namely happiness, sadness, fear, disgust, surprise and anger. They developed the Facial Action Coding System (FACS) for describing facial expressions. It is appearance-based. FACS uses 44 action units (AUs) for the description of facial actions with regard to their location as well as their intensity. Individual expressions may be modelled by single action units or action unit combinations. FACS codes expression from static pictures.

In the nineties, more works emerge started from Cottrell et al. (1990) who described the use of a Multi-Layer Perceptron Neural Networks for processing face images. They presented a number of face images to the network and train it to perform various tasks such as coding, identification, gender recognition, and expression recognition. During this procedure, face images are projected onto a subspace in the hidden layers of the network; it is interesting to note that this subspace is very similar to the eigenfaces space. However, an important difference is that, in this case, the face subspace is defined according to the application for which the system is to be used. Correct identification rates of up to 97 percent were reported when the system was tested using a database of images from 11 individuals. Mase et al. (1991) then used dense optical flow to estimate the activity of 12 of the 44 facial muscles. The motion seen on the skin surface at each muscle location was compared to a pre-determined axis of motion along which each muscle expands and contracts, allowing estimates as to the activity of each muscle to be made. Recognition rates of 86% were reported. Subsequently, Matsuno et al. (1994) presented a method of facial expressions recognition using two dimensional physical models named *Potential Net* without using feature extraction. *Potential Net* is a physical model which consists of nodes connected by springs in two dimensional grid configurations. Recognition is achieved by comparing nodal displacement vectors of a net deformed by an input image with facial expression

vectors. It recognized four kinds of facial expressions, happiness, anger, surprise and sad and the hit rate is about 90%.

After that, Yacoob and Davis (1996) continues the research in optical flow computation to identify the direction of rigid and nonrigid motions that are caused by human facial expressions. The approach is based on qualitative tracking of principal regions of the face and flow computation at high intensity gradient points. Three stages of the approach are: locating and tracking prominent facial features, using optical flow at these features to construct a mid-level representation that describes spatio-temporal actions, and applying rules for classification of mid-level representation of actions into one of the six universal facial expressions. In the mean time, Rosenblum et al. (1996) proposed a radial basis function network to learn the correlation of facial feature motion patterns and human expressions. The network was trained to recognize “smile” and “surprise” expressions. Success rate was about 83%~88%. The work explores the use of a connectionist learning architecture for identifying the motion patterns characteristic of facial expressions. Essa and Pentland (1997a) also described facial structure using optimal estimation optical flow method coupled with geometric, physical and motion-based dynamic models. It is an extension variance of FACS, called FACS+ which is a more accurate representation of human facial expressions. Another work by Lanitis’s group (Lanitis, 1997) used an active shape model to locate facial features and then used shape and texture information at these locations to classify expressions, obtaining a recognition rate of 74% for expression of the basic emotions and also neutral expressions. In the following year, Otsuka and Ohya (1998) proposed a method for spotting segments displaying facial expression from image sequences. The motion of the face is modeled by HMM in such a way that each state corresponds to the conditions of facial muscles, e.g., relaxed, contracting, apex and relaxing. Off-line and on-line experiments were carried out. Off-line experiment obtains the optimum threshold values and to evaluate relation of recognition rate and frame rate. On-line experiments were used to evaluate recognition performance for sequences of multiple instances of facial expressions. Experiments showed that the segments for

the six basic expressions can be spotted accurately near real time. In 1999, Chandrasiri et al. (1999) proposed a facial expression space which is made by the same person's peak facial expression images based on multidimensional scaling and a method for deriving trajectory of a facial expression image sequence on it. The main advantage of this method is that it focuses on both temporal and spatial changes of personal facial expression.

More variances of recognition approaches emerge after the year 2000, and some are tested with video sequences (Bourel, 2002; Pardas, 2002; Barlett et al, 2003). Shin et al. (2000) extract pleasure/displeasure and arousal dimension emotion features using hybrid approach. The hybrid approach used feature clustering and dynamic linking to extract sparse local features from edges on expression images. The expressions are happiness, surprise, sadness, disgust, fear, satisfaction, comfort, distress, tiredness and worry. They concluded that the arousal-sleep dimension may depend on the personal internal state more than pleasure-displeasure dimension, that is to say, the relative importance of dimension can have an effect on facial expression recognition on the two dimensional structure of emotion. Fasel and Lüttin (2000) described a system that adopts holistic approach that recognizes asymmetric FACS Action Unit activities and intensities without the use of markers. Facial expression extraction is achieved by difference images that are projected into a sub-space using either PCA or ICA, followed by nearest neighbor classification. Recognition rates are between 74~83%. Bourel (2002) investigated the representation of facial expression based on spatially-localised geometric facial model coupled to a state-based model of facial motion. The system consists of a feature point tracker, a geometric facial feature extractor, a state-based feature extractor, and a classifier. The feature extraction process uses 12 facial feature points. The spatio-temporal features are then created to form a geometric parameter. The state-based model then transforms the geometric parameter into 3 possible states: 'increase', 'stable', 'decrease'. The classifier makes use of k-nearest neighbour approach. Another work on video sequence recognition is by Pardas (2002) who described a system that recognizes emotion based on MPEG4 facial animation parameters. The system is

based on HMM. They defined a four-state HMM. Each state of each emotion models the observed FAPs using a probability function. Kim et al. (2003) then proposed a method to construct a personalized fuzzy neural networks classifier based on histogram-based feature selection. Recognition rate is reported to be in the range of 91.6%~98.0%. The system proposed in (Bartlett et al., 2003) detects frontal faces in the video stream and classifies them in seven classes in real time: neutral, anger, disgust, fear, joy, sadness, and surprise. An expression recognizer receives image regions produced by a face detector and then a Gabor representation of the facial image region is formed to be later processed by a bank of SVMs classifiers.

