Saliency-driven Real-time Tactile Effects Authoring

2012

Myongchan Kim

Master's Thesis

Saliency-driven Real-time Tactile Effects Authoring

Myongchan Kim (김명찬)
Division of Electrical and Computer Engineering
Pohang University of Science and Technology

2012

영상주목도 기반 진동 효과 자동 저작

Saliency-driven Real-time Tactile Effects Authoring

Saliency-driven Real-time Tactile Effects Authoring

by

Myongchan Kim

Division of Electrical and Computer Engineering

POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

A thesis submitted to the faculty of
Pohang University of Science and Technology
in partial fulfillment of the requirements
for the degree of Master of Science
in the Division of Electrical and Computer Engineering

Pohang, Korea
2012. 6. 27.
Approved by

S.M. Choi

Seungmoon Choi, Academic Advisor

Saliency-driven Real-time Tactile Effects Authoring

Myongchan Kim

The undersigned have examined this thesis
and hereby certify that it is worthy of acceptance
for a master's degree from POSTECH

06. 27. 2012

Committee Chair	**Seungmoon Choi (Seal)**
Member	**Seungyong Lee (Seal)**
Member	**Sung Ho Han (Seal)**

MECE 김명찬, Myongchan Kim, 영상주목도 기반 진동 효과 자동 저작,
20100582 Saliency-driven Real-time Tactile Effects Authoring,
Division of Electrical and Computer Engineering, 2012, 53 p,
Advisor: Seungmoon Choi. Text in English

Abstract

New-generation media such as the 4D film have appeared lately to deliver immersive physical experiences, yet the authoring has relied on content artists, impeding the popularization of such media. An automated approach for the authoring becomes increasingly crucial in lowering production costs and saving user interruption. This paper presents a fully automated framework of authoring tactile effects from existing video images to render synchronized visuotactile stimuli in real time. The spatiotemporal features of video images are analyzed in terms of visual saliency and translated into tactile cues that are rendered on tactors installed on a chair. A user study was conducted to evaluate the usability of visuotactile rendering against visual-only presentation. The result indicated that the visuotactile rendering can improve the movie to be more interesting, immersive, appealing, and understandable. However because the initial algorithm had a weakness in extracting salient areas in real-world situations, the algorithm was revised for improvement. Another user study was conducted to evaluate the usability of visuotactile rendering against the initial method, manual authoring, and visual-only presentation. The result elucidated that the new approach fixed the previous problem and improved the visualtactile environment.

Contents

List of Figures

Chapter 1

Introduction

Recent advances in haptics technologies have proven its worth as an effective communicative source in a wide variety of applications. While the majority of multimedia contents are being mediated through visual and auditory channels, recent research and industrial applications, such as 4D films, are extending beyond the bimodal interaction to encompass diverse physical experiences including vibration, breeze, smell, mist, or tickler [7, 16]. Such new-generation films provide better experiences in terms not only of immersion and entertainment, but also of better content delivery through unallocated haptic sensory channels. As the commercialization of stereoscopic TV was perceived as a particularly successful occasion, it is not far away to *feel physical movies* in everyday life. In creating such immersive experiences, the haptic sensation is one of the crucial components to bring about pervasive changes in multimedia.

1.1 Thesis Motivation & Objective

4D movies can be found at theater easier than before, these contents may be the future interactive entertainment by enriching the audiovisual effects but currently it is difficult to find it elsewhere than theater because they still carry a couple of limitations in terms of gaining popularity from the public due to its

Figure 1.1 D-Box is one of popular motion systems specially designed to offer the 4D effects for the entertainment industry

high cost and developing time.

Because each 4D effect chair at theater is sold more than $50,000, it is not easy for family to own the 4D effect system for their home entertainment yet. 4D contents are also limited in terms of their volumes and contents because they are only possibly produced by limited number of companies. In addition, 4D effects are very dependant on hardware that the current production is processed by creating 4D contents based on the client's specific requirements and hardware. Therefore 4D contents are difficult to reuse it in other systems. The second problem of 4D contents is effect creation authoring consumes much hours. There is no automated process for authoring 4D effects yet, content designers must watch the movie frame by frame, match the sync and duration of tactile signal and actuate the effects at the appropriate time frame. This manual authoring process requires much hours when the effects need to adapt with the surrounding environment. For instance, when a car moves from right to left and the user needs to feel the physical movement from right to left, the physical effect at the right area should be gradually decreased while the physical effect at the left area should be increased, this drives the authoring process more

Figure 1.2 4D theater in Republic of Korea, known as 'CJ 4DPlex'

complex. Also by considering of the fact of the widely known knowledge that different aged people may understand the physical effects differently, physical effects should be authored to deliver the designer's intention to most people.

Unlike computer vision or graphics fields, many haptics related researches are too difficult to participate in. The history of haptics is not only so long but also the cost of the hardware is not appealing. This may push out researchers to join in the haptics community, however, I believe the 4d effects with commercially acceptable approaches may trigger the more researches to participate in haptics.

In order for 4D effects gain more popularity, the authoring process should be simplified and a solution, method and apparatus for automatic tactile effect authoring will be suggested in this thesis. The automatic authoring will no longer require the tactile effect designers work excessive hours for creating patterns. Also the 4D tactile display solution using vibratory motors would save the apparatus development cost and generate enough vibratory feedbacks to the user to understand the intended physical patterns. I hope the design of automatic software approach and inexpensive hardware design would set a beginning milestone for motivating future 4D haptic researchers.

1.2 Thesis Contribution

This paper reports the research on algorithms to transform streaming visual signals to tactile cues using the visual saliency, a real-time tactile display built upon an array of tactors installed on a chair, and a user study that evaluated the usability of our system. Furthermore, the system is aimed at a real-time interaction system unlike the previous researches, having the great benefit of distributing haptic contents without manual pre-encoding. To my knowledge, this is the first attempt for an automatic, real-time tactile effect authoring system making use of movies. The system can play synchronized visuotactile effects in real time directly from a movie source. In addition, it has a potential advantage for human-aided design. An initial seed can be provided rapidly by our system, and then designers can take it over and enhance the tactile scenes, thereby, reducing production costs significantly. This can be a viable alternative considering no semantics is taken into account in our system. In addition, our study uses low-cost vibration motors for tactile rendering. This choice greatly elevates the practicability of our system, but it comes with a large actuation latency to take care of. For synchronous visuotactile stimulation, our system uses asynchronous commanding, that is, issues tactile commands earlier than visual commands by pre-calibrated differences between the display latencies.

1.3 Thesis Outline

This thesis structures as follows: the related work for this thesis will be reviewed in the chapter 2. General saliency algorithm and hardware implementation will be introduced in chapter 3. In chapter 4, the initial rendering algorithm for automatic visuotactile effects and its subjective evaluation will be discussed. In chapter 5, alternative rendering method is proposed and also the new subjective evaluation will be discussed. In chapter 6, the overall discussion will be stated and the thesis is concluded in chapter 7.

Chapter 2

Related Work

A couple of research efforts have been devoted to manually extract tactile information from visual images and to convey the tactile cues to stimulate different body locations to inform the user the system intention. For instance, in the early 1970s, binary pictorial information in a static black-and-white image was used to generate tactile cues, which was delivered by the 400-points two dimensional tactile stimulator mounted on the back cushion of the chair [4]. More recently, streaming 2D or 3D videos were used to generate haptic feedback based on the objects' depth information analyzed from the video scene [1]. The haptic feedback was used to provide shapes of the object with the haptic interface while watching the videos at the same time. Such manual postprocessing was later extended into the area of tactile glove system to give a movie-watcher synchronized visuotactile feedback [11]. The system was focused more on translating important motion segments to tactile feedbacks instead of the objects' shape, which motivated our work with automated approach. One of the notable recent approaches is the use of predefined tactile animation in interactive systems such as video games [8]. Their event-driven authoring opened a new possibility for the authoring of tactile cues for dynamic scenarios.

2.1 4D Movies

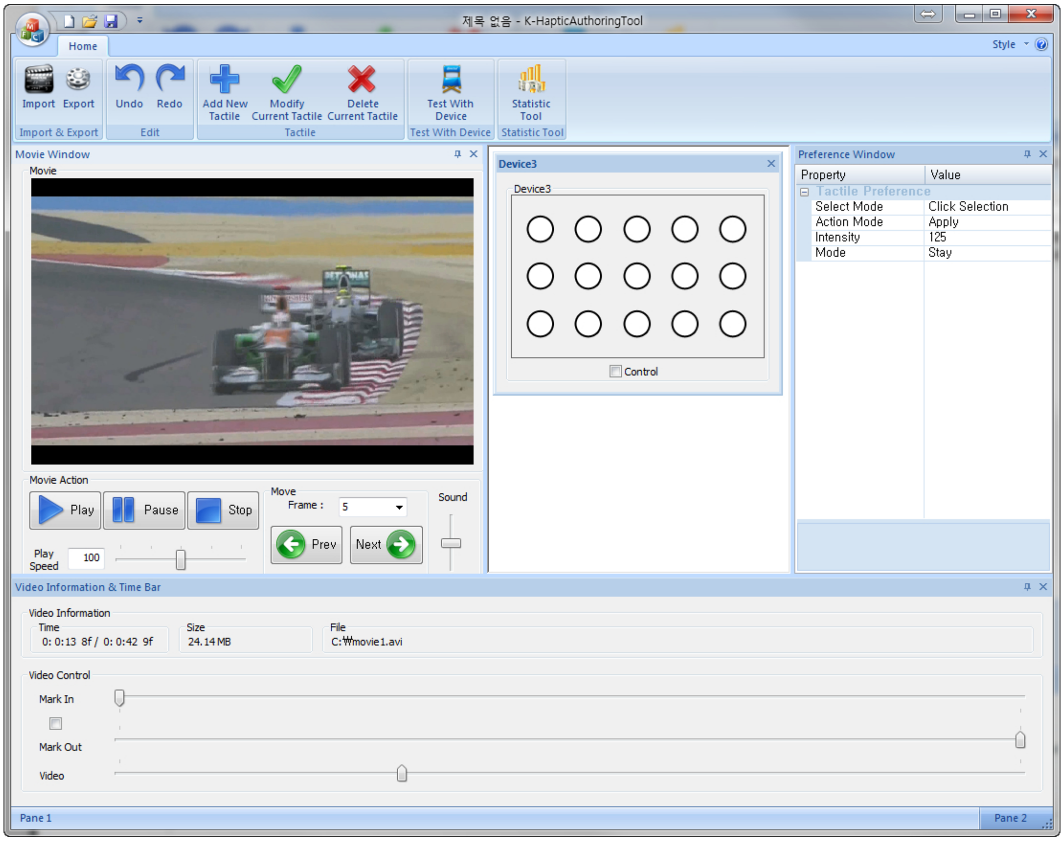
When the haptic or tactile feedback are combined and synchronized with audiovisual media, the system can provide an increasingly immersive experience. The 4D theater is a widely known attempt of implementing such a system, which provides the physical sensation in addition to audiovisual sensation. When the movie is played, pre-written physical effects are provided to the audience simultaneously to emphasize visual cues while keeping them interested and entertained. The audience can send himself a journey through a character narrated throughout the movie as if he is actually playing a role in the scene and sensing realistic motions from various physical effects equipped in the 4D display. Commercial 4D theater systems introduced diverse physical elements, including breeze, vibration, mist, smell, light, tickler or motion control, for serving richer contents and more involvement in the story narrated from the movie [7, 16]. Although commercial 4D theater has already been popularly seen in a few countries including Republic of Korea, Japan and United States of America because of its expensive production process, the entrance fee for is entertainment is not low and not many contents are available yet.

