9.0 CONCLUSIONS

This research addressed the problem of automatic facial expression recognition based on FACS AUs. We have developed a computer vision system that automatically recognizes facial expressions with subtle differences and also estimates expression intensity. Three methods are used to extract facial motion information and estimate motion intensity: feature point tracking using the coarse-to-fine pyramid method, dense flow tracking together with the principal component analysis, and high gradient component analysis in the spatio-temporal domain, and is then used to discriminate subtle differences in facial expressions observed from image sequences of varying lengths. To determine the optimum HMM topology, a method has been developed and successfully applied to the facial expression recognition. Facial expressions of different types, intensities, and durations from a large number of untrained subjects have been tested. The results show that our system has high accuracy in facial expression recognition.

9.1 Contributions

The major contributions of this dissertation are summarized below.

This is the first automatic facial expression recognition system that has the capability of recognizing nine facial expressions based on FACS AUs: 3 subtly different expressions occurring in the upper face region, and 6 occurring in the lower face region. This system has successful trained with a significant number of image sequences containing various expressions from different individuals. It has been tested yielding good recognition performance.

Three methods have been developed for automatic extraction of information in facial expressions from an image sequence. The first method is facial feature point tracking by applying the coarse-to-fine pyramid approach to track facial feature motion starting from selected feature points pertinent to facial expressions. It has subpixel accuracy for both subtle feature motion and large facial motion. A facial expression is then represented by a displacement vector sequence formed by concatenation of displacement vectors of facial feature points. The process is automatic except the interactive initialization of the feature points selection. It may run in real time. The second method is the wavelet-based multi-resolution level dense flow tracking over an entire region. The principal component analysis is then applied to the computed motion field to compress the information in terms of weighted eigenflows. A facial expression is then abstracted as an appropriate weight vector sequence. The third method is the high gradient component analysis in the spatio-temporal domain to automatically extract motion lines and edges corresponding to furrows appearing in face regions. This information is abstracted as a vector sequence where each vector consists of means and variances of high gradient components in 16 blocks for each frame region.

It is our pioneering effort to apply HMM for automatic recognition of facial expressions in image sequences of arbitrary lengths. For each type of the facial motion extraction and with the encoded symbol sequences obtained therefrom as inputs, two HMMs have been constructed: one models 3 facial expressions in the upper face region, and the other models 6 facial expressions in the lower face region. These models are of low order with simple feedforward connections and are used in the maximum likelihood decision. They have been successfully trained and tested yielding high recognition rate.

A method has been developed to determinate an “optimal” HMM topology for modeling facial expressions, “optimal” in the sense of plausible fewest elements of model parameters. Each state of the HMM corresponds to one level of motion of facial expressions. The number of states required for each HMM is proportional to some measure of variation in motion among the training data of different classes. Applying these method gives a 2nd-order 3-state left-right HMM for modeling the upper facial expressions and a 3rd-order 4-state left-right HMM for modeling the lower facial expressions, respectively.

Differences in facial expression intensity can be used to discriminate between deliberate and spontaneous expressions, measure the degree of emotion, and analyze and synthesize facial expressions for teleconferencing or MPEG-4 applications. Expression intensity estimation may also provide an index for expression segmentation. We have given a method for estimating expression intensity based on each type of the extracted expression information and either a nonlinear mapping or a minimum distance criterion.

9.2 Suggestions for Future Work

Potential applications of our automatic facial expression recognition system include: assessment of nonverbal behavior in clinical and research settings such as psychological research of facial behavior coding, the communication between parents and preverbal infants, biomedical applications such as pre/post surgical path planning or clinical improvement prediction, law enforcement for lie detection, tiresome detecting, detection of tired drivers to help avoid car accidents, lip-reading to compliment speech recognition (audio-vision analysis), teleconferencing and MPEG-4 by analyzing and synthesizing facial signals, animation of the face, and the human-computer interface/interaction to make a computer able to “see.”

Based on the work in this dissertation, the following problems are suggested for further research:

1. Recognition of more facial expressions by adding more “expression units” of individual AUs and AU combinations into the automatic recognition system.

2. Separation of non-rigid facial motion from rigid head motion when the latter involves a motion greater than $\pm 30^0$.

3. Detail comparison of three methods of extraction of facial expression information by using a larger but the same set of training and testing data.

4. Studies on Hidden Markov Model integration either from the view point of a multi-dimensional HMM with multiple inputs or from the view point of decision integration of multiple HMMs.

5. Investigations on computational issues with regard to (a) both 2-level and more than 2-level wavelet-based dense flow estimation, and (b) eigenflow computation.

6. Standardization of expression intensities.

7. Automatic segmentation of facial expression subsequences from a video sequence based on, for example, expression intensity estimation, for use in a real-time system.

8. Implementation of a real-time system. At the present time, the facial feature point tracking is the most likely method to approach running in real time; automatic segmentation of facial expression subsequences will be one of the key problems to be solved.