Ji (2005) based on FACS and developed a system that adopted a dynamic and probabilistic framework based on combining Dynamic Bayesian Networks (DBN) with FACS for modeling the dynamic and stochastic behaviors of spontaneous facial expressions. The three major components of the system are facial motion measurement, facial expression representation and facial expression recognition. Wu et al. (2005) modeled uncertainty in facial expressions space for facial expression recognition using fuzzy integral. The fuzzy measure is constructed in each facial expression space. They adopted Active Appearance Models (AAM) to extract facial key points and classify based on shape feature vector. Fuzzy C-means (FCM) was used to build a set of classifiers. The recognition rates were found to be 83.2% and 91.6% on JAFFE and FGnet databases respectively. Yeasin et al. (2005) compared the performances of linear and non-linear data projection techniques in classifying six universal facial expressions. The three data projection techniques are Principal Component Analysis (PCA), Non-negative Matrix Factorization (NMF) and Local Linear Embedding (LLE). The system developed by (Anderson and McOwan, 2006) characterized monochrome frontal views of facial expression with the ability to operate in cluttered and dynamic scenes, recognizing six emotions universally associated with unique facial expressions, namely happiness, sadness, disgust, surprise, fear, and anger. Faces are located using a spatial ratio template tracker algorithm. Optical flow of face is subsequently determined using a real-time implementation of gradient model. The expression recognition

system then averages facial velocity information. The motion signatures produced are then classified using Support Vector Machines. The best recognition rate is 81.82%. Zeng et al. (2006) classified emotional and non emotional facial expressions occurred in a realistic human conversation setting-Adult Attachment Interview (AAI). The AAI is a semistructured interview used to characterize individuals' current state of mind with respect to past parent-child experiences. Piecewise Bezier Volume Deformation (PBVD) was used to track face. They applied kernel whitening to map the data to a spherical symmetrical cluster. Then Support Vector Data Description (SVDD) was applied to directly fit a boundary with minimal volume around the target data. Experimental results suggested the system generalize better than using PCA and single Gaussian approaches. Xiang et al. (2007) utilized fourier transform, fuzzy C means to generate a spatio-temporal model for each expression type. Unknown input expressions are matched to the models using Hausdorff distance to compute dissimilarity values for classification. The recognition rate was found to be 88.8% with expression sequences.

In general, a facial expression recognizer comprises of 3 stages, namely feature extraction, features selection, and classification. Feature extraction involves general manipulation of the image. The raw image is processed to provide a region of interest (human face without hairs and background) for the second stage to select meaningful features. Some noise reduction, clustering, labelling or cropping may be done in this stage. For some ready data which are taken off the shelf from online database, the first stage is unnecessary.

Features selection is an important module. Without good features, the effort made in the classification stage would be in vain. Fasel and Luetin (2003) provides a detailed survey on facial expression, they classified the feature extraction methods into several groups. Some of the methods were highlighted like Gabor wavelets, Active Appearance Model, Dense Flow Fields, Motion and Deformable Models, Principal Component Analysis, High Gradient Components and etc. They group the facial features into 2 types: intransient facial features and transient facial



features. Intransient features are like eyes, eyebrow and mouth which are always present in the face. Transient features include wrinkles and bulges.

Neural network may be a popular choice for classification. Most of them fall under supervised learning. The other classification methods underscored by Pantic and Rothkrantz (2000) are Expert System Rules, Discriminant functions by Cohn et al. (1998), Spatio-temporal motion energy templates by Essa and Pentland (1997b), Thresholded motion parameters by (Black & Yacoob, 1997), and Adaptive Facial Emotion Tree Structures by Wong and Cho (2006). In the past two decades, research has focused on how to make face recognition systems fully automated by tackling problems, such as, localization of a face in a given image or video clip and extraction of features such as eyes, mouth, etc. Meanwhile, significant advances have been made in the design of feature extractions and classifiers for successful face recognition. Both the appearance-based holistic approaches and feature-based methods have the strength and weaknesses. Compared to holistic approaches, feature-based methods are less sensitive to variations in illumination and viewpoint and to inaccuracy in face localization. Several survey papers offer more insights on facial expression systems which can be found in (Andrea et al., 2007; Kevin et al., 2006; Zhao et al., 2003; Fasel & Luetin, 2003; Lalonde & Li, 1995)

### **3. Facial Expression Systems**

Most computer vision researchers think of motion when they consider the problem of facial expression recognition. An often cited study by Bassili (1978) showed that humans can recognize facial expressions above chance from motion, using point-light displays. In contrast to the Bassili's study in which humans were barely above chance using motion without texture, humans are nearly at ceiling for recognizing expressions from texture without motion (i.e. static photographs).

Appearance-based features include Gabor filters, integral image filters (also known as box-filters, and Haar-like filters), features based on edge-oriented histograms and those based on Active Appearance Models (Edwards et al., 1998). A common reservation about appearance-based features for expression recognition is that they are affected by lighting variation and individual difference. However, machine learning systems taking large sets of appearance-features as input, and trained on a large database of examples, are emerging as some of the most robust systems in computer vision. Machine learning combined with appearance-based features has been shown to be highly robust for tasks of face detection (Viola & Jones, 2004; Fasel et al., 2005), feature detection (Vukadionvic & Pantic, 2005; Fasel, 2006), and expression recognition (Littlewort et al., 2006). Such systems also do not suffer from issues of initialization and drift, which are major challenges for motion tracking.

Figure 4 shows an outline of the real-time expression recognition system developed by Bartlett and colleagues (Barlett et al., 2003; Littlewort et al., 2006). The system automatically detects frontal faces in the video stream and codes each frame with respect to 7 dimensions: neutral, anger, disgust, fear, joy, sadness, surprise. The system first performs automatic face and eye detection using the appearance-based method of Fasel et al. (2005). Faces are then aligned based on the automatically detected eye positions, and passed to a bank of appearance-based features. A feature selection stage extracts subsets of the features and passes them to an ensemble of classifiers which make a binary decision about each of the six basic emotions plus neutral. According to their results, they found that Gabor wavelets and ICA (Independent Component Analysis) gave better performance than PCA (Principal Component Analysis), LFA (Local Feature Analysis), Fisher's linear discriminants, and also outperformed motion flow field templates. More recent comparisons included comparisons of Gabor filters, integral image filters, and edge-oriented histogram (e.g., Whitehill & Omlin, 2006), using SVMs and AdaBoost as the classifiers. They found an interaction between feature-type and classifier, where AdaBoost performs better with integral image filters, while SVMs perform better with Gabors. The

difference may be attributable to the fact that the pool of integral image filters was much larger. AdaBoost performs feature selection and does well with redundancy, whereas SVMs were calculated on the full set of filters and do not do well with redundancy.