2.2 Tactile Authoring

Providing convenient and sophisticated interfaces for authoring haptic and tactile contents are also important in widely diffusing the haptic experience to novice users. Several research efforts have explored such tools. One of such approaches is authoring, editing, and storing passive haptic media using the binary format extension of MPEG-4 and MPEG-5 standards [2, 20]. This approach was primarily intended for haptic broadcasting, and can be an initiative to spread the pre-encoded haptic media widely. It is also of importance to support intuitive GUI for authoring, in particular for lengthy media and a number of actuators. Lemmens et al. proposed a pattern editor of 64 vibration actuators in tactile jacket to systematically program and manipulate the desired tactile experience [14]. In the tactile glove system proposed by Kim et al. [11], spatiotemporal tactile cues were promisingly produced by manually drawing lines on visual scenes represented on the regular grid. Fig. 2.1 illustrates Kim's latest manual authoring UI. Since this process needs to be carried out carefully by inspecting visual scenes frame by frame, it is a laborious and time-consuming task. In this thesis, it aims at providing automated framework to rapidly create the required tactile cues at minimum costs and seamlessly synchronizing the generated tactile cues with visual cues.

2.3 Saliency Algorithm

Finally, the neuro-scientific background on visual saliency and its computational implementation is briefly reviewed, which is the main motivation to automate the tactile authoring. It is well known that attentional allocation involves reflexive *bottom-up* capture of visual stimuli, in the absence of user's volitional shifts [17, 5]. Whereas humans are generally efficient in searching visual information from complex scene, this does not necessarily mean that everything is perceived simultaneously. After pre-attentive primitives such as color and lightness are first detected in parallel and separately encoded into feature maps, a slow serial conjunctive search is followed to integrate the feature maps into a single topographical *saliency map* [18, 6]. Neuronal mechanisms of early vision underlying bottom-up attention, called the ”center-surround” difference, give us an important insight for detecting salient areas. The periphery of retinal zone (surround) suppresses neuronal activation in narrow receptive fields of the highest spatial acuity (center). Thereby, visual structures are particularly visible well if they are occupying a screen region popping out of its local neighborhood. Inspired by the ”center-surround” mechanisms, Itti et al. made significant contribution in computational models of the *saliency map* based on multiscale image pyramids in the area of computer vision [12, 9]. Their model is



Frame : 5

Play Speed 100

Property	Value
Select Mode	Click Selection
Action Mode	Apply
Intensity	125
Mode	Stay

Time: 0: 0:13 8f / 0: 0:42 9f

Size: 24.14 MB

File: C:₩movie1.avi

Figure 2.1 Kim et al.,'s tactile feedback manual authoring UI. [11] Designed to author tactile effects for 3 x 5 array of tactors. In the top left pane, the video frame can be selected for editing; the tactile cues can be inserted in the middle pane; the tactile cue intensities and the type of insertion can be selected in the right pane; shows the overall timeline and current position in the bottom pane known to be effective for analyzing gaze behaviors in an image, and our saliency model exploits their principle to efficiently attain plausible outcome on visual saliency.

Chapter 3

Implementation

The whole research includes the software implementation and the hardware design. Before the detailed information of two big sections is narrated, a brief overview of the system is illustrated which gives a light head-up to readers of this thesis.

Second, the neuroscientific background on visual saliency and its implementation will be reviewed. It is well known that attentional allocation involves the reflexive (bottom-up) capture of visual stimuli, in the absence of user's volitional shifts [17, 5]. Albeit humans are generally efficient in searching visual information from complex scenes, this does not necessarily mean that everything is perceived simultaneously. Preattentive primitives such as color and lightness are first detected in parallel and then separately encoded into feature maps. A slow serial conjunctive search is followed to integrate the feature maps into a single topographical *saliency map* [18, 6]. Neuronal mechanisms of early vision underlying bottom-up attention give us an important insight for detecting salient areas. The periphery of retinal zone (surround) suppresses neuronal activation in narrow receptive fields of the highest spatial acuity (center). Therefore, visual structures are particularly well-visible when they occupy a region popping out of its local neighborhood.

Lastly, the systematic actuation design will be introduced. Eccentric rotated motor (ERM) was used in this research due to a couple of reasons. One is its affordability and the other is its relatively strong vibratory feedbacks. A circuit was built for actuating multiple vibratory motors attached on a custom made tactile display using signals generated from the computer via Ni-DAQ device.

3.1 Overview of Framework

In this section, a brief perspective on the system will be provided. Fig. 3.1 illustrates the pipeline of our system. In the system, visual and tactile renderings are asynchronously executed using two different threads. For every frame, the thread for visual display runs as usual, but meanwhile, the thread builds the saliency map that spatiotemporally abstracts perceptual importance in a visual scene. The resulting saliency map is translated and stored in tactile buffers, the resolution of which is identical to that of a physical tactor array. In the other thread for tactile rendering, the tactile buffers are read into a tactile map at a lower frame rate, e.g., 5 Hz. This tactile map is mapped to the actuation commands to be sent to the tactors. In particular, the tactile commands are issued in advance to the visual commands with compensation of vibration latency. Each step is detailed in the following sections.

3.2 Tactile Movie Generation using Visual Saliency

Main framework of this thesis includes the use of visual saliency. The idea was first suggested and brought into the algorithm by Professor Sungkil Lee from Sungkyunkwan university thus I could extend his work. I am grateful to Professor Lee for his contribution in this research.

The extraction of visual saliency from an input image basically relies on the typical computational method proposed by Itti et al. [12, 9], which has been distinguished for its effectiveness and plausible outcome in analyzing gaze be-

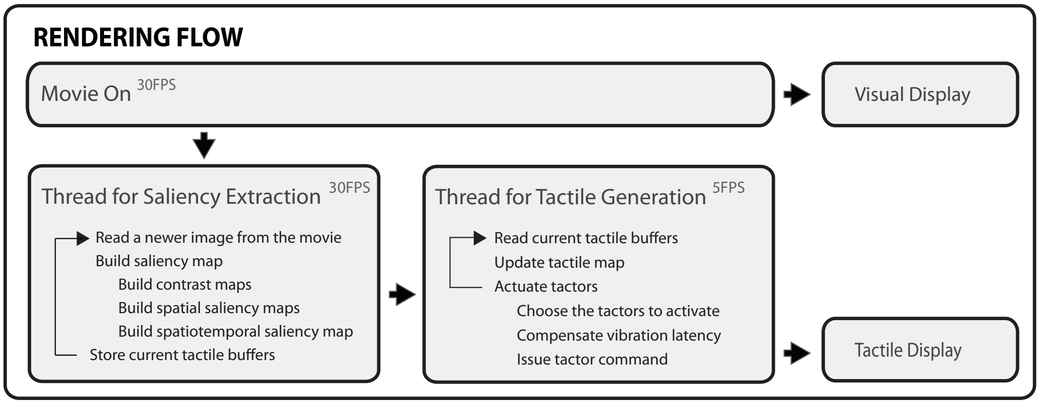


Figure 3.1 Rendering flow of our system. Two threads for visual and tactile display run simultaneously but in different frame rates.

haviors. The key idea of their algorithm is finding salient regions by subtracting a pair of images each other, spatially convolved over different kernel sizes; they further accelerate this procedure using the image pyramid. The image averaged with a smaller blur kernel preserves finer structures than that with a larger blur kernel. Therefore, the image difference between them effectively captures spatially salient areas differing from local neighborhood, which simulates the biological process in the human visual system. Such image difference is called the *center-surround* difference. First their basic algorithm will be described in more details, and then extensions on the use of CIE L*a*b* color space, temporal saliency, and binary thresholding will be described. See also Fig. 3.2 for the whole pipeline of the algorithm.

The basic algorithm of Itti et al. is as follows. Given an input RGB image, visual feature maps (e.g., color, luminance, and orientation) are first extracted. The image pyramid of each feature map is built by successively downsampling an image to the $1/4$ size of its predecessor until reaching the coarsest image of 1×1. For example, for an image with a resolution of $2^N \times 2^N$, the levels of its image pyramid ranges from 0 (the finest image) to N (the coarsest image). For each image pyramid, six pairs of center (c) and surround (s) levels are defined; the common configuration from the previous studies was used, $c \in \{2, 3, 4\}$ and $s = c + \delta$, $\delta \in \{3, 4\}$ [9]. For all the center-surround pairs, cross-scale image differences (i.e., *center-surround* differences) are computed (the result is called as *contrast maps*). Computationally, the center-surround is realized by upsampling a coarser surround image to the finer center image and subtracting each other. Finally, the contrast maps are linearly combined to yield single topographical saliency map. Further details can be found in [9, 13]. The parameters used here are commonly accepted when using a visual saliency map, and the choice of them is decoupled from a particular tactile rendering algorithm and hardware. Since optimal parameters to further enhance tactile sensation are unknown, finding such parameters would be an interesting direction for future research.

This research contributes on improving the definition of visual features (e.g, color, luminance, and orientation) using CIE L*a*b* color space (in short, Lab space), a widely known perceptual color space [3]. One common challenge in using the saliency map is to find appropriate weights for linear combination of different contrast maps. One could use unit weights, but there is no guarantee that this is an optimal selection. To cope with this, the Lab color space wherein a Euclidean distance between points roughly scales with their perceived color difference is used. Instead of multiple feature maps, a single Lab image is defined as a feature map. This allows to evaluate the perceptual color difference at a single step without the linear combination issue more efficiently.

