論文の内容の要旨

論文題目　Early Facial Expression Recognition with Subtle Feature Analysis

微細な特徴の解析による顔表情の早期認識に関する研究

氏名　蘇　鷺梅

The early facial expression recognition task is one to recognize facial expressions as early as possible. It differs significantly from the conventional facial expression recognition task that is aimed as recognizing fully displayed facial expressions at the apex. The conventional method often performs poorly at recognizing subtle facial expressions, and this means that early facial expression recognition can make important contributions for developing techniques for natural human-computer interfaces and human affective computing.

Three issues play important roles in determining the performance of early facial expression recognition. Firstly, the facial deformation is subtle in the early stage of displaying an expression, and this makes it difficult to extract subtle facial features from 2D images or 3D dense data. Secondly, the subtle facial features are very sensitive to noise due to their low intensity, and the noise would unavoidably affect the early recognition results. Lastly, it is difficult to develop an early facial expression classifier that can achieve the conflicting goals of higher classification accuracy and early recognition time simultaneously.

This thesis focuses on developing techniques to resolve the above mentioned issues in early facial expression recognition. In particular, a feature magnification method is proposed to extract discriminative facial features from subtle facial deformation. Moreover, to reduce the influence of noise on the expression recognition performance, two feature refinement methods are developed by analyzing

the characteristics of noise existing in subtle facial features. Furthermore, an early classifier that best suits the early facial expression classification problem is constructed by maximally utilizing the expression information in subtle facial features.

The first part describes an expression category- and intensity- dependent feature magnification method to extract subtle facial features. Because the facial deformations are so subtle, it is extremely difficult to extract discriminative features from different facial expression categories. By and large, the existing facial feature extraction methods, ranging from simple feature descriptors like the pixel intensity difference and facial geometry variations to facial motion models like optical flow and the active appearance model, extract obvious facial deformations. To extract discriminative features from subtle facial expressions, an expression category- and intensity- dependent feature magnification approach was proposed to magnify the captured subtle expressions into the corresponding exaggerated expressions due to the complexity of the motions in facial expressions. By considering the cases of pre-magnified subtle expressions, an expression category- and intensity- dependent magnification factor can be prepared for each possible subtle expression with a certain expression category and certain intensity level. Based on these magnification factors, subtle facial expressions can be reliably transformed into much more discriminative exaggerated expressions and thereby a significant improvement can be made in the performance of subtle expression recognition. Experiments corroborated that the expression category- and intensity dependent feature magnification method outperforms methods without motion magnification and even worked well at recognizing exaggerated facial expressions.

The second part describes two subtle feature refinement approaches to reduce the influence of noise on subtle facial features. The low intensity of subtle facial expression features (deformations) makes them very sensitive to noise and the noise can easily affect the early recognition result. Conventional facial expression recognition mainly focuses on recognizing obvious facial expressions and often ignores the influence of noise on feature classification. On the other hand, existing feature refinement approaches (such as principal component analysis and filtering) cannot successfully reduce the influence of noise on subtle facial features. This is because they only work when the facial features compose most of the principle component in feature spatial space or the noise is distributed at higher frequencies. In the case of subtle facial expression recognition, the noise is probably nearly equal in intensity to the subtle facial features in spatial domain and is probably distributed at lower frequencies. Therefore, to alleviate the influence of noise on early facial expression recognition, two feature refinement methods were devised to enhance subtle facial features. One is adaptive wavelet spectral subtraction, which spatial-temporally refines subtle facial expression deformation with an estimated noise model. In particular, a wavelet packet method is used to analyze the spatial-temporal characteristics of the noise in subtle features. To the best of our knowledge, this is the first effort that refines subtle facial features in the spatial-temporal

domain. The estimated noise model is then used to adaptively reduce the noise not only at high frequencies but also at low frequencies. The other subtle feature refinement method is LDA-based support vector machine, which combines the idea of linear discriminant analysis (LDA) with support vector machine (SVM). The LDA-based SVM method refines subtle features by compacting noise and maximizing the class separability of subtle features without requiring a noise model. The margin of the LDA-based SVM can be enlarged, and consequently, the classification performance improves. The final goal of feature refinement is improving the classification performance. Generally speaking, feature refinements are independently performed before feature classification. Therefore, the improvement in classification performance from the feature refinements cannot be directly evaluated. The LDA-based SVM integrating together feature refinement and feature classification can improve the classification performance by directly reducing the influence of noise on feature classification. Experiments corroborated that the above described feature refinement methods outperform other feature refinement methods by enhancing the discriminability of subtle facial expression features and consequently make correct recognitions earlier.

In the third part, an early classifier based on early RankBoost is proposed for early facial expression classification. Previous early classifiers are based on structured frame classifiers, and each frame classifier is trained to classify similar frame features. Since subtle facial features are similar because of the subtlety of the corresponding facial changes, these frame classifiers are poor at finding discriminative features. On the other hand, to improve the performance of frame classifiers, a temporal alignment or interpolation compensation is usually performed on expression sequences with different lengths and speeds in order to warp similar frame features. However, such methods cannot be used with a real-time human computer interface. An early classification method based on early RankBoost was devised to solve the early facial expression recognition problem. In most cases, the facial expression intensity increases monotonically from neutral to apex. This observation was exploited to develop a solution for early facial expression recognition. In order to find the most discriminative features of subtle facial expressions, frame rankers are introduced to learn the temporal variation of pair-wise subtle facial expression features in accordance with their temporal order. A weight propagation strategy is then applied to boost the frame ranker into an early recognizer. In this way, the early RankBoost method is capable of finding the most discriminative features from subtle facial features. The rank order of facial features can be naturally decided by the frame order of facial features in training facial expression sequences. The early RankBoost method can also learn and recognize facial expression sequences with different lengths and speeds without requiring temporal alignment or interpolation compensation. Experiments corroborated that the early classifier outperforms other early detection methods and gives promising results on the Cohn-Kanade database and our own dataset built using a high-speed motion capture system.