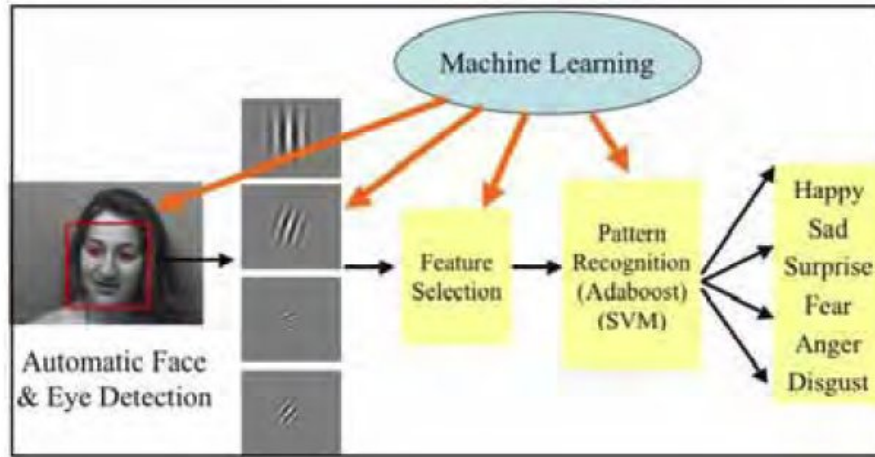


Figure 4: Outline of the real-time expression recognition system of Littlewort et al. (2006)

In this chapter, we will present our recent works in recognizing facial expressions. Our work attempts to use a boosting Naïve Bayes Classifier (NBC) to classify facial expressions. Both theoretical and practical studies have often been carried out to understand the predictive properties on this boosting NBC method. The Bayesian classifier is the most popular classifier among the probabilistic classifiers used in the machine learning community. Furthermore, Naive Bayesian classifier is perhaps one of the simplest yet surprisingly powerful techniques to construct predictive models from labeled training sets when comparing to other supervised machine learning methods. NBC can also provide a valuable insight to the training data by exposing the relations between attribute values and classes besides good predictive accuracy. The resulting of NBC are often robust to a degree where they match or even outperform other more complex machine learning methods despite NBC's assumption of conditional independence of attributes given in the class. The study of probabilistic classification is the study of approximating a joint distribution with a product distribution. Probabilistic classifiers operate on data sets where

each example  $x$  consists of feature values  $\langle a_1, a_2, \dots, a_i \rangle$  and the target function  $y$  can take on any value from a pre-defined finite set  $V = (v_1, v_2, \dots, v_j)$ . Bayesian rule is used to estimate the conditional probability of a class label  $y$ , and then assumptions are made on the model, to decompose this probability into a product of conditional probabilities. The formula used by the simple Bayesian classifier is:  $P(a_i | v_j)$  and  $P(v_i)$  which can be calculated based on their frequency in the training data, under the assumption that features values are conditionally independent given the target value. On the other hand, one can model the component marginal distributions in a wide variety of ways for numerical features. The zero counts are another problem with this formula. Zero counts are obtained when a given class and feature value never occur together in the training set, and is problematic because the resulting zero probabilities will wipe out the information in all the other probabilities when they are multiplied. Incorporating a small sample correction into all probabilities is one of the solutions to this problem. Its probability is set to  $1/N$ , where  $N$  is the number of examples in the training set if a feature value does not occur given some class,. The assumption of independence is clearly almost always wrong. However, a large scale comparison of simple Bayesian classifier with state-of-the-art algorithms for decision tree induction and instance-based learning on standard benchmark datasets found that simple Bayesian classifier sometimes is superior to each of the other learning schemes even on datasets with substantial feature dependencies. An explanation on why simple Bayesian method remains competitive, even though it provides very poor estimates of the true underlying probabilities. NBC algorithm always had the best accuracy per needed training time and it predicts the class feature in a very short time.

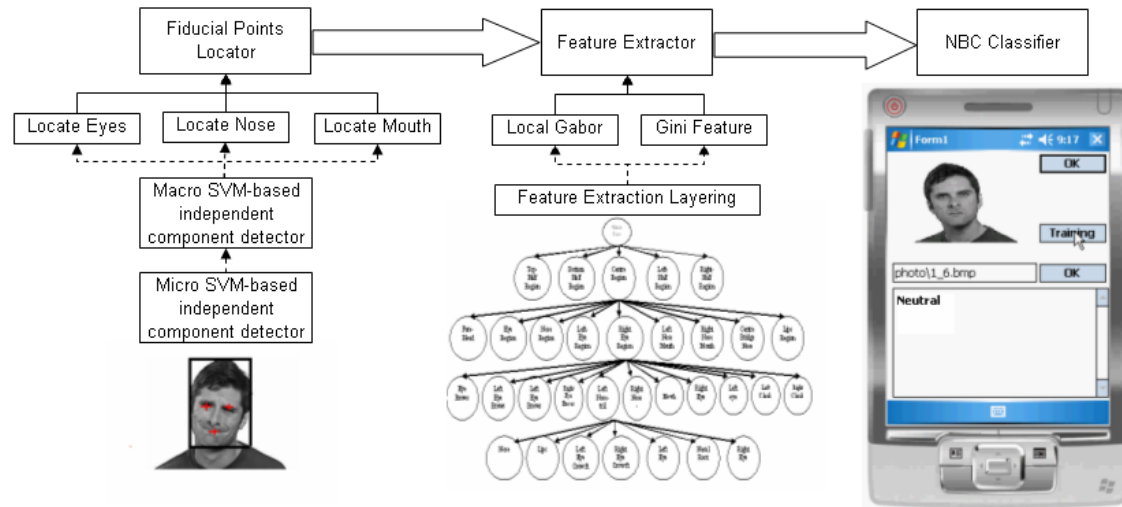


Figure 5: A system is proposed to recognize four types of facial expressions using Naïve Bayesian Boost

In this section, we outline the structure of our developed facial expression recognizer. The proposed model is going to recognize four types of facial expression, namely, neutral, joy, sad and surprise. A probabilistic based approach using the Naïve Bayesian Boost is adopted to recognize these four types of facial expressions. Figure 5 demonstrates the block diagram of the entire system. The system is composed of three major blocks, i.e., the feature locator, the feature extractor, and the classifier. The feature locator finds crucial fiducial points for subsequent feature extraction processing. We adopted Gabor features and Gini extractions which will be discussed in the following sub-sections. Finally the meaningful features are classified into the corresponding class.

#### 4. Feature Extraction

The first step in the system is face detection, i.e., identification of all regions in the scene that contain a human face. The problem of finding faces should be solved regardless of clutter, occlusions, and variations in head pose and lighting conditions. The presence of non-rigid movements due to facial expression and a high degree of variability in facial size, color and

texture make this problem even more difficult. Numerous techniques have been developed for face detection in still images (Yang et al., 2002; Li & Jian, 2005).

In our work, we used the most commonly face detector proposed by Viola and Jones (2004) to detect the face region from either a still image or a video stream. This detector consists of a cascade of classifiers trained by AdaBoost. Each classifier employs integral image filter, also called “box filters”, which are reminiscent of Haar Basis functions, and can be computed very fast at any location and scale. This is essential to the speed of the detector. For each stage in the cascade, a subset of features is chosen using a feature selection procedure based on AdaBoost. After the process of face detection, we need to extract the fiducial point to locate at the facial components such as eyes, nose and mouth. The component-based feature detector is used in this stage. It has two levels, namely, the micro SVM based independent component detector and the macro SVM based independent component detector. The micro level uses linear SVM based independent component detection. Each component classifier was trained on a set of extracted facial components (the 4 key fiducial components) and on a set of randomly selected non-face patterns. The macro level uses the maximum outputs of the component classifiers within rectangular search regions as inputs to a combination SVM classifier. The macro SVM performs the final detection of the face component regions.