Also, the previous definition of the visual saliency along the *temporal* dimension is augmented. Since the saliency map was initially designed for static image analysis, it only deals with spatial dimension. Therefore, directly applying it to streaming images with dynamic scenes may not be appropriate. For instance, salient yet static objects may be less significant than dynamic objects in a scene. Hence, in order to preclude such static spots and track dynamic

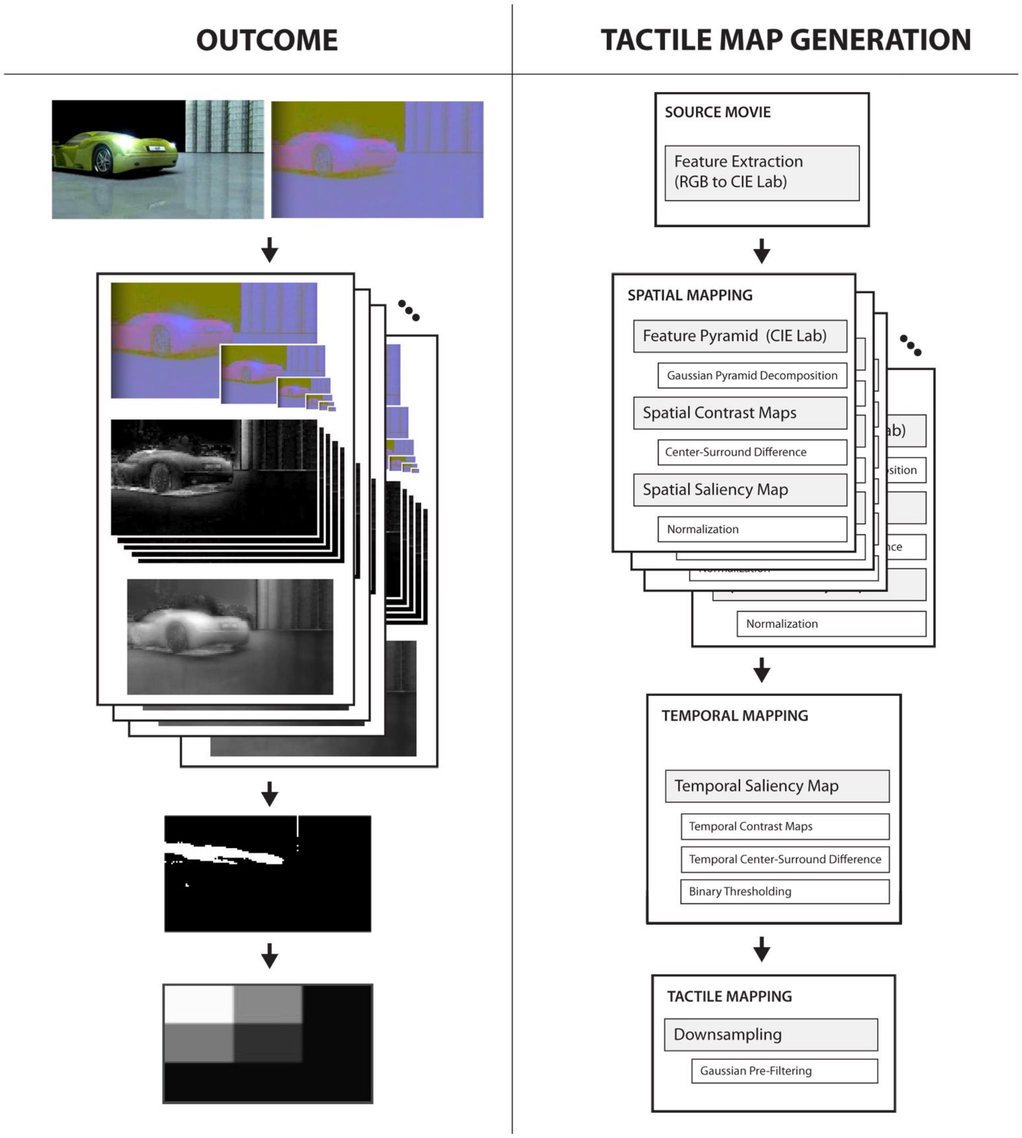


Figure 3.2 Overall pipeline of the visuotactile mapping algorithm.

motions, temporal saliency should be used. The principle is the same as that of the spatial saliency, but the *temporal* image pyramids are built along the time. In other words, a level-n image averages the spatial saliency maps of 2^n

previous frames including the current frame. The temporal image pyramids have the same image size as the spatial saliency map has, and are built on the fly. The final temporal saliency map is computed using the *temporal* center-surround differences, whereas the temporal center levels $\{0, 1, 2\}$ are used instead of $\{2, 3, 4\}$.

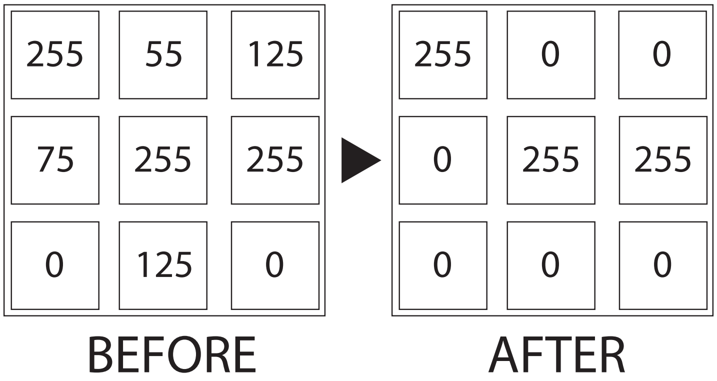
The resulting spatiotemporal saliency map was globally scaled using the nonlinear normalization operator as was done in [9], which uses the ratio between the mean of local and global maxima. While fine details of a saliency map promote the gaze analysis of a static image, they often inhibit the maximum saliency response in an image. To draw more focus on the most salient spots, it is more effective to discard excessive details. Thus, binary thresholding under a certain cutoff value is used. (in our case, 50 percentage of the maximum level) [19]

The next step is relating the final saliency map to tactile cues to actuate tactors. To abstract tactile display hardware, a tactile map such as a 3×3 array is used. The resolution of the saliency map is usually much higher than tactors, and hence, the mapping between the saliency map and tactile map needs to be defined. Therefore a simple linear downsampling with Gaussian prefiltering is used in the implementation, leaving room for better mapping that considers scene semantics and the expectation of user's volitional factors. In addition, a crispy feeling of tactile feedback can be generated by limiting the number of tactile cells actuated at once. These cells are selected based on the largest intensities and the unselected cells are changed to zero as shown in Fig 3.3. For this research, I selected three cells with largest intensities.

3.3 Actuation Design

The hardware implementation is performed for stimulating tactile feedbacks from the tactile map acquired from extracted tactile map using visual saliency.

There were a number of options to choose actuators for picturing tactile feed-



255	55	125
75	255	255
0	125	0

BEFORE

255	0	0
0	255	255
0	0	0

AFTER

Figure 3.3 Cell selection method to generate a crispy tactile feedback

back onto the human body. Widely known tactile actuators include linear resonance actuators (LRA) and eccentric rotated motors (ERM). Two actuators have different characteristics that LRA's output acceleration is decided based on the vibrating frequency and its input voltage while ERM can be only manipulated by input voltages. In the meanwhile, ERM has advantages over LRA in cost and relatively bigger output intensity.

In order to make a tactile display as simple and affordable as possible, Coin-type Eccentric Rotated Motor (ERM) is used as vibratory motor with its radius of 18mm and the thickness of about 3.5mm. It has a voltage range of 0 Volts to 3.0 Volts and can produce more than 10,000 rpm at its maximum voltage, it also requires a circuit and signal amplifier to actuate. OP-AMP, OPA544T manufactured from Texas Instrument was used because it has a high current amplifying functionality. Therefore the Ni-DAQ Max library was first used to generate the actuating signals for vibratory motor, it is later amplified through the designed circuit. See also Fig. 3.4 for the circuit design for actuating multiple actuators at the same time.

While the vibratory motor can be used to stimulate the tactile feelings to user, various ways of placing motors in tactile display were suggested. One of

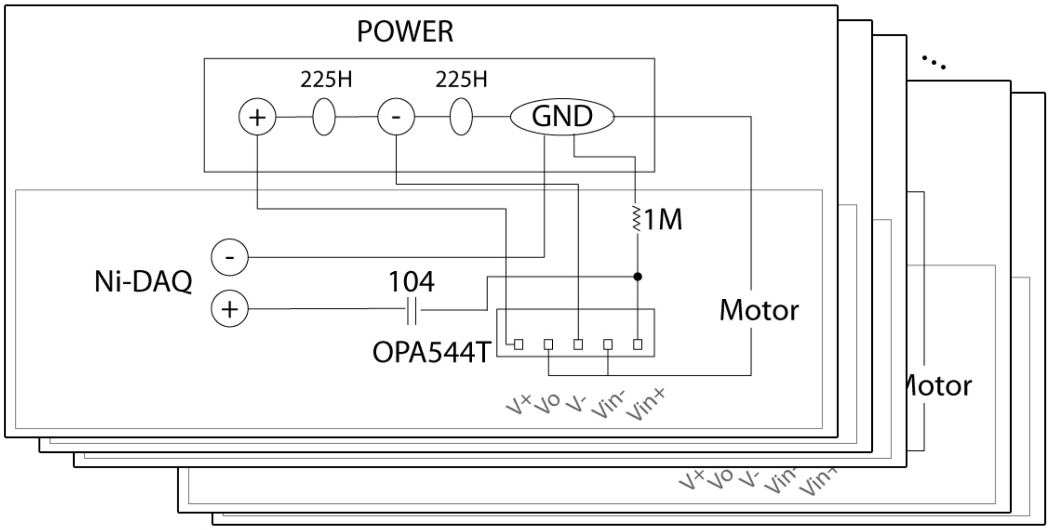


Figure 3.4 Overall circuit design for actuating multiple vibratory motors

raised dilemma during designing display is how actuators should be used. Because tactile feedback has to be transmitted through skin, tactile display must be in contact with subjects. However many previous studies showed not many people like wearing unnecessary clothes including vest, jacket or even gloves that it gives them an uncomfortable and unsatisfying environment. Therefore avoiding wearable environment became one of top objectives when implementing tactile display, non-wearable display was to be considered.

A notion of people spending their most of a day while seated leads that a chair can be a option for tactile display. A chair usually is structured from two surfaces of hip and back. Back surface has slight advantages over hip surface. Back may have wider flat surface compared to the hip surface and how image from monitor is shown to people can be geometrically equal to the surface of the back chair people touches. Therefore back cushion of chair is selected for tactile display and actuators are installed and contactable when people sit and lean back to the chair. Another problem raised after deciding where to put actuator is the best number of actuators for the display. Morrell and Wasilewski have installed tactors 60mm apart both horizontally and vertically on the back of the car seat for the vibrotactile seat system [15], Jones and Ray also used the tactor inter distance of 60mm for the vibration identification research [10] Based on the facts, placing three or five actuators in each row seems reasonable and three actuators are placed in each column leads to the tactile display with 3 x 3 or 3 x 5 size. In the meanwhile, people are only able to identify the limited number of tactile cues regardless of the resolution of the tactile display which suggest the maximum number of tactile cues in a given time should be controlled algorithmically.

Because many image analysis based algorithms are notorious for its complexity and high cost, it is necessary to analyse the computational performance of the haptic content generation algorithm. The system was implemented on an Intel i5 2.66 GHz with Intel OpenCV library. For most movie clips, up to the resolution of 1300×900, the system performed more than 30 FPS, the common requirement for real-time rendering. For HD resolutions such as 1080p, parallel GPU processing can be exploited on demand, as was done in [13].

