Emocap: Video Shooting Support System for Non-Expert Users

Hiroko Mitarai, Japan Advanced Institute of Science and Technology, Japan
Atsuo Yoshitaka, Japan Advanced Institute of Science and Technology, Japan

ABSTRACT

Authoring quality video content is difficult since proper camerawork is required for delivering content which appropriately reflects user’s expressive intentions. Based on an experimental result on non-expert users, the incremental interaction model was formerly proposed. The authors propose a system based on the model which compensates for user’s lack of cinematographic knowledge or skills by relating affective information such as atmosphere or mood to shooting techniques. After selecting a specific type of atmosphere to express non-verbal information, the system analyzes the shooting image and the camera operation including the camera angle and the zooming speed to assist the user. The appropriate parameters for shot sizes, camera angles and zoom speed are based on the analysis of extracted features from ten highly-evaluated films. The system evaluation has indicated that the system assists the user in reflecting user intention of the shot appropriately; therefore it enables the user to capture shots more appropriately and effectively without cinematographic knowledge or skills.

Keywords: Affective Information, Film Grammar, Multimedia IS (Information Systems), Video Authoring, Video Production

INTRODUCTION

Video authoring is becoming more and more accessible for non-professional users. Types of media distribution technologies have branched out as well; aside from TV stations and movie theaters, the internet has enabled broadcasting for nonprofessional users. Technical advances in video cameras have enabled some automated functions such as auto-focus and white balance correction. In video editing, increasing hard disk space has yielded more editing space; people can now easily store dozens of captured movies and perform non-linear editing on their personal computers. Video content production has penetrated to ordinary families along with dissemination of audiovisual equipment.

However, footage shot by nonprofessional users on occasions such as weddings or a baby’s first steps is often left without editing and seldom watched (Kirk, Sellen, Harper, & Wood, 2007). Adams states that inexperienced users tend to move cameras in random directions while shooting, which is one of the factors in disconnect between user’s intention and the delivered media (Adams, Venkatesh, & Jain,
One of the possible reasons for this is that the users shoot subjects too carelessly, and do not consider enough if the footage shot is worth watching, and reflects their intention properly. Mei proposed the notion of capture intention, emphasizing the importance of intention of the user (Mei, Hua, Zhou, & Li, 2007). Even if the footage captured the subject revealing strong emotions, these emotions cannot be expressed effectively with dull wide-angle shots. Impressive moments cannot be expressed without the proper rendition of the scene, using appropriate camerawork.

Psychological impact of visual images is enormous. Professionals skillfully express their opinions utilizing film/video production techniques, also referred to as film grammar or film language, cultivated by film directors and theorists. However, as this method is called grammar, there are rules to follow in order to create visually effective story development, such as where to place the camera when shooting and how the captured footage should be placed when editing. This is not always easy to learn. Since determining appropriate camerawork is difficult for unskilled or inexperienced users, this difficulty often hinders the smooth visual communication between the authors and viewers. By utilizing an appropriate visual information expression, viewers can understand the shot content more appropriately.

In this paper, we implemented a system based on the incremental interaction model to assist users in shooting the subject more appropriately by camera operation recognition, utilization of affective information and the film production techniques to apply to actual shots, based on the skills that film professionals have (Mamer, 2000) and the film grammar (Arijon, 1991). After a user enters affective words representing affective information such as emotion, strength or tension/excitement, the system shows corresponding guidance in order to express specified affective information. Since the system follows the incremental interaction model (Mitarai & Yoshitaka, 2010), it assists the user by displaying the guidance according to the capturing images and camera operation of the user such as camera angles, zooms and face size of the subject. Appropriate camera angles, zoom speed and face size of the subject have been determined according to the parameter analysis on ten highly-evaluated Hollywood movies.

While the user follows the guidance, and the system continually senses the change of the camera operation and gives guidance accordingly, therefore nonprofessional users can shoot the scenes in a proper way so as to express desired atmosphere. The system is comprised of a video camera, acceleration sensor, zoom speed detection unit, LCD panel with touch screen and a laptop PC. A system performance evaluation was carried out on ten graduate students. Experimental results showed that the proposed method is effective for inexperienced users to shoot expressing desired affective information.

The main advantages of the system are that it is the first implementation of affective shooting support system, and that it has sought a legitimate relationship between film grammar and affective information. Although it is significant to capture emotional moments appropriately, affective shooting support had not been investigated prior to our work. Although the implementation yet has some place for improvement, it showed that the system efficiently supports the user to shoot. We focused attention on the relationship between film grammar (Arijon, 1991) and affective information.

RELATED WORK

There are two levels of content understanding in the field of video content retrieval: cognitive level and affective level (Hanjalic & Xu, 2005). Similarly, we suggest that there are two types of assistance in content creation: technical assistance and affective assistance. The technical assistance includes alleviation of camera shakes or correction of optical parameters such as exposure or focus. This type of assistance aims at making the shots easy to watch. The affective assistance includes assistance for appropriate expression of nonverbal
Evaluation of Situations Causing Split of Attention in Multimedia Learning Environments via Eye-Tracking Method
www.igi-global.com/chapter/evaluation-of-situations-causing-split-of-attention-in-multimedia-learning-environments-via-eye-tracking-method/133004?camid=4v1a

Client-Side Relevance Feedback Approach for Image Retrieval in Mobile Environment
www.igi-global.com/article/client-side-relevance-feedback-approach/54461?camid=4v1a

Emoticon Recommendation System to Richen Your Online Communication
www.igi-global.com/article/emoticon-recommendation-system-to-richen-your-online-communication/109076?camid=4v1a