After the presence of a face has been detected in the observed scene, the next step is to extract the information about the displayed facial signals. Most of the existing facial expression analyzers are directed toward 2D spatiotemporal facial feature extraction. The feature extractor then adopted Gabor wavelet feature extraction. Gabor wavelet is a popular choice because of its capability to approximate mammals’ visual cortex. The primary cortex of human brain interprets visual signals. It consists of neurons, which respond differently to different stimuli attributes. The receptive field of cortical cell consists of a central ON region is surrounded by 2 OFF regions, each region elongates along a preferred orientation (Daugman, 1985). According to

Jones and Palmer (1987), these receptive fields can be reproduced fairly well using Daugman's Gabor function.

The Gabor wavelet function can be represented by:

$$g(x, y) = g_1(x, y) \exp(j2\pi Wx) \quad (1)$$

where

$$g_1(x, y) = \left( \frac{1}{2\pi\sigma_x\sigma_y} \right) \exp\left( -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right) \quad (2)$$

We consider the receptive field (RF) of each cortical cell consists of a central ON region (a region excited by light) surrounded by two lateral OFF regions (excited by darkness) (La Cara et al., 2003). Spatial frequency ( $W$ ) determines the width of the ON and OFF regions.  $\sigma_x^2$  and  $\sigma_y^2$  are spatial variances which establish the dimension of the RF in the preferred and non-preferred orientations. As shown in Figure 6a, the Gabor wavelets are represented with different orientations and frequencies. These Gabor wavelets act as stimuli to the system. Figure 6b gives an overview of the optimal stimuli for the first 48 units resulting from one typical simulation. Since the processing at the retina and LGN (Lateral Geniculate Nucleus) cortical layers are used by Gaussian kernels applying to some spectral bands to simulate local inhibition, most stimuli resemble Gabor wavelets. The input image is in HSI domain and is projected into different Gabor wavelets to generate the output signals that resemble electrical signals in visual cortex. Different orientations and special frequencies produce different wavelets. After the convolution of input and wavelets, a set of feature vectors is formed that acts like the 'hypercolumns' as described by Hubel and Wiesel (1962).

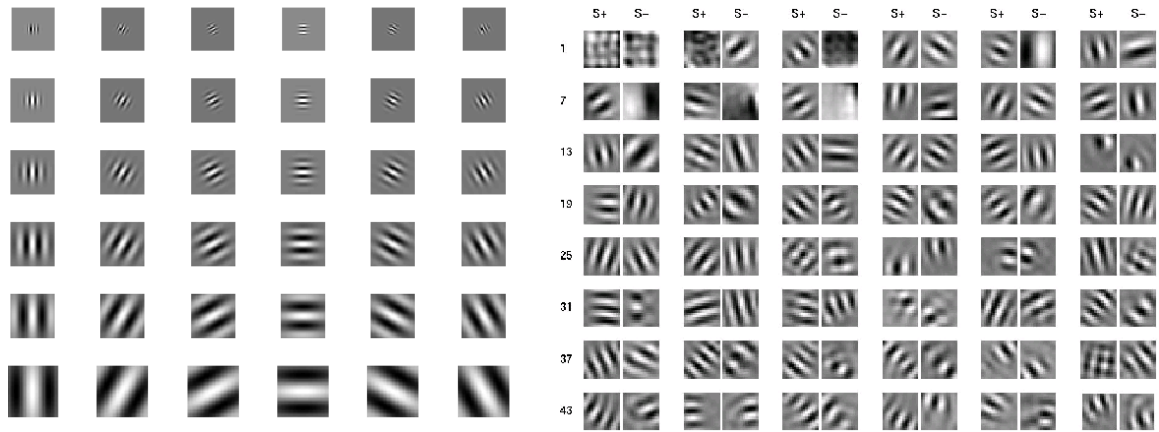


Figure 6: Gabor wavelets representation and the overview of optimal excitatory and inhibitory stimuli (S+ res. S-).

## 5. Feature Selection

Feature selection plays an important role of the automatic facial expression recognition system. The objectives of feature selection include noise reduction, regularization, relevance detection and reduction of computational effort. It is likely a learned image filters, such as, ICA, PCA and LFA which are based on unsupervised learning from the statistics of large image databases. There are two main models of feature selection, the filter model and the wrapper model. The filter model is generic, which is not optimized for any specific classifier. It is modularized and may sacrifice classification accuracy. The wrapper model is always tailored to a specific classifier and it may lead to better accuracy as a result. The strength of the wrapper model is that it differentiates irrelevance, strong and weak relevance and it improves performance significantly on some datasets. One weakness of the wrapper model is that calling the induction algorithm repeatedly may cause overfitting. Another weakness is that its computational cost is expensive. In our work, we used the filter approach for the feature selection in the facial expression recognition. Two methods will be discussed in this section.



### 5.1. *T-test Feature Selection*

In many existing feature selection algorithms, feature ranking is often used to show which input features are more important (Guyon and Elisseeff, 2003; Wang and Fu, 2005), especially when datasets are very large. T-test (Devore and Peck, 1997) is a common type of feature ranking measures which is often used to assess whether the means of two classes are statistically different from each other by calculating a ratio between the difference of two classes means and the variability of the two classes. The T-test has been used to rank features for microarray data (Jaeger et al., 2003; Su et al., 2003) and for mass spectrometry data (Wu et al., 2003; Levner 2005). These uses of T-test are limited to two class problem. For multi-class problems, a T-statistic value can be calculated as eqn. (3) for each feature of each class by evaluating the difference between the mean of one class and the mean of all the classes, where the difference is standardized by the within-class standard deviation.

$$t_{ic} = \frac{\bar{x}_{ic} - \bar{x}_i}{M_c \cdot (S_i + S_0)} \quad (3)$$

$$S_i^2 = \frac{1}{N - C} \sum_{c=1}^C \sum_{j \in c} (x_{ij} - \bar{x}_{ic})^2 \quad (4)$$

$$M_c = \sqrt{1/n_c + 1/N} \quad (5)$$

Here  $t_{ic}$  is the T-statistics value for the  $i$ -th feature of the  $c$ -th class;  $\bar{x}_{ic}$  is the mean of the  $i$ -th feature in the  $c$ -th class, and  $\bar{x}_i$  is the mean of the  $i$ -th feature for all classes;  $x_{ij}$  refers to the  $i$ -th feature of the  $j$ -th sample;  $N$  is the number of all the samples in the  $C$  classes and  $n_c$  is the number of samples in class  $c$ ;  $S_i$  is the within-class standard deviation and  $S_0$  is set to be the median value of  $S_i$  for all the features. T-statistics is usually used to shrink class means toward the mean of all classes to constitute a nearest shrunken centroid classifier, but do not rank features

regard to all the classes. In our work, another feature ranking method called GINI feature ranking for feature selection is used as discussed in the next section.