At the following tactile rendering stages, the intensities of the tactile map are interpreted as the levels of vibration, and actuate the tactors whose dimensions are same as the tactile map. The actuation is performed in a continuous form, since the tactile map is also streaming along with the source video. The mapping between the tactile map and commands is straightforward, and thus, software commands for issuing haptic signals are virtually negligible. However, the mechanical latency takes longer than time for a single video frame, and thus, it requires to be compensated and this will be discussed in the next section.

Chapter 4

Method & Evaluation I

4.1 Tactile Rendering I

In order to test the saliency-based algorithm for visuotactile mapping, although it is independent of particular hardware platforms, a test platform for tactile display was built. The display is designed to provide vibrotactile stimuli onto the lower back of a user sitting on a chair. ERMs are installed on a chair in a 3×3 array. Each tactor is 10 cm apart from neighboring tactors and independently connected to a customized control circuit. The maximal voltage to actuate tactors is 3 V, which can provide the vibration intensity up to 49 G at a 77-Hz vibration frequency. When tactors are fastened on a solid chair, the tactor's vibration may be propagated onto neighboring tactors and be dislocated when they vibrate. To alleviate this problem, the tactors are installed in the tactor housings of a chair cushion cover made of thin nylon, and used the cover to wrap the sponge fabric of the chair cushion (see Fig. 4.1).

The procedure of tactile rendering is as follows. The tactile map for rendering reads tactile buffers which store the translated results of the video thread. Since visual and tactile renderings run at different threads with different refresh rates, latency from vibrotactile actuators requires suitable compensation to avoid confusions in visuotactile presentation. The lag is mainly caused by

Figure 4.1 Our test platform for tactile display.

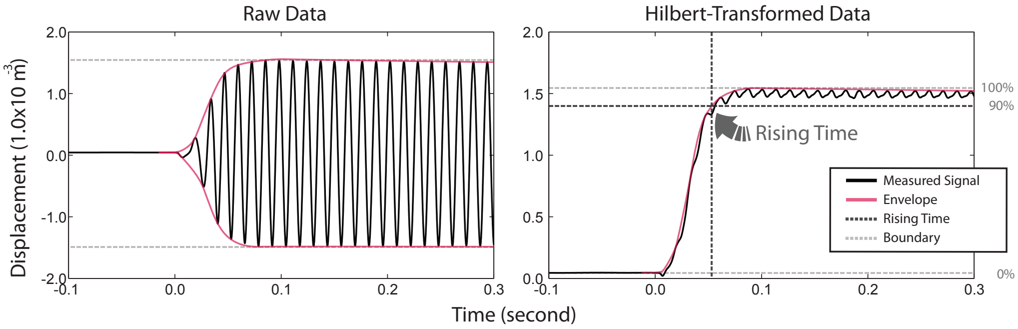


Figure 4.2 Example of a vibration intensity response for the definition of rising time.

the mechanical actuation delay of triggering vibration on the tactors. When a delay falls short of the duration required to play a single visual frame (e.g., 33 ms), the latency is in an acceptable range in practice and does not need to be compensated.

A simple remedy for the latency problem is pre-issuing tactile commands a few frames earlier than the corresponding visual signals. To do so, the latency values for a number of starting and target vibration voltages was measured. When a target voltage was smaller than a starting voltage, i.e., when the motor

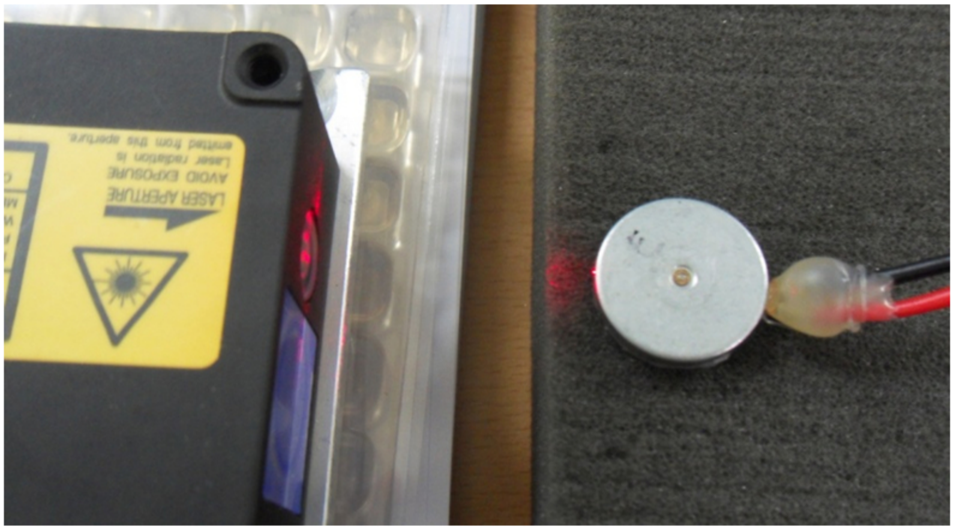


Figure 4.3 How the laser vibrometer is used to measure vibration

was decelerated, the maximum latencies in the acceptable range were observed and thus, decided not to consider those falling times further. However, the acceleration process required for the target voltage larger than the starting voltage showed a noticeably longer delay. Hence, the rising time of actuation as the time period required to reach the 90 percent of steady-state vibration level was defined, as shown in Fig. 4.2, and was compensated in tactile rendering.

The vibration intensity was noninvasively measured by looking at the tactor's vibration displacement using a laser vibrometer (SICK, model: AOD5-N1). During the measurements, the vibration motor was fixed on a flat sponge 30 mm next to the vibrometer. Starting voltages and offsets to the target voltages were sampled in the range from 0 to 3.0 V by a step size 0.1 V. The recorded data were fed to Hilbert transform to reconstruct their signal envelope for accurate amplitude estimation (Fig. 4.2).

For the rendering purpose, a function of parametric form is more convenient than interpolating the latency data in real time. Thus, the rising time data to

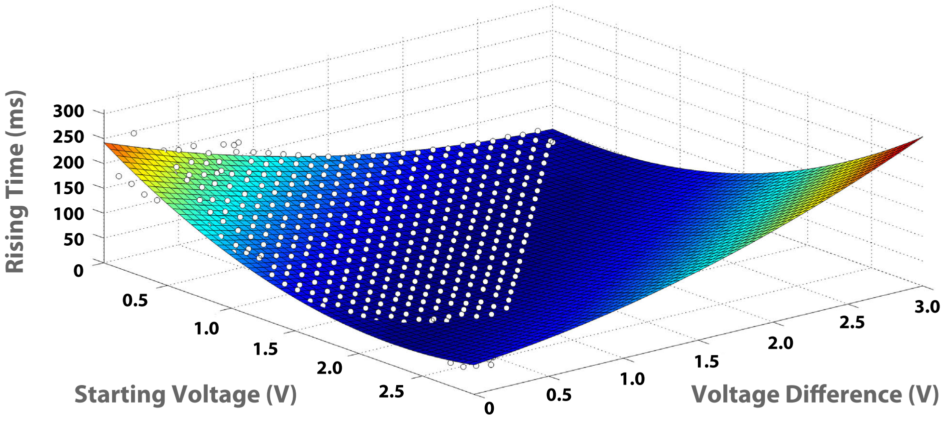


Figure 4.4 Example of vibration displacement measured using a laser vibrometer.

a quadratic form ($R^2 = 0.98$) was regressed as, such that:

$$f_r(V_s, V_d) = 241.9 - 175.7V_d - 117.7V_s + 38.86V_d^2 + 49.59V_dV_s + 18.07V_s^2, \quad (4.1)$$

where $f_r(V_s, V_d)$ is the estimated rising time, and V_s and V_d are the starting voltage and the voltage offset, respectively. In practice, it does not have to consider all the target voltages. Weak voltage commands less than a certain threshold (e.g., 0.5 V) can be discarded. By excluding such inputs, a set of rising times with 150 ms or less are obtained, in which the maximum resolution of tactile cues can be up to five video frames. Accordingly, a single data block for tactile signals is filled out of five video frames. For instance, when 60 ms is found as the rising time for the next tactile data block, the first three frames of the data block are written with the previous input voltage and the remaining two frames are filled with the target voltage at the next data block.

The estimation of falling time, $f_f(V_d, V_s)$, was found as follows.

$$f_f(V_d, V_s) = 4.111 + 9.328V_d - 13.52V_s - 3.588V_d^2 - 0.9188V_dV_s - 2.505V_s \quad (4.2)$$

where the R^2 was 0.82, $f_r(V_s, V_d)$ is the estimated rising time, and V_s and V_d are the starting voltage and the voltage offset, respectively. Falling time is shorter than the estimated rising time that most of the time required from starting voltage to target voltage falls within $33ms$ which is the time required for processing a single image. Therefore it is not necessary to consider compensating the falling time delay in tactile data buffer.

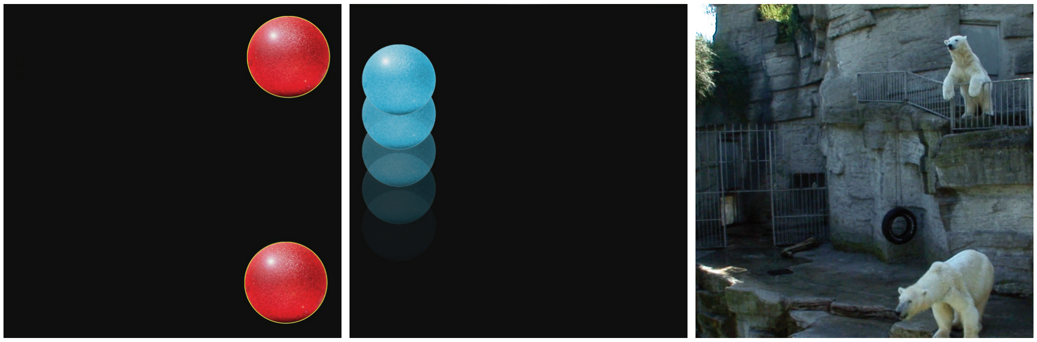
A concern about force discontinuity might arise here in issuing discrete force commands within the tactile data block. However, it does not manifest itself, since the low-bandwidth dynamics of the actuator is likely to interpolate the abrupt changes of the vibration intensity. Hence, this simple strategy can effectively compensate for the motor delay of vibrotactile rendering to synchronize the visual and tactile stimulations, while avoiding torque discontinuity.

4.2 Subjective Evaluation I

A user experiment was conducted to assess the usability of our system. Six usability items comparing visual-only and visuotactile presentations for different types of movies were collected via questionnaire. This section reports the methods used in our evaluation and experimental results.

Twelve paid undergraduate students (6 males and 6 females; 19–30 years old with average 22.3) participated in the experiment. The participants were asked to wear a thin T-shirt, leaning back on a chair to directly feel the vibration. Also, they wore earplugs and a headset to isolate them from tactor noises. They all clearly know the experiment procedure and have no problems in detecting vibrations in their backs.

