論 文 の 内 容 の 要 旨

論文題目　Subjective Impression Prediction by Complexity Related Features
（複雑さに関連した特徴を用いた主観的印象予測）

氏　名　孫　理天

As the amount of multimedia content shared and consumed online is growing exponentially, it is becoming more and more important to develop a system that is capable to understand and interpret such content in a way that is consistent with human beings. In addition, from the perspective that the best way to understand ourselves is to build a machine that could think similar with us, predicting users' impression towards such content is of great meaning in both psychology and computer science. Our aim in this thesis is to predict human impressions towards photographs and videos through analyzing both the content and the viewer under the inspiration of psychological works.

Complexity is considered as an important indicator in cognitive process. And a stimulus with a moderate complexity level leads to the most positive cognitive experience. In this work we predict the viewers' impression towards photographs by analyzing the complexity of the content. In addition, the level of cognitive load caught by the complexity of a stimulus could be measured through the viewer's eye movement. Thus, we predict individual impression towards video lectures using gaze information. However, psychological theories concerning complexity are only verified on limited situations, and the relationship between complexity and viewer's experience on extensive scope of application is not yet clear. To these end, we propose a series of complexity related features, verify the relationship between complexity and viewer impression, and predict the subjective impression for both photographs and video lectures.

Firstly, we evaluate the role of complexity played in aesthetic assessment and verify the relationship between complexity and aesthetics on large-scale photographs through computational methods. We designed an experiment to collect human ratings on the complexity of various photos. We proposed a set of visual complexity operators taking reference of the factors used in psychological experiments and extract visual complexity properties of the photograph from the aspects of composition, shape and distribution. We extract a set of visual

complexity features using these operators from various perception cues (VCPC). And we applied gradient boost trees regression on these features to set up the complexity model and showed that the complexity level calculated from the proposed features have a near-monotonic relationship with human beings' beauty expectation on thousands of photos. After that we calculated complexity levels for large-scale photo database, and analyzed the relationship between public aesthetics ratings and complexity level.

Secondly, we built up a hierarchical framework to extract structures of different size and intensity contrast, and applied the visual complexity operators to extract the visual complexity features from hierarchical structure (VCHA). We then applied the VCHA features to estimate the aesthetic quality for photographs. There is no standard training and testing protocol for the public aesthetics dataset, so we conducted various experiments under different conditions in order to ensure fair comparisons with state-of-the-art methods. The experimental results demonstrated that the proposed visual complexity features could outperform existing manually prepared features and even better than deep features for balanced training samples. In addition, the proposed features can be extracted directly from samples without tedious learning stage required by deep features.

Thirdly, we use features extracted from gaze information to predict individual rating for video talks. We constructed a dataset of eye movements during video lecture watching together with viewer's rating. Then we proposed a set of gaze features, which not only include the conventional distribution features but also include the analysis of the relationship between visual saliency and gaze point in both static and dynamic aspects. By doing so, we set up a baseline for researches in personal rating prediction for video lectures using gaze.