## 5.2. *GINI feature selection*

Apart from the statistical based feature selection likes T-test to rank the features for selection, another type of feature ranking is commonly used which is based on information theories. This type of feature ranking is normally called correlation based feature selection. It uses a correlation based heuristic to evaluate the worth of features (Hall and Simth, 1998). In order to perform the evaluation, a heuristic will be used. This heuristic takes into account of the usefulness of individual features for predicting the class label along with the level of intercorrelation among them. The hypothesis on which the heuristic can be stated as: Good feature subsets contain features highly correlated with the class, yet uncorrelated with each other.

Classification tasks in machine learning often involve learning from categorical features, as well as those are continuous or ordinal. A measure based on information theory estimates the degree of association between nominal features. If  $X$  and  $Y$  are discrete random variables, the entropy of  $Y$  before and after observing  $X$  are given as:

$$H(Y) = -\sum_{y \in Y} p(y) \log_2 p(y), \quad (6)$$

$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2 p(y|x). \quad (7)$$

The amount by which the entropy of  $Y$  decreases reflects the additional information about  $Y$  provided by  $X$  and is called the information gain (Quinlan, 1993). Information gain is given by:

$$\begin{aligned} \text{gain} &= H(Y) - H(Y|X) \\ &= H(Y) - H(X|Y) \\ &= H(Y) + H(X) - H(X, Y) \end{aligned} \quad (8)$$

Information gain is a symmetrical measure – that is, the amount of information gained about Y after observing X is equal to the amount of information gained about X after observing Y. Unfortunately, information gain is biased in favour of features with more values, that is, attributes with greater numbers of values will appear to gain more information than those with fewer values even if they are actually no more informative. The purpose of feature selection is to decide which of the initial features to include in the final subset and which to ignore. If there are  $n$  possible features initially, then there are  $2^n$  possible subsets. The only way to find the best subset would be to try them all – this is clearly prohibitive for all but a small number of initial features. GINI feature ranking method is a simple way to task for the feature selection even if there are a lot of possible features to be selected.

Gini Index selects features based on information theories (Hall and Smith, 1999; Zhou and Dillion, 1988). It measures the impurity for a group of labels. Gini Index for a given set  $S$  of points assigned to, for example, two classes  $C_1$  and  $C_2$  is given below:

$$GINI(s) = 1 - \sum_{j=1,2} [p(C_j | s)]^2. \quad (9)$$

With  $p(C_j | s)$  corresponds to the frequency of class  $C_j$  at set  $S$ . The maximum value as  $1 - \frac{1}{nc}$  occurs when features are equally distributed among all classes, which implies less interesting information. On the other hand, the zero value occurs when all points belong to one class which represents the most interesting information. We then sort the  $n$  features over different classes of samples in ascending order based on their best Gini index. Low Gini index corresponds to high ranking discriminative features.

## 6. Classification

After the feature components are ranked and selected, the facial expression recognition tasks are essential to make use of the feature vectors to discriminate and identify each of the expressions. Various pattern classifiers such as *Support Vector Machines (SVM)*, *K-nearest neighbors (K-NN)*, *Naïve Bayes Algorithm* and *Artificial Neural Networks* could be used for emotions recognition. SVM is a learning technique developed by Vapnik (1995) which was strongly motivated by results of statistically learning theory. SVM operates on the principle of induction, known as structural risk minimization, which minimizes the upper bound of the generalization error. K-nearest neighbors method is a nonparametric technique in pattern recognition, which is used to generate  $k$  numbers of nearest neighbors' rules for classification. Naïve Bayes algorithm is based on the Bayesian decision theory, which is a fundamental statistical approach to the problem of pattern classification. This approach is based on quantifying the tradeoffs between various classification decisions using probability and the costs that accompany such decisions.

Naive Bayesian classifier is a simple probabilistic classifier based on applying Bayesian' theorem with strong (Naive) independence assumptions. Independent feature model is a more descriptive term for the underlying probability model. Parameter estimation for Naive Bayesian models uses the method of maximum likelihood in many practical applications; that is, one can work with the Naive Bayesian model without believing in Bayesian probability or using any Bayesian methods. Naive Bayesian classifiers can be trained very efficiently in a supervised learning setting, depending on the precise nature of the probability model. Naive Bayesian classifiers often work much better in many complex real-world situations than one might expect, in spite of their Naive design and apparently over-simplified assumptions. An advantage of the Naive Bayesian classifier is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because

independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix. Detailed analysis of the Bayesian classification problem has shown that there are some theoretical reasons for the apparently unreasonable efficacy of Naive Bayesian classifiers (Rish, 2001) recently. One should notice that the independence assumption may lead to some unexpected results in the calculation of posteriori probability. When there is a dependency between observations, the above-mentioned probability may contradict with the second axiom of probability by which any probability must be less than or equal to one. The Naive Bayesian classifier has several properties that make it surprisingly useful in practice despite the fact that the far-reaching independence assumptions are often inaccurate. For example, the decoupling of the class conditional feature distributions means that each distribution can be independently estimated as a one dimensional distribution. This will help to alleviate problems stemming from the curse of dimensionality, such as the need for data sets that scale exponentially with the number of features. It arrives at the correct classification as long as the correct class is more probable than any other class; hence class probabilities do not have to be estimated very well. In other words, the classifier is robust enough to ignore serious deficiencies in its underlying Naive probability model.

Given a set of test data  $X$  and a posteriori probability of a hypothesis  $H$ ,  $P(H|X)$ , it may follow the Bayesian theorem (Rish 2001) as:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}. \quad (10)$$

Informally, this can be written as  $\text{posteriori} = \text{likelihood} \times \frac{\text{prior}}{\text{evidence}}$ . It predicts  $X$  belongs to

$C_2$  iff the probability  $P(C_2|X)$  is the highest among all the  $P(C_k|X)$  for all the  $k$  classes. The detailed steps are as follow: Firstly, let  $D$  be a training set of tuples and their associated class labels, and a test tuple that is represented by an  $n$ -D attribute vector  $X = (x_1, x_2, \dots, x_n)$ .