The experiment used a two-factor within-subject design. The first factor was the presence of tactile cues while playing a movie. The other factor was the type of movies to be presented. Three movies, two synthetic and one natural, were used in the experiment (see Fig. 4.5 for each). The video resolution of 1000×1000 at 30 Hz was used commonly. One synthetic movie showed the static motions in that one or two balls repeatedly appeared at various locations



(a) Static motion (b) Dynamic motion (c) Real-world example

Figure 4.5 Examples of (a) static-motion movie (two balls appear in various locations), (b) dynamic-motion movie (a single ball moves in various speed), and (c) real-world movie (a documentary film of two bears in a zoo).

and remained at the same position more than one second. The other synthetic movie contained dynamic motions with a ball moving around at various speeds with sudden stalling motions. The last one was a real-world movie that shows bears in a zoo. By combination, each participant went through the total of six successive experimental sessions. Their presentation order was balanced using Latin squares.

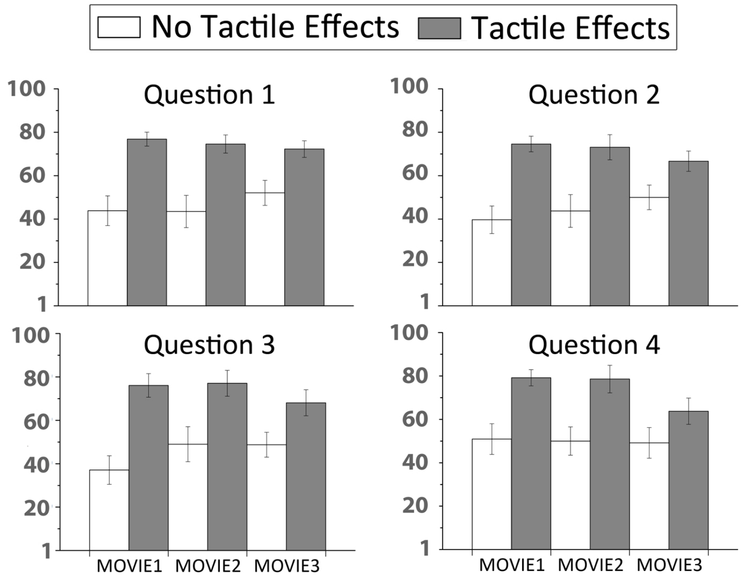
After each session, a break longer than two minutes was provided to the participant to fill out the questionnaire. The questionnaire consisted of six questions. Four questions were common to all the sessions, and the other two were tactile-specific questions given only in the conditions where tactile cues were presented. One additional survey asking a free evaluation of the overall system was followed. The common questions included: Q1. How interesting did you find the system? Q2. How much did you like the whole system? Q3. How much were you immersed in the movie with the given system? Q4. How well did you understand the contents of the movie? The tactile-specific questions were: QA. How well were the vibrations matched with the movie? QB. How much did the vibrations improve the immersion into the movie? Each question except the survey used a 100-point continuous scale, where 100 represents the strong agreement to the question, 50 a neutral opinion, and 1 the strongest disagreement.

4.3 Evaluation I Result & Discussion

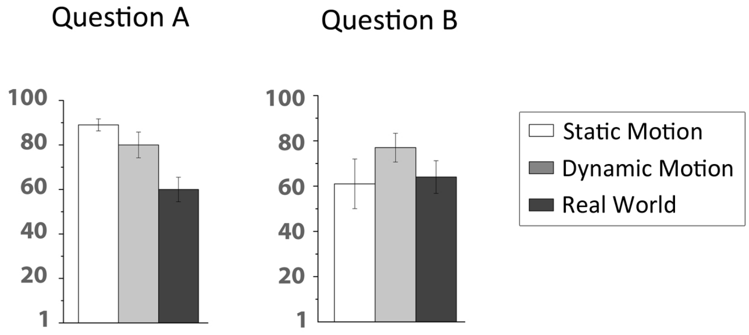
The subjective ratings of the participants obtained in the experiment are plotted with standard errors in Fig. 4.6a and Fig. 4.6b. Overall, the presence of tactile stimulation elicited much positive responses. The participant preferred the tactile-enabled movies to the original movies. Further, the tactile cue supports significantly enhanced the immersion and content delivery as well. I applied ANOVA to see the statistical significances of these differences between the visuotactile and visual-only presentations. Statistical significances were found for all the four common questions; all the p values were less than 0.001 and their F-statistics were $F_{1,11} = 39.95$, $F_{1,11} = 33.05$, $F_{1,11} = 30.87$, and $F_{1,11} = 21.79$, respectively. In contrast, none of the responses resulted in statistical significance for the factor of movie type. The tactile-specific question, QA, examined the system performance of the motion extraction; as expected, static-motion movie was evaluated as the most well matched (mean score 89, SE=2.68), while the real-world movie had the weakest performance (mean score 60, SE=5.48). The question on the absolute measurement of immersion, QB, showed that the real-world movie was the least favored over the other two movies.

The user study proved that the synchronized tactile display system improved the multimedia presentation; the participants were more engaged and immersed in the movie experience; tactile cues were effective in improving content delivery. The positive responses elucidated that the tactile cues were translated relatively well and it would emphasize the motions of salient spots in images.

More qualitative evaluation, regarding the overall system, was also collected via additional survey. The participants saw the system innovative and new, while the tactile cues are fairly well generated according to the salient spots



(a) Common Questionnaires



(b) Tactile Specific Questionnaires

Figure 4.6 Average subjective ratings measured in the user experiment.

of the movies. On the other hand, some of them saw the system abnormal or difficult to follow when they first tried it out. One suggestion for revising the rendering algorithm was limiting the duration of tactile cues, which makes them more memorable and avoids excessive tactile events as well.

As seen in the experimental results, the quality of translation in creating tactile cues is the key in conveying better experiences. The simple configurations of the synthetic movies were translated almost perfectly and preferred in terms of immersion with higher ratings, while the real-world movie was evaluated rather unorganized.

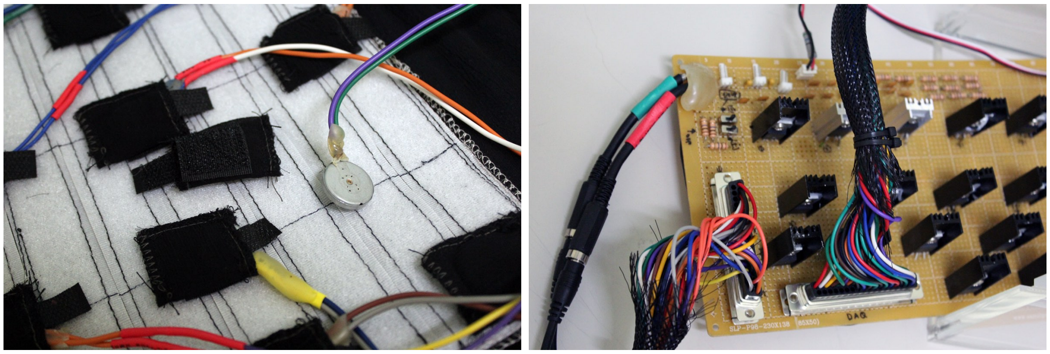
There is no doubt the tactile rendering method can enhance the theatre environment, participants responded tactile experience at the right place and crispy feeling are pleasing because they are very understandable and reasonable. On the other hand, excessive and wrong translation would be a natural consequence of the system because the algorithm does not have computations for the high-level semantics of objects presenting in the scene. When tactile feedbacks are generated in a wrong timeline, for instance, if the tactile effect occurs five seconds after the bomb exploded in the movie, the tactile feedbacks lose their purpose of enhancing the tactile experiment but rather bringing the confusion. Therefore it must generates the tactile effects with the visual effect in terms of the correct time frame. The algorithm was tested on both ball and bear movie, the quantitative result shows the ball has much higher scores than the bear's. The reason behind this result is that the ball's is much easier to understand than bear movie. Also the ball movie has one or two salient object to see at once while the bear movie has many potential salient objects including sun, bears, rocks and the salient background. Thus effects must be located in the geometrically same location from the visual scene. The algorithm may bring out the subjective attention of excitement or enjoyment while watching movie, it has a limitation of inaccuracy in both time and its location.

Chapter 5

Enhanced Tactile Rendering

5.1 New Approach of Tactile Rendering

The initial tactile back display uses a series of one of the most popular and inexpensive actuators, installed in the chair cushion as a 3×3 array format. However a 3×3 array tactile display did not give a rich sensation and should have a larger resolution to support crispier tactile feeling at specific location. For solving this problem, implementing a larger resolution system was necessary. Because many media contents these days have a larger horizontal size over the vertical size, known as wide screen media, it seems reasonable to place more actuators in a row than a column. Also the back is known to have a poor discriminating ability in vertical direction than horizontal direction. Therefore a 3×5 array tactile display was developed. Also because people's back are not wide enough that the new display cannot have a gap between actuators more than 6 cm. Actuators used are same as the previous tactile display system with its maximal input Voltage of 3 V. Also for expansionability, a new chair cushion cover was made. It is made of thin nylon cloth with Velcro hooks so ERM housings made with Velcro loops can be attached onto the cloth. Velcro hooks are very durable that it does not break even if more than 20 tactors actuating at the same time. However the cloth with Velcros and tactors' wires do not look



(a) Chair cover with Velcros (b) Circuit Design

Figure 5.1 Examples of (a) cushion cover cloth made with Velcros and (b) a proposed circuit design with 15 OP-Amps for actuating multiple ERMs.

pleasing so a cushion cover was used to wrap the cover. 5.1

Because the saliency map often carries too much information, binary thresholding was applied to create a filtered map discarding potential unnecessary noises in the tactile rendering I. However, the binary thresholding carries a couple of major problems. The best cutoff parameter for binary thresholding varies and the rejected information is not guaranteed as unnecessary. The image does not always use the entire spectrum of histogram; for instance, when a limited lower range spectrum of the histogram is used to represent the image, the intensity cells for entire image are changed as zero after the thresholding is applied. It removes the potential-meaningful saliency information unintentionally. The second major problem is that it is impossible to guarantee whether the saliency information after the filter is appropriate information when too much saliency information is presented. By using binary thresholding method, it may bring a misdirection that the potential important information disappears or the unnecessary information becomes emphasized so it clouds the appropriate necessary information. Therefore, the extraction can be improved by using other filtering techniques instead of binary thresholding and a regular downsampling procedure as shown in Fig. 5.2. To obtain a more distinguished salient map,

