Context-aware Automatic Haptic Effect Generation Algorithm for Improved Content Viewing Experiences

Yaxuan Li, Yongjae Yoo, Antoine Weill--Duflos, Jeremy R. Cooperstock[1]

***Abstract*— We propose an automatic vibrotactile content creation pipeline considering contextual information based on the movie's audiovisual content for improving the overall hapto-audio-visual (HAV) experience. The algorithm extracts significant features from video files and generates the haptic stimuli corresponding to the video content. We implement the algorithm on a low-cost system with nine actuators attached to the chair and will evaluate the system performance with 20 participants using six movie clips in the future.**

I. INTRODUCTION

Multimedia content, such as movies and television programs, is increasingly offering enriched user experiences thanks to the advances of immersive technologies. One such augmentation, the inclusion of haptic effects, has been adopted to provide greater realism and immersion in content viewing experiences in movie theaters and museums. Haptic modalities of vibration, temperature, force, and proprioception have been explored and evaluated, using a variety of hardware configurations from simple, low-cost setups to more bulky and dedicated ones. To achieve a high-quality overall experience, haptic content creation, matching the audiovisual content, is essential. The current practice of such content authoring predominantly relies on the intuition and experience of expert designers. Despite the investigation of automated haptic authoring approaches [1], manual design of haptic effects by practitioners remains the norm. In the present extended abstract, we summarize our ongoing research on automatic haptic effect generation as a step toward improving the adoption of automatic authoring tools.

II. METHODS

A. Audiovisual feature extraction

As auditory features, we exploited psychoacoustic measures to determine relevant attributes of important events from the scene. Our intuition is that violent and highly dynamic scenes, such as those involving explosions or fighting, often contain more high-frequency auditory components than other content. We therefore picked sharpness with an appropriate threshold to determine the keyframes for haptic rendering, as illustrated in Fig.1(b). With regard to visual features, salient region extraction is effective at determining where the events occur on screen [2]. We built a convolutional neural network (CNN) model to detect the salient area of every frame, as illustrated in Fig.1(c), and use this to determine the spatial distribution of vibrotactile feedback to be rendered.

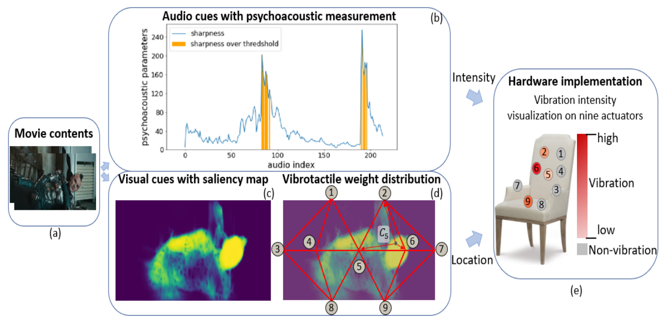

Fig. 1: Haptic effects synthesis algorithm.

B. Hardware implementation and rendering

For a feasibility study, we integrated nine vibration motors (RB-See-403, Seeed Studio) into a chair. Considering the layout of the actuators, we divide the saliency map into 14 triangles, as illustrated in Fig.1(d). In each triangle, we calculate the centroid C_i of the saliency shape, which is predicted by the CNN model and shown as yellow pixels in Fig.1(c) and (d), $i \in \{1,\ldots,14\}$, and the distance to three adjacent vertices. Fig.1(d) provides an example of C_5 in $\triangle A_2 A_5 A_6$. We assign weights that are inversely proportional to the distance between the centroid and adjacent vertices and determine the vibration intensity of the corresponding actuators for haptic rendering. Contextual information from audio and visual content is fused with psychoacoustic features to determine the time (or frames) and intensity of vibration and saliency map for spatial rendering as shown in Fig.1(e).

III. CONCLUSION AND FUTURE WORK

This introduces our ongoing research of developing an automated haptic authoring algorithm that extracts contextual information and features from audiovisual contents. Our next step is to evaluate the hapto-audio-visual experience in terms of immersion, preference, discomfort, and harmony by 20 participants with six short (approximately one-minute) movie clips.

*Research supported by InterDigital Corporation.

[1]Yaxuan Li, Yongjae Yoo, Antoine Weill--Duflos, and Jeremy R. Cooperstock are with the Department of Electrical and Computer Engineering, Centre for Intelligent Machines (CIM), and Centre for Interdisciplinary Research in Music, Media and Technology (CIRMMT), McGill University, 3480 Rue University, Montréal, H3A 0E9, Quebec, Canada {yaxuan, yyoo, antoinew, jer}@cim.mcgill.ca

REFERENCES

[1] J. Lee, B. Han, and S. Choi, "Motion effects synthesis for 4d films," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 10, pp. 2300–2314, 2015.

[2] M. Kim, S. Lee, and S. Choi, "Saliency-driven tactile effect authoring for real-time visuotactile feedback," in *Proc. Eurohaptics'12 Part I*, ser. LNCS, vol. 7282, 2012, p. 258–269.