Suppose there are  $m$  classes  $C_1, C_2, \dots, C_m$ . The classification is to derive the maximum posteriori, i.e., the maximal  $P(C_i|X)$ . Alternatively, a simplified assumption: attributes are conditionally independent (i.e., no dependence relation between attributes). This greatly reduces the computation cost: Only counts the class distribution. The formula is stated as below:

$$\begin{aligned} P(X|C_i) &= \prod_{k=1}^n P(X_k|C_i) \\ &= P(X_1|C_i) \times P(X_2|C_i) \times \dots \times P(X_n|C_i) \end{aligned} \quad (11)$$

Bayesian networks can represent joint distributions we use them to compute the posterior probability of a set of labels given the observable features, and then we classify the features with the most probable label. The idea is to use a strategy that can efficiently search through the whole space of possible structures and to extract the ones that give the best classification results.

Consider the problem of classifying facial expression by features, for example into joy and non-joy. Imagine that image are drawn from a number of classes of facial features which can be modeled as sets of words where the (independent) probability that the  $i$ -th feature of a given image occurs in a image from class  $C$  can be written as  $P(W_i|C)$ . For this treatment, we can simplify further by assuming that the probability of a feature in an image is independent of the total number of features, or that all image contain same number of features. Then the probability of a given image  $E$ , given a class  $C$  is

$$P(E|C) = \prod_i^n P(W_i|C). \quad (12)$$

Bayesian' theorem manipulates these into a statement of probability in terms of likelihood given by:

$$P(C|E) = \frac{P(C)}{P(D)} \times P(E|C). \quad (13)$$

Our proposed algorithm adopted the Naïve Bayesian with boosting. Each iteration of boosting uses the Gini reduction and selection method and to remove redundant features. Gini uses probability of separation on classes for each features, whereas T-test uses the distance based separation for ranking each features. For Gini ranking, the most discriminative feature will give the lowest value. It is a measure to examine whether the data has clear discrimination between the classes. The main reason for using boosting NB approach is that the embedded feature selection technique makes the data more suitable for classification. The algorithm is summarized as below:

---

**Boosting Naïve Bayesian Algorithm:**

---

*GINI feature selections:*

1. Assign each training sample with weight=1.
2. For ten iteration (ten features):
  - Sort features index S.
  - Split S.
  - Break if GINI criterion is satisfied.

*BNB classification:*

1. Apply simple Bayesian to weighted data set.
  2. Compute error rate.
  3. Iterate the training examples.
    - Multiply the weight by  $\frac{e}{1-e}$ .
    - Normalize the weight
  4. Add  $-\log \frac{e}{1-e}$  to weight of class predicted
  5. Return class with highest sum
- 

We assess the ability of the system to recognize different facial expressions. We have adopted the Mind Reading DVD , a computer-based guide to emotions, developed by a team of psychologists led by Prof. Simon Baron-Cohen (2004) at the Autism Research Centre, University of Cambridge. The database contains images of approximately 100 subjects. Facial images are of size 320x240 pixels, 8-bit precision grayscale in PNG format. Subjects' age ranges from 18 to 30. Sixty-five percent were female, 15 percent were African-American, and three percent were Asian or Latino. Subjects were instructed by experimenter to perform a series of facial displays as an example shown in Figure 7. Subjects began each display from a neutral face. Before performing

each display, the experimenter described and modeled the desired display. The model recognizes four types of facial expression: neutral, joy, sadness and surprise. Twenty images were used for training in which 5 images were used for representing each expression. The facial expression recognition result is shown in Figure 8. The confusion matrix is included in the figure as well where the column of the matrix represents the instances in a predicted class, in which each row represents the instances in an actual class. The system correctly recognizes 76.3% of neutral, 78.3% of joy, 74.7% of sad and 78.7% of surprise expressions amongst 100 subjects in the database, although some facial expressions do get confused by the wrong class, however at an acceptable range of less than 12%. In addition, comparisons with other approaches are necessary for us to investigate how the recognition performance of our approach can be benchmarked with others. Table 1 shows the recognition results for facial expression recognition using T-test and GINI as feature selection and Euclidean, k-nearest neighbour (kNN) as classifiers to compare our boosting Naïve Bayesian (GINI and Naïve Bayesian) approach. Gini processes 854 raw features and shrink down the dimension to 20 features to be further processed by classifier. These 20 features are used as it provides the most optimal results. According to the results in the table, the boosting Naïve Bayesian approach achieves the most optimal result. The T-test is able to assess whether the means of different groups are statistically different from each other. kNN is a classification method for classifying objects based on closest training examples in the feature space. We have used  $k=5$  for kNN classifier based on the problem domain. These approaches are generally used for bench-marking. Our approach that combines Gini and Naïve Bayesian achieves average of 75% in which it outperforms the others. The computational time is recorded about 2.1 frames per second in real time implementation.





Figure 7: Four categories of facial expressions. (a) Neutral, (b) Joy, (c) Sad, and (d) Surprise

Table 1: Comparison of feature selection techniques for facial expression recognition. Three feature selection options are compared using Naïve Bayesian and kNN as the classifiers

Feature selection	Naïve Bayesian	kNN	Euclidean
None	53%	57%	58%
T-test	57%	59%	59%
GINI	<b>75%</b>	57%	63%

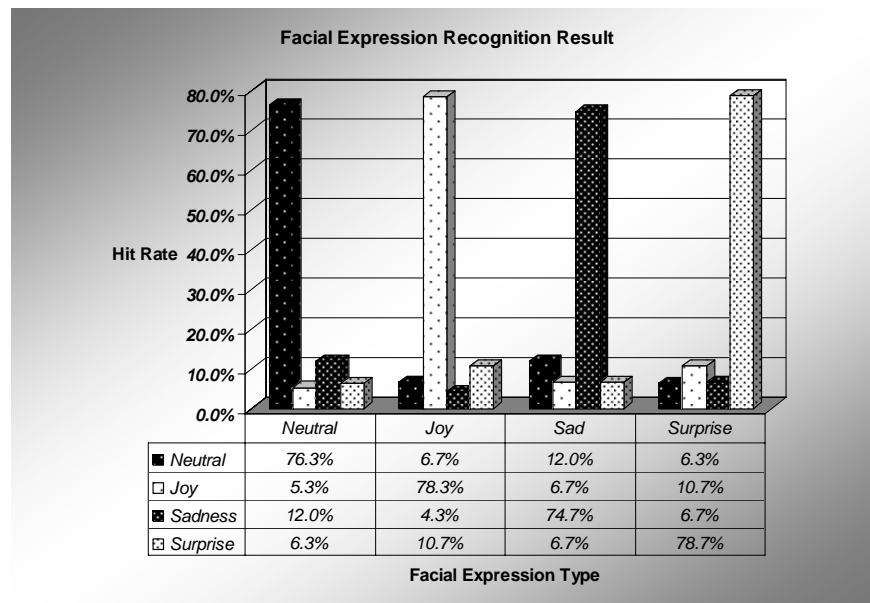


Figure 8: Facial Expression Recognition Result of the System

## 7. Future Trends and Conclusions

Automating the analysis of facial signals, especially rapid facial signals (facial expressions) is important to realize more natural, context-sensitive (e.g., affective) human-

computer interaction, to advance studies on human emotion and affective computing, and to boost numerous applications in fields as diverse as security, medicine, and education. This chapter introduced recent work of our group in this research field.