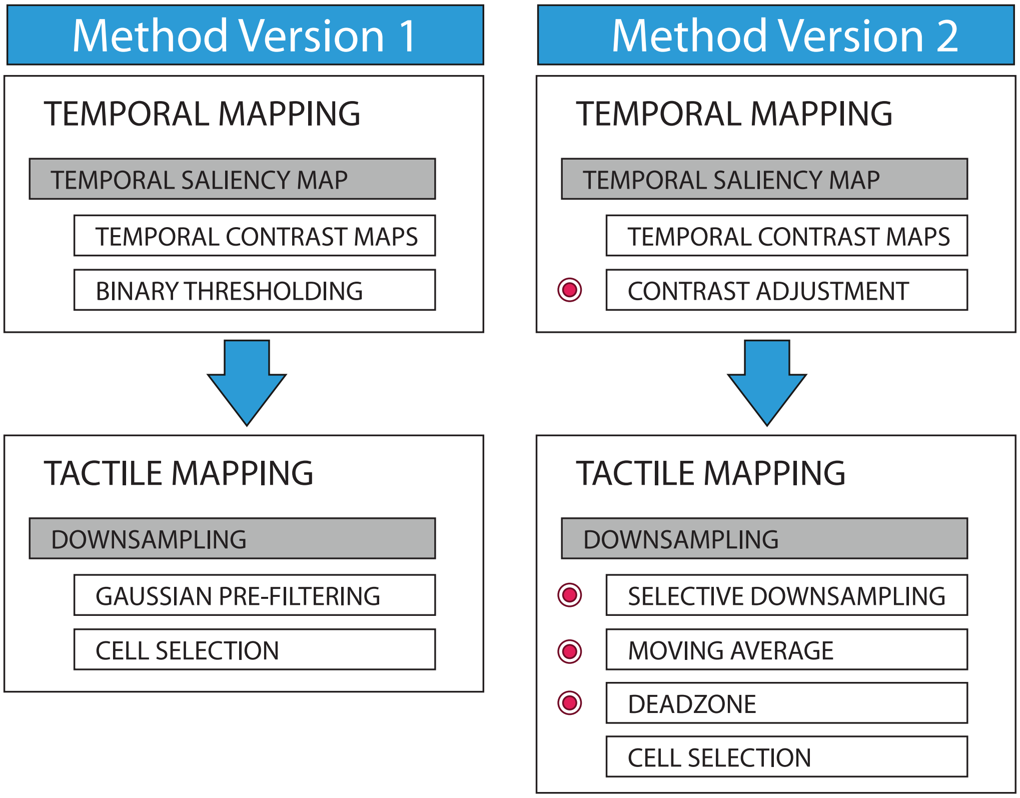
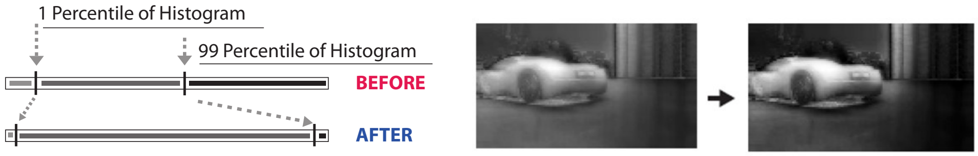


Figure 5.2 Tactile rendering methods of the version 1 and the version 2 are compared. Difference is marked with the red circle in Version 2.

another image processing technique is applied while a number of thresholding methods are used in the tactile mapping stage.

Because of the first version's problems, it is required to come up with a new approach that can deliver the saliency information more efficiently. Based on the fact that the audience-thought salient area is included in the saliency map most of the time, understanding how to extract it would help delivering the user-considered tactile effects. If an image color spectrum is spread over the entire histogram, it would lower the computation loads and more options for post-processing to result the meaningful saliency map are possible. Also ac-



(a) Contrast Adjustment (b) Application of contrast adjustment

Figure 5.3 The visual presentation of image contrast adjustment technique

quiring the contrasted image is important because it shows the precise salient information in an obvious way so the tactile map conversion becomes easier.

In order to satisfy conditions mentioned above for obtaining the precise salient information, the image contrast adjustment method commonly used in image processing field is applied. The image contrast adjustment method can be applied by analysing the histogram, each single percentile of the highest and lowest histogram is converted into the maximum intensity of the pixel, 255, and the minimum intensity, 0. The rest of 98 percentile of the histogram is stretched and their intensities are remapped from 1 to 254. This attempts to enhance the image contrast while it does not lose the original salient information of the image.

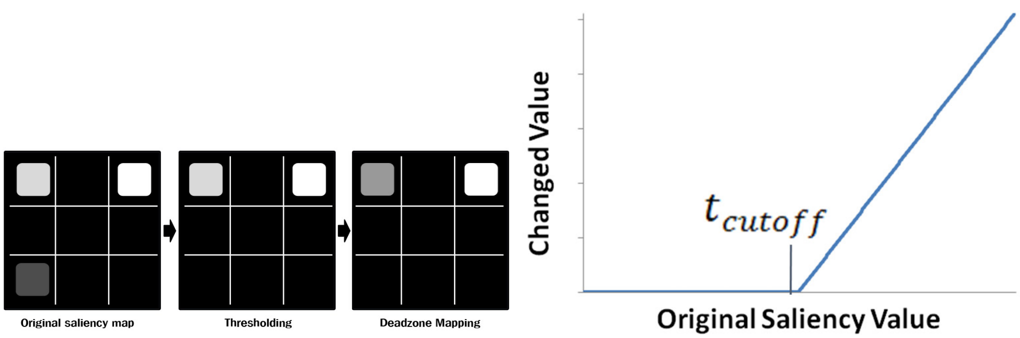
Once the salient information is emphasized by the image enhancement technique, it is necessary to convert it to the tactile-friendly map to control the tactile feedbacks to avoid the confusion of misdirected vibration. This can be simply implemented by analysing the global intensity of the salient information observed. For instance, when downsampling the saliency map to be used as tactile map, saliency map will be divided as the number of ERMs in tactile display. Then each of divided cells has corresponding pixels with its intensity. All the pixel intensities at each cell are added up and local maxima can be obtained. From the local maxima, the global maximum can be obtained. The saliency map is downsampled into the tactile map and the converted tactile map becomes zero if its global maximum is less than the certain threshold. Thus, the selective downsampling can be equated as, such that:

$$T = \begin{cases} 0 & \text{if } \mathrm{Max}(T) \leq t_{threshold} \\ T & \text{Otherwise} \end{cases}$$

where, tactile map $T = \{t_1 \ldots t_n\}$; n = the number of cells in the map; $t_i = i^{th}$ cell of the tactile map; $t_{threshold} = t_{max}$ x 0.1; maximum cell intensity, $t_{max} = 255$. This process is necessary due to two reasons that it can emphasize the salient area by eliminating the low salient areas and also let the image analyzed more easily in following tactile rendering stages.

The selective downsampling stage is considered as hard thresholding. Therefore a series of collected tactile maps after such thresholding, the unstabilized and time-incontinuous saliency map will be detected. People need enough duration of vibrations to realize its availability, but the tempered experience may bring the uncomfortable experience to the user; therefore, it is necessary to consider a way of smoothing the saliency shifts. Moving average filter is one solution that can be used in moderating rapid saliency area changes, the filter modulates the saliency map's intensity gradually and slowly changed so ERMs have enough time to be actuated and users would not have difficulty in understanding the tactile signals. Because the two threads are used between the main visual buffer and the tactile buffer, the tactile buffer holds future and past information for the specific visual buffer frame, it can create a non-causal system with the moving average filter.

Lastly, deadzone mapping is used to emphasize the salient information that it is applied to the processed tactile map using methods mentioned above. Fig. 5.4 With the pre-defined cutoff threshold, cell intensity of the processed tactile maps lower than a certain cutoff threshold are rejected from tactile rendering. While the cell intensities of higher than the cutoff threshold are stretched again from zero to the maximum intensity, 255. In addition, different cutoff threshold can result the user to have different tactile feelings; higher the cutoff threshold is, users feel crispier and shorter duration of tactile feelings. The best cutoff threshold is not defined therefore it will be found through the evaluation



(a) Example of filter application (b) Graphical representation of the filter

Figure 5.4 The visual presentation of image contrast adjustment technique

section. In addition, the cell selection stage is also included for generating a crispy tactile feedback as used in the first rendering method. 3.3

5.2 Subjective Evaluation II

A user experiment was conducted to assess the usability of our new system. In order to test the current algorithm's performance, the manually created tactile effects and the tactile effects created from the previous method were compared. Since there has never been a similar research, there is no strict guide line for who is capable of authoring manual tactile effects, but a person who has been doing similar authoring researches seems reasonable. One candidate had been selected for manual authoring that he has been working on researches of tactile feedback authoring based on auditory medium, with work experiences of tactile-related industry for more than two years. He was asked to create tactile effects for tested movies using the tactile effect authoring tool, initially developed by Kim et al., [11]. The six different methods included three approaches of the new method with different deadzone mapping's cutoff thresholds including a quarter of the maximum intensity, as method 1; a half of the maximum intensity, as method 2; and three quarters of the maximum intensity, as method 3.

(a) Movie 1

(b) Movie 2

(c) Movie 3

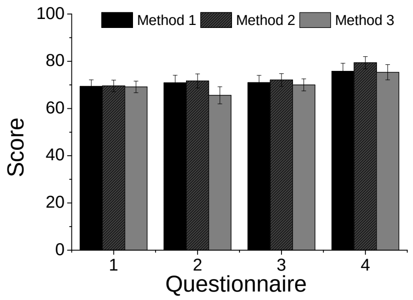
Figure 5.5 Three tested movies. Movie (a) is a combined video of initial static and dynamic videos used in the first subjective evaluation. Movie (b) is the dynamic real-world movie of racings. Movie (c) is the documentary style of real-world movie with dynamic scenes

For this experiment, parameters of 64, 128 and 192 were used due to its maximum intensity of pixel as 256. Also three additional methods were compared, including one approach of the initial algorithm, known as method 4, one approach of the manual authoring, method 5 and no tactile feedbacks as method 6.

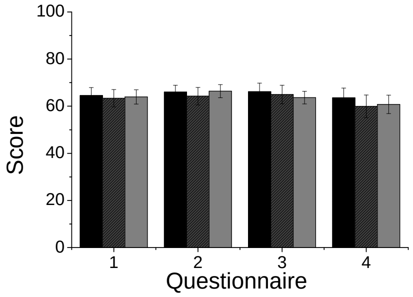
The experiment used a two-factor within-subject design, similar to the evaluation executed in the previous chapter. The first factor was the type of movies and the second factor was the method of the tactile cue generation while playing a movie. Three different types of movies and six different methods were used for comparison. Videos were chosen from totally different genres. One is the synthetic ball movie used in the previous algorithm (Fig. 5.5a), another is a racing movie of two racing cars driving on the track (Fig. 5.5b) and the other movie is the action genre with explosive scenes (Fig. 5.5c); their video resolutions are 1000x1000, 1200x720 and 1300x800 at 30Hz.

The first video of the synthetic ball is the combination of static motion and dynamic motion. The salient enough a red ball appears for one second and disappears. It appears again after one second but at different locations. It also includes the dynamic motion of a ball moving in different location with various speed. It contains motions of a simple linear moves in a vertical or diagonal direction as well as a circular motion. The ball in the video is very salient and easy for audiences to find a designer's intention and also the maximum number of balls appeared at each scene is two and the color of a ball is vibrant while the background of the video is black. The scene takes a place in the racing track in sunny day in the second video that the video has dramatical scenes of two cars racing each other as passing by the cameras, which allows the user to think as if he is actually in the track watching the game. There are a few salient areas besides racing cars, but generally the scene is not very complicated. The third video is the documentary genre that the video narrates the story from the cameraman's perspective. Audiences will understand the scene as the camera moves. Camera usually stays at a fixed point while it rotates either horizontally or vertically. The setting is at late night that there are only a few salient scenes in the entire video. One of the salient areas is found when explosion begins, gives audiences clear idea of salient areas moving in a linear direction of left to right. The last video may be viewed as most dynamic over other videos however it can give audiences a clear idea of where to look.

