学位報告 4

報告番号	※甲　　第　　号

主　論　文　の　要　旨

論文題目　Context-aware User-dependent Viewpoint Recommendation for Multi-view Field Ball Sports Videos
(フィールド球技の多視点映像における
コンテキストアウェアなユーザ依存型視点推薦法)

氏　　名　汪　雪婷 (WANG　Xueting)

論 文 内 容 の 要 旨

As an extension of the widely used single-view video, multi-view videos by multiple cameras offer more viewpoint options. They are playing an important role in representing events in areas of education, surveillance, communication, and entertainment. To make the multi-view video technology more feasible in practice, many research works have been conducted on multi-view video capturing, compression, streaming, and delivery technologies. When multi-view video capturing and streaming are developed in practice, the representation technology on how to show the multi-view contents to users becomes an important issue. Therefore, this thesis focuses on viewpoint recommendation of multi-view videos among the multi-view video processing research targets.

In this thesis, the field ball sports content is focused as the target application, especially the soccer game, which is a typical sport game containing the common components of field ball sports.

Multi-view videos can be customized according to users' preferences, so that users can enjoy interesting content based on their own choices, instead of the fixed contents provided by professional TV broadcasters. However, there exists the problem of continual appropriate viewpoint selection, which leads to the annoyance and stress of a completely manual process if only existing multi-view video interfaces are provided to users. Thus, automatic user-dependent viewpoint recommendation is important for enhancing the advantage of multi-view video viewing.

The goal of the works presented in this thesis is to achieve automatic user-dependent viewpoint recommendation adapted to diverse video contexts and

user contexts. Here, the video contexts are defined as representations of what is happening in the view, consisting of scene contexts and production contexts. The scene contexts include the size and number of visible objects, objects' spatial distribution, and scene events. The production contexts include camera location and optical axis, viewpoint switching angle and frequency. Meanwhile, users' viewpoint selection tendencies are considered as the user contexts, which are related to user personalities, interests, and experience level on the viewing content, and favorite objects. The users' viewpoint selection tendencies could be not only the common viewpoint selection tendency of most users, but also the unique tendency of a specific user or of a group of users with similar tendencies.

To effectively recommend viewpoints with high similarity to users' viewpoint selection tendencies, the recommendation task is considered as a supervised learning problem. The following factors are considered in the recommendation framework: context-dependent learning scheme, spatio-temporal feature representation, and user modelling for group-based recommendation, respectively. The effectiveness of recommendation is evaluated by the degree of similarity between the recommendation and actual users' viewpoint selection records.

This thesis is organized with the following chapters.

Chapter 1 introduces the background, motivation, challenges, and the purpose of the work presented in this thesis. Contributions of the work are also summarized in this chapter.

Chapter 2 reviews related works. An overview of related approaches on automatic video editing and recommendation, and viewpoint selection methods for multi-view videos are provided, and the differences with the proposed methods are discussed.

Chapter 3 describes the creation of a multi-view video dataset used in this thesis, together with user viewpoint selection records acquired through a multi-view video editing experiment.

Chapter 4 proposes a context-dependent learning scheme to adapt to various video contexts and users' viewpoint selection tendencies. Viewpoint evaluation and transition processes are used to represent the video contexts including scene contexts and production contexts by appearance features and transition costs of viewpoints. The proposed method uses different weight parameters to combine the appearance features and the transition costs, and optimizes them for each scene context by minimizing the difference between the generated viewpoint sequences and users' viewpoint selection records. This context-dependent learning scheme can provide both common and personal recommendation by using common or personal viewpoint selection records of users. The effectiveness of this method is confirmed by comparing the generated context-dependent and the independent viewpoint recommendations with actual selections made by users. Chapter 5 proposes a method to improve the

video context representation using spatio-temporal feature representation. It was found that different trajectory distributions of focused objects cause a difference in the viewpoint selection for different users by analyzing users' viewpoint selection tendencies. Thus, the recommendation scheme is trained by learning the relationship between the users' viewpoint selection tendencies and the spatio-temporal scene context represented by object trajectories. Three methods were compared to find the best one to represent this relationship, and the Gaussian Mixture Models (GMM) based method achieved the best performance by assessing the degree of similarity between the viewpoint recommendation and user's selection records. These methods uses users' personal viewpoint selection records as the learning label, but can also be applied to common recommendation by considering common viewpoint selection records of multiple users.

Chapter 6 describes a method that focuses on the cold-start problem of user-dependent recommendation based on a user model for group recommendation. The clod-start problem here is the lack of sufficient user records for model construction. The relationship between viewpoint selection tendency and user attributes is analyzed and applied to solve the problem. A group-based recommendation framework consisting of a user grouping approach based on the similarity with existing users' viewpoint selection records, and a member group estimation approach based on the classification by user attributes are proposed in this chapter. The generated group-based recommendation yielded similar recommendation effectiveness to the personal recommendation, so it can be used when user's personal records are insufficient. The users with high emotional stability in personality and high-level experience in the target sport trend to have stable viewing pattern and receive better user-dependent viewpoint recommendations.

Finally, Chapter 7 concludes this thesis and presents directions for future work.

学位関係