In summary, although most of the facial expression analyzers developed so far target human facial affect analysis and attempt to recognize a small set of prototypic emotional facial expressions like happiness and anger, some progress has been made in addressing a number of other scientific challenges that are considered essential for realization of machine understanding of human facial behavior. Existing methods for machine analysis of facial expressions discussed throughout this chapter assume that the input data are near frontal-view face image sequences showing facial displays that always begin with a neutral state. In reality, such assumption cannot be made. The discussed facial expression analyzers were tested on spontaneously occurring facial behavior, and extract information about facial behavior in less constrained conditions such as an interview setting. However deployment of existing methods in fully unconstrained environments is still in the relatively distant future. Development of robust face detectors, head and facial component trackers, which will be robust to variations in both face orientation relative to the camera, occlusions, and scene complexity like the presence of other people and dynamic background, forms the first step in the realization of facial expression analyzers capable of handling unconstrained environments.

To date, we have looked into the several aspects of facial expression recognition which are published in separate publications (Cho et al., 2007; 2008; 2009). The achieved developments thus far include the unsupervised learning of facial emotion categorization, the tree structured model of classification and the deployment of the system in hand-held mobile devices. There are two aspects still unsolved. The first issue is how the grammar of facial behavior can be learned and how this information can be properly represented and used to handle ambiguities in the observation data. Another issue is how to include information about the context in which the observed expressive behavior was displayed so that a context-sensitive analysis of facial behavior

can be achieved. Meanwhile, we will also look into explicit modeling of noise and uncertainty in the classification process. The explicit modeling may consist of the temporal dynamic of facial expressions, spontaneous facial expressions, multimodal facial expression classification (Zeng et al. 2009). These aspects of machine analysis of facial expressions form the main focus of the current and future research in the field. Yet, since the complexity of these issues concerned with the interpretation of human behavior at a deeper level is tremendous and spans several different disciplines in computer and social sciences, we believe that a large, focused, interdisciplinary, international program directed towards computer understanding of human behavioral patterns (as shown by means of facial expressions and other modes of social interaction) should be established if we are to experience true breakthroughs in this and the related research fields.

## References

- Anderson, K., & McOwan, P. W. (2006). A real-time automated system for the recognition of human facial expressions. *IEEE Transactions on Systems Man and Cybernetics Part B*, 36(1), 96-105.
- Andrea F. Abate, Michele Nappi, Daniel Riccio, Gabriele Sabatino (2007), 2D and 3D face recognition: A survey, *Pattern Recognition Letters*, Volume 28, Issue 14, Image: Information and Control, 15 October 2007, Pages 1885-1906,
- Bartlett, M. S., Littlewort, G., Fasel, I., & J. R. Movellan. (2003). *Real time face detection and facial expression recognition: Development and applications to human computer interaction*. Paper presented at the CVPR, Madison.
- Bassili, J.N. (1978). Facial motion in the perception of faces and of emotional expression. *J. Experimental Psychology*, Vol. 4, No. 3, pp. 373-379. Black, M. J., & Yacoob, Y. (1997). Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion. *International Journal of Computer Vision*, 25(1), 23-48.
- Bourel, F. C., C.C.; Low, A.A.;. (2002). *Robust facial expression recognition using a state-based model of spatially-localised facial dynamics*. Paper presented at the IEEE International Conference on Automatic Face and Gesture Recognition, 20-21.
- Chandrasiri, N. P., Park, M. C., Naemura, T., & Harashima, H. (1999). *Personal facial expression space based on multidimensional scaling for the recognition improvement*. Paper presented at the Proceedings of the Fifth International Symposium Signal Processing and Its Applications, 22-25.

- Cho S.Y. and J.-J. Wong (2008), "Human face recognition by adaptive processing of tree structures representation", *Neural Computing and Applications*, 17 (3), pp. 201-215, 2008.
- Cho S.Y. and Y.-Y. Nguwi (2007), "Self-Organizing Adaptation for Facial Emotion Mapping", *The 2007 International Conference on Artificial Intelligence*, June 2007, Las Vegas, US.
- Cho S.Y., T.-T. Teoh and Y.-Y. Nguwi (2009), "Development of an Intelligent Facial Expression Recognizer for Mobile Applications", in *First KES International Symposium on Intelligent Decision Technologies*.
- Cohn, J. F., Zlochow, A. J., Lien, J. J., & Kanade, A. T. (1998). *Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression*. Paper presented at the International Conference on Face & Gesture Recognition.
- Cottrell, G. W., & Fleming, M. K. (1990). *Categorisation of Faces Using Unsupervised Feature Extraction*. Paper presented at the Int'l Conf. Neural Networks, San Diego.
- Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*: J. Murray, London.
- Daugman J. (1985) Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Amer.*, vol. 2, no. 7, pp. 1160-1169.
- Devore, J. and Peck, R. (1997). *Statistics: The Exploration and Analysis of Data* (third edition). Duxbury Press, Pacific Grove, USA.
- Edwards, G.J., Cootes, T.F. & Taylor, C.J. (1998). Face Recognition Using Active Appearance Models, *Proc. European Conf. Computer Vision*, Vol. 2, pp. 581-695.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J. Personality Social Psychol*, 17(2), 124-129.
- Elkman P., Huang T.S., Sejnowski T.J., & Hanger, J.C., (Eds.), (1993). *NSF Understanding the Face*, A Human Face eStore, Salt Lake City, 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(7), 757-763.
- Fasel, I.R. (2006). *Learning Real-Time Object Detectors: Probabilistic Generative Approaches*. PhD thesis, Department of Cognitive Science, University of California, San Diego, USA.
- Fasel, I.R., Fortenberry, B. & Movellan, J.R. (2005). A generative framework for real time object detection and classification. *Int'l J Computer Vision and Image Understanding*, Vol. 98, No. 1, pp. 181-210.
- Fasel, B., & Lütting, J. (2000). *Recognition of Asymmetric Facial Action Unit Activities and Intensities*. Paper presented at the Proceedings of the International Conference on Pattern Recognition, Barcelona, Spain.