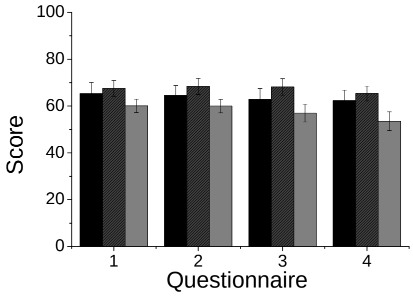
Twenty four paid undergraduate students (19 males and 5 females; 18–27 years old with average 22.0) participated in this experiment. The participants were asked to wear a thin T-shirt, leaning back on a chair to directly feel the vibrations. Also, they wore earplugs and a headset to isolate tactor noises. They all clearly understood the experiment procedure and had no problems in detecting vibrations in their backs. Videos were played 18 times in total and after each session, a break longer than one minute was provided to the participant to fill out the questionnaire. After all sessions, the participant was asked to do a survey of overall thoughts on this system. The questionnaire was made of four common questions and two tactile-specific questions. The common questions included: Q1. How much were you immersed in the movie with the given system? Q2. How much did you like the whole system? Q3. How interesting did you find the system? Q4. How well did you understand the contents of the movie? The tactile-specific questions include QA. How well were the vibrations matched with the movie? QB. How much did the vibrations improve the immersion into the movie? Common quetions used a 100-point continuous scale, where 100 represents the strong agreement to the question, 50 a neutral opinion, and 1 the strongest disagreement. For the tactile-specific questions, it also used a 100-point continuous scale but the subject was asked to answer points less than 50 if the tactile effects is negative to the system, on the other hand, points more than 50 could be given if the tactile effects were useful to the system.



(a) The result of movie 1



(b) The result of movie 2



(c) The result of movie 3

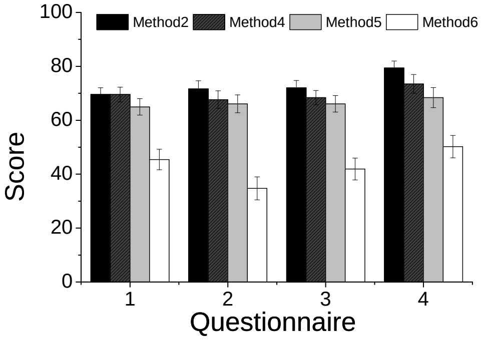
Figure 5.6 All three methods are the version 2 of the tactile rendering algorithm. While method 1 has a deadzone mapping cutoff threshold of 64, method 2 has 128 and method 3 has 192. The three methods are compared in movie 1, movie 2 and movie 3.

5.3 Subjective Evaluation II Result & Discussion

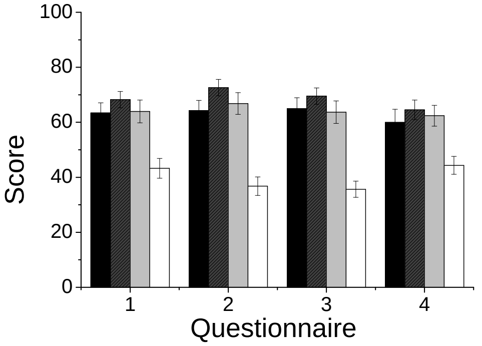
Before comparing tactile methods altogether, three newly suggested methods were analysed. Subjective responses for the newly suggested methods are plotted in Fig 5.6. Method 1, method 2 and method 3 are different from the parameter used in deadzone mapping. Although the result does not have a specific pattern or significant differences between each other, the method 2 elicited most positively responded over the other methods. While it has a better score in Q4 for the movie 1 and movie 3, the method 2 only has a slightly lower score in movie 2, which is a totally reasonable result. Because of the characteristic of deadzone mapping, higher the parameter is, sharper the audience may feel, and vice versa. Since the method 2 had the most positive scores over newly suggested methods, it was picked to compare with other tactile methods.

Fig 5.7 shows the subjective evaluation of the newly suggested algorithm–method 2, previous algorithm–method 4, the manual authoring–method 5, and no tactile effect–method 6. Throughout the three movies, there was a significant difference between method 2,4,5 and method 6. This implies the subjects prefer having the tactile effects while watching the videos. On the other hand, method 2 has the highest score over method 4 and method 5 with the movie 1.

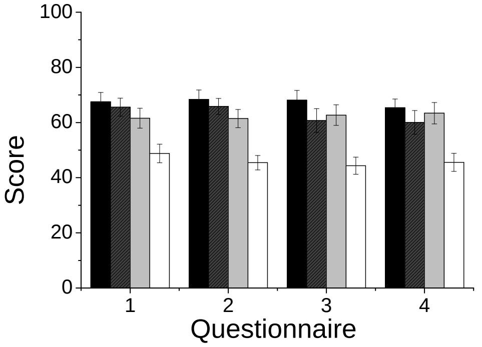
The movie 2 results differently than movie 1's. The method 4 has the highest score mean over other methods. For better explanations, nature of the movie and algorithm should be discussed. The first rendering algorithm has disadvantages when there are many salient areas or bright background in the scene, massive and uninformative tactile effects were generated without specific intention, confusing audiences from too much information. The video 2 is unlike other test videos, it takes scenes at the racing track. It seems uninformative and continuous tactile effect generations with absence of exact spatial information somehow match with the given racing film, allowing the user to feel as if they are at the racing track. This result derives that with certain situation, audiences may just want immersions with the movie but less care about the exact spatial information. According to the movie 3, three tactile methods are not significantly different but the method 2 has a higher mean than other methods. Movie 3 is another complicated video with a number of dramatic areas



(a) The result of movie 1



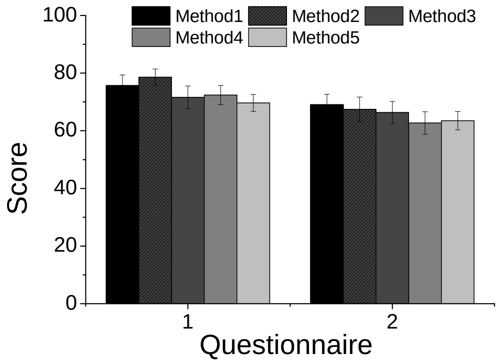
(b) The result of movie 2



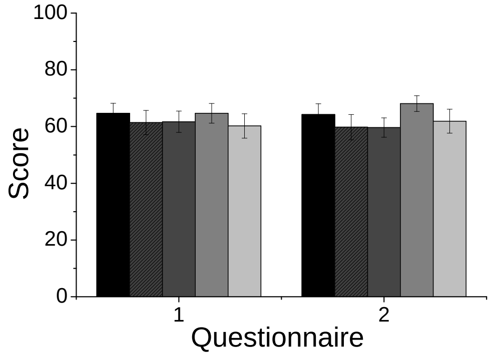
(c) The result of movie 3

Figure 5.7 The result compares all feedbacks of the new algorithm–method 2, the initial algorithm–method 4, the manual authoring–method 5, visual only–method 6 that may grab audiences' attention during the play. However, the most salient area is when the explosive scene takes into place, offers an audience the salient impact within the short period of time. The result implies audiences probably prefer method 2 because it generates tactile effects at the right time frame of salient events

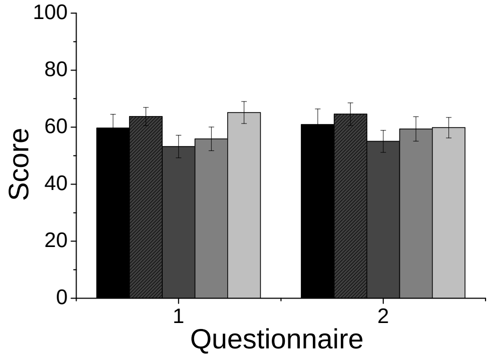
To support more on this finding, tactile specific questions were asked as 5.8 The result is very alike from the common questionnaire found in Fig. 5.6 and Fig. 5.7. Over five tactile methods, the newly suggested methods have the highest scores over other methods in movie 1 which was found similar to Fig. 5.7a However Fig. 5.7b implies audiences sometimes prefer enough vibration feedbacks regardless of the presence of the spatial or temporal information. From this result, contents of the movie is another factor that finds out what type of vibration should be given to audiences. From Fig. 5.7c, the newly proposed method and the manual authoring method acquired the highest scores. This responses elucidated that the tactile cues were translated from visual scenes at the appropriate time frame and gives audiences harmonic sensations.



(a) The result of movie 1



(b) The result of movie 2



(c) The result of movie 3

Figure 5.8 The tactile only result. All methods except the visual only method are compared. It has a similar score tendency of the Fig. 5.7

Chapter 6

Discussion

6.1 Enhancing Tactile Feedback

The visuotactile translation can be enhanced by tweaking the obtained tactile map in addition to the newly suggested method. For instance, simple linear mapping between the visual and tactile viewports can be dynamically adapted. The study of cinematography commonly tries to place salient ones in the middle of the scene to draw attention. However, this may cause tactile feedback occurring in the same location for a long period, and generate an unpleasant concentrated feeling to the user. In addition, the periphery of the screen is often of less importance, and small movements in the center often require to be exaggerated. To consider this, the region of interests can be dynamically widen and narrowed down according to the present context, and the provision of tactile cues can be limited in a short period, solely for important shots in scene semantics.

As briefly mentioned in the beginning of this chapter, the cinematography often places salient objects in the middle of the scene. This may draw an attention from users but endless tactile feedbacks would be turned into unpleasant feelings. Therefore when the algorithm detects that tactile effects are generated in the middle for more than a few second, the map is literally zoomed in and its larger resolution are analysed. Acquiring a tactile map by zooming in the scene generally leads to an opportunity to have a tactile map with detailed information. Tactile effects using this detailed information can avoid the predicted feedback and generate the informative feedbacks.

While zoom-in method exists to acquire the detailed information of the saliency image, the cell penalized method also enhances the user feedback using the predicted information. The cell penalized method does not equally divide the saliency map and create the corresponding tactile map, but it tries not to pick the cell positioned in the middle by penalising the middle cells to have a fewer cells assigned. The corner has a larger cell space than the one in the very middle or the ones in the middle and outer area. By performing this operation, it gives a better chance for corner or outer cells to be picked. Although the penalization of inner cell is applied by reducing the size of the cells as half, the penalization ratio can be tentative.