- Fasel, B., & Luetten, J. (2003). Automatic facial expression analysis: a survey. *Pattern Recognition*, 36(1), 259-275.
- Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3: 1157-1182.
- Hall, M. A. and Smith, L. A. (1998). Practical Feature Subset Selection For Machine Learning. In Proceedings of the 21<sup>st</sup> Australasian Computer Science Conference, 181-191. Springer.
- Hall M. A. and Smith L. A. (1999) Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper. in FLAIRS Conference, pp. 235–239.
- Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *J. Physiol.*, 160, 106-154.
- Jaeger, J., et al. (2003). Improved gene selection for classification of microarrays. *Pac. Symp. Biocomput.* 53-94.
- Ji, Y. Z. Q. (2005). Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5), 699-714.
- Jones J.P., L.A. Palmer. (1987) An evaluation of the Two-Dimensional Gabor Filter model of simple Receptive fields in cat striate cortex, *J. Neurophysiol.*, vol. 58 (6), pp. 1233-1258.
- Kevin W. Bowyer, Kyong Chang, Patrick Flynn, A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition, *Computer Vision and Image Understanding*, Volume 101, Issue 1, January 2006, Pages 1-15, ISSN 1077-3142, DOI: 10.1016/j.cviu.2005.05.005.
- Kim, D.-J., Bien, Z., & Park, K.-H. (2003). *Fuzzy neural networks(FNN)-based approach for personalized facial expression recognition with novel feature selection method*. Paper presented at the IEEE International Conference on Fuzzy Systems.
- La Cara G.E., M. Ursino, M. Bettini. (2003) Extraction of Salient Contours in Primary Visual Cortex: A Neural Network Model Based on Physiological Knowledge. Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE, Vol. 3, 17-21 Sept. pp. 2242 – 2245.
- Lanitis, A. T., C.J.; Cootes, T.F. (1997). Automatic interpretation and coding of face images using flexible models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7), 743-756.
- Levner, I. (2005). Feature selection and nearest centroid classification for protein mass spectrometry. *BMC Bioinformatics* 6: 68.
- Li, S.Z. & Jain, A.K., (Eds.), (2005). *Handbook of Face Recognition*, Springer, New York, USA.
- Littlewort, G., Bartlett, M.S., Fasel, I., Susskind, J. & Movellan, J. (2006). Dynamics of facial expression extracted automatically from video. *J. Image & Vision Computing*, Vol. 24, No. 6, pp. 615-625.

- Mase, K., & Pentland, A. (1991). Recognition of facial expression from optical flow. *IEICE Trans.*, 74(10), 3474-3483.
- Matsuno, K., Iee, C.-W., & Tsuji, S. (1994). *Recognition of Human Facial Expressions Without Feature Extraction*. Paper presented at the ECCV.
- Otsuka, T., & Ohya, J. (1998). *Spotting segments displaying facial expression from image sequences using HMM*. Paper presented at the IEEE Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, Japan.
- Pantic M. (2006). Face for Ambient Interface, Lecture Notes in Artificial Intelligence, vol. 3864, pp. 35-66.
- Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: the state of the art. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(12), 1424-1445.
- Pardas, M. B., A.; Landabaso, J.L. (2002). *Emotion recognition based on mpeg4 facial animation parameters*. Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing.
- Quinlan, J. R. (1993). C4.5: Program for Machine Learning. Morgan Kaufmann.
- Rish I. (2001). An empirical study of the naïve Bayes classifier”, Technical Report RC 22230.
- Rosenblum, M., Yacoob, Y., & Davis, L. S. (1996). Human Expression Recognition from Motion Using a Radial Basis Function Network Architecture. *IEEE Transactions on Neural Networks*, 7(5), 1121-1138.
- Sabatini, S. P. (1996). Recurrent inhibition and clustered connectivity as a basis for Gabor-like receptive fields in the visual cortex. In R. M. Joseph Sirosh, and Yoonsuck Choe (Ed.), *Lateral Interactions in the Cortex: Structure and Function*. Austin, TX: The UTCS Neural Networks Research Group.
- Shin, Y., Lee, S. S., Chung, C., & Lee, Y. (2000, 21-25 Aug 2000). *Facial expression recognition based on two-dimensional structure of emotion*. Paper presented at the International Conference on Signal Processing Proceedings.
- Simon Baron-Cohen, Ofer Golan, Sally Wheelwright, and Jacqueline J. Hill (2004). *Mind Reading: The Interactive Guide to Emotions*, London: Jessica Kingsley Publishers.
- Su, Y. et al. (2003). RankGene: identification of diagnostic genes based on expression data. *Bioinformatics* 19: 1578-1579.
- Vapnik V.N. (1995), *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.
- Viola, P. & Jones, M. (2004). Robust real-time face detection. *J. Computer Vision*, Vol. 57, No. 2, pp. 137-154.

- Vukadinovic, D. & Pantic, M. (2005). Fully automatic facial feature point detection using Gabor feature based boosted classifiers, *Proc. IEEE Int'l Conf. Systems, Man and Cybernetics*, pp. 1692-1698.
- Wang, L. and Fu, X. (2005). *Data Mining with Computational Intelligence*. Springer, Berlin, Germany.
- Whitehill, J. & Omlin, C. (2006). Haar Features for FACS AU Recognition, *Proc. IEEE Int'l Conf. Face and Gesture Recognition*, 5 pp.
- Wong, J.-J., & Cho, S.-Y. (2006). *Facial emotion recognition by adaptive processing of tree structures*. Paper presented at the Proceedings of the 2006 ACM symposium on Applied computing, Dijon, France.
- Wu, B. et al. (2003). Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics* 19: 1636-1643.
- Wu, Y., Liu, H., & Zha, H. (2005). *Modeling facial expression space for recognition*. Paper presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems.
- Xiang T., M. K. H. L. a. S. Y. C. (2007). Expression recognition using fuzzy spatio-temporal modeling. *Pattern Recognition*, 41(1), 204-216.
- Yacoob, Y., & Davis, L. S. (1996). Recognizing Human Facial Expressions from Long Image Sequences using Optical Flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(6), 636-642.
- Yang, M.H., Kriegman, D.J. & Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, pp. 34-58.
- Yeasin, M., & Bulot, B. (2005). *Comparison of linear and non-linear data projection techniques in recognizing universal facial expressions*. Paper presented at the IJCNN.
- Zeng, Z., Fu, Y., Roisman, G. I., Wen, Z., Hu, Y., & Huang, T. S. (2006). *One-class classification for spontaneous facial expression analysis*. Paper presented at the International Conference on Automatic Face and Gesture Recognition.
- Zeng Z., M. Pantic, G.I. Roisman and T.S. Huang (2009), 'A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions', in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39-58.
- Zhao W., R. Chellappa (2003), A. Rosenfeld, P.J. Phillips, Face Recognition: A Literature Survey, *ACM Computing Surveys*, 2003, pp. 399-458
- Zhou X. J. and Dillion T. S.. (1988) A Heuristic - Statistical Feature Selection Criterion For Inductive Machine Learning In The Real World. in *Systems, Man, and Cybernetics*, Proceedings of the 1988 IEEE International Conference on, vol. 1, Aug 1988, pp. 548-552.