More than enough tactile stimulus would not be widely preferred if the contents have no salient areas, especially the stimulus from back is what people usually did not have and the right amount of stimulus should be given to the user. As the cinematography technique mentioned above, stimulus from salient areas can be generated at the same location more than enough may bring the user uncomfortableness and boredom. One method to avoid this is to adjust the time of tactile stimulus generation. When the tactile cells have more than certain amount of salient intensity for a certain period of time, the cells intensities can be exponentially decreased regardless of the saliency they have in the future. For instance, when the very middle cell generates the maximum tactile stimulus of three voltages for more than five seconds, at the period of three second, the two-third of the maximum tactile stimulus will be only given to the user and at the period of five second, no stimulus will be provided.

6.2 Overall Discussion

From the two stages of tactile rendering and evaluation, the traditional manual tactile authoring process for 4D films can be replaceable with automated visuotactile feedback system. The proposed method could fulfil the purpose of offering physical effects when video is played, the algorithm's credence was evaluated through two evaluations. The first proposed method proved the tactile feedback's prospects on replacing the current authoring trend. The result showed with three different movies that the user preferred the automatic approach to the visual-only presentation that the tactile feedback's commercial success was already proven by the commercial 4D products. This attempt submitted insights that the automatic approach can offer the user to experience visuotactile immersivenes,s which is the research's objective.

The first proposed method had a weakness in extracting the salient features when distinguishing from backgrounds, therefore meaningless vibration regardless of saliency information were often provided. This was presented because the algorithm targets at simplifying the saliency map unlike other visual saliency algorithms with multiple features. Although it loses opportunities to look at the image from many perspectives, the fact of simplicity can still offer the possibility of real-time visuotactile feedback. The user should be given visual and tactile related feedback in terms of spatial and temporal manner for enjoying the synchronized system. To reduce the confusion from visuotactile feedback, a new way of extracting the tactile map in automated manner was proposed and this was narrated in the chapter 5. The newly proposed method included a few tactile map moderating stages that it enhances the tactile map's spatial and temporal information, thus the new method provided the time-correct and space-dedicated tactile effects than before. However spatially and temporally correct tactile cue generation was very sensitive to the contents of the movie requiring the movie-sensitive parameters for filtering out the unnecessary tactile cues to be set. The second proposed method was investigated through a similar experiment found from the chapter 4. It compared with the first proposed method and the manual authoring approach. The result between all methods together indicated marketability of the automatic approach since the manual tactile authoring offers what people wanted to have from the tactile feedback. Although the overall result between tactile generation methods is not statistically different but the second approach has positive responses than the first and manual tactile authoring. This hopeful result shows the automated approach with a few minor revisions will result statistically better than the manual authoring in the near future.

The participants responded at the after-experiment survey session stating the movie can be another important factor for the tactile effects. The synthetic movie is easier to understand the designer's intention of saliency information, other real-world movies are more tricky to figure out. For instance, the tactile effect helped them understanding the contents of the second movie of the experiment II, the racing movie, participants enjoyed the tactile effects offered at the dramatical occasion such as the third movie. However some responded visuotactile feedback did not enhance the interests of watching video too much and occasional visually unmatched tactile effects from the automated approach troubled some of them from understanding the contents. Also some preferred the physical effects generated before the actual time frame to give a head up to the participants, feedback must be smooth in terms of time, and they must be offered at the exact time frame.

Although the subjective evaluation showed the upbeat response for automated tactile generation approach, the entire process was designed too complicated than supposed to be. A number of thresholding stages were used for rendering, some of them could be joined as one. For instance, the selective downsampling was used as hard thresholding and another thresholding was used in the deadzone mapping stage. They can be combined together and still carry out the similar result. Participants also suggested that the future evaluation should include the auditory effects in addition to the traditional evaluation process; many of them believed tactile and auditory feedback are highly related and can be a good factor for deciding the tactile feedback's usefulness towards the traditional visuoauditory system.

On the other hand, excessive and wrong translation would be a natural consequence of our system because the entire algorithm does not hold the high-level semantics of objects present in the scene. Without such scene semantics, the translation is hardly perfect. One good way to improve this problem is combining automated processing and manual authoring in the postprocessing. In the automated step, it aims at generating rather excessive tactile cues to some extent without filtering. Given tactile cues automatically generated from the visual information, the tactile content designer tries to prune out redundant cues to provide more focused feedback or to add some missing semantics during the translation. Also, diverse tastes from different cultures and favors can be incorporated in this postprocessing stage. I envision in the near future the tactile authoring will be automated in this fashion to provide sophisticated tactile cues.

Chapter 7

Conclusion

I presented an automated framework of tactile effect authoring to provide synchronized visuotactile stimuli in real time. The visual saliency served as a basis for extracting spatiotemporal importance from existing visual media and translating the visual importance to tactile cues. Vibrotactors installed on a chair were used to render tactile effects synchronously, along with the compensation of vibrotactile latency. The user study found that visuotactile rendering were preferred to visual-only presentation, eliciting more immersion and involvement, and better understandings of content.

Since the proposed method is independent from particular rendering methods, the approach has special importance for haptic content creators and interaction designers, who strive to create online or offline physical contents inducing spatially-present experience. Haptic cues extracted automatically from the existing media can facilitate the rapid manipulation of a tactile movie in the postproduction stage. In the future, I am planning to include more sophisticated supports for the semi-automatic postproduction.

요 약 문

영상주목도 기반 진동 효과 자동 저작

3D효과에 물리적효과가 추가 된 4D영상은 4D영화관이나 써드파티등의 하드웨어를 통해서 손쉽게 감상할 수 있다. 영상을 감상하면서 시각 및 촉각적인 느낌이 동시에 제공 되는 경우 정보전달 및 흥미유발에 도움을 주지만 그 제작에 있어 많은 노력과 시간이 필요하다.

본 논문에서는 동영상 내 주변보다 주목성을 갖는 두드러진 대상에 대하여 공간적 및 시간적 주목도를 이용하여 영상의 시각적 주목도를 자동으로 만들고 이를 바탕으로 촉각적 주목도를 완성하여 촉각 자극을 주며, 이렇게 만든 촉각 자극과 주어진 하드웨어의 성능적 한계를 인식하여 인지적 조화에 문제가 없이 시촉감각적 자극을 동시에 제공하는 알고리즘을 제안한다.

시각적주목도를 바탕으로 이진역치방법과 다운샘플링을 통해서 구해진 촉감렌더링 방법은 가장 기본이 되는 공이 나타났다 사라짐을 반복하는 영상과 공의 영상에 선형적 동작이 추가 된 영상, 그리고 실제 영상으로 동물원에서 곰이 나타나는 것을 촬영한 영상, 총 세 가지 영상에 대한 사용자평가실험을 진행했다. 시촉감각적 자극을 동시에 제공하기 위해 특별 제작된 햅틱 의자를 이용하였다. 하드웨어의 특성인 구동지연시간때문에 시각과 동기화가 어려운 점을 고려하여 하드웨어의 성격을 재어 하드웨어 지연시간 보상 알고리즘을 구성해주었다. 또한 사용자에게 보다 뚜렷한 효과를 주기 위해 구한 촉감지도 중 그 크기가 가장 큰 세 가지의 셀을 골라 해당하는 진동을 제공해주었다. 이렇게 구성한 알고리즘을 바탕으로 변환된 촉각 주목도 지도들과 함께 촉각 효과를 사용자에게 실시간으로 전할 수 있다. 사용자평가

실험 결과 비촉감렌더링 방법보다 몰입감을 증가시키고, 흥미도를 증가시키고, 선호도가 증가되고, 이해도 역시 증가한다고 나왔다. 하지만 마지막영상인 실제 영상에서의 시촉감각적 효과 제공은 알고리즘이 정확하게 시각효과를 촉각효과로 변환해주지 못하여 사용자에게 되려 혼돈을 주기도 하였다. 이에 따라 촉감렌더링방법 개선이 필요하게 되었다.

개선된 방법은 영상의 대비조정향상방법과 역치방법을 거쳐서 다운샘플링을 하여 촉감지도를 완성하였으며 연속적인 촉감효과를 제공하기 위해 이동평균필터를 적용하였으며 다시 데드존 과정을 통해 촉감효과를 강조하기 위해 주목도가 낮은 부분을 제거하여 사용자에게 촉감효과를 제공하였다. 개선된 알고리즘 또한 사용자에게 보다 뚜렷한 효과를 주기 위해 구한 촉감지도 중 그 크기가 가장 큰 세 가지의 셀을 골라 해당하는 진동을 제공해주었다. 마지막으로 촉감효과를 제공해주는 촉감의자의 진동 해상도를 높여 사용자에게 풍부한 느낌을 제공하였다. 개선된 알고리즘의 유용성을 알아보기 위해 사용자평가실험을 진행하였으며 이를 위해 이전 사용자평가실험에서 사용되었던 공 영상, 레이싱차들이 경주하고 있는 레이싱영상, 그리고 폭탄이 터져 폭발하는 극적인 영화영상 등 총 세가지 영상이 사용되었다. 비교한 방법으로는 개선한 방법, 첫번째 방법, 수동으로 촉각효과가 저작된 방법, 영상만 제공되는 방법 등 네 가지 방법이 사용되었으며 사용된 설문으로는 기존의 설문과 동일하며 사용자평가 실험 결과 개선된 알고리즘이 촉감효과 비제공방법과 비교하여 통계적으로도 유의하게 좋다는 응답을 얻었다. 하지만 기존의 알고리즘과 비교한 결과는 그 평균수치만 높을 뿐 통계적으로 유의하게 좋다는 결론을 얻지는 못했다.

4D효과 저작에는 시간이 많이 들며 이제는 손쉽게 사용자가 체험할 수 있으며 기존의 연구에서는 촉감효과를 저작하기 위해 저작툴을 만들어서 사용자에게 촉감효과를 제공했으나 본 연구에서는 촉감효과 저작에 걸리는 시간을 단축시키기 위하여 촉감효과를 자동으로 저작하여 사용자에게 제공한다는 점에서 의의가 있다.

Bibliography

[1] J. Cha, S.-y. Kim, Y.-s. Ho, and J. Ryu. 3D Video Player System with Haptic Interaction based on Depth Image-Based Representation. *IEEE Transactions on Consumer Electronics*, 52(2):477–484, May 2006.

[2] J. Cha, Y. Seo, Y. Kim, and J. Ryu. An Authoring/Editing Framework for Haptic Broadcasting: Passive Haptic Interactions using MPEG-4 BIFS. In *Eurohaptics*, pages 274–279. Ieee, Mar. 2007.

[3] CIE. *C.I.E: Recommendations on uniform color spacescolor-difference equations, psychometric color terms*. Supplement Publication No 2 to CIE Publication No 15 (E-1.3.1), 1978.

[4] C. Collins. Tactile Television-Mechanical and Electrical Image Projection. *IEEE TRANSACTIONS ON MAN-MACHINE SYSTEMS,*, MM(1):65–71, 1970.

[5] C. E. Connor, H. E. Egeth, and S. Yantis. Visual attention: bottom-up versus top-down. *Current biology : CB*, 14(19):R850–R852, Oct. 2004.

[6] R. Desimone and J. Duncan. Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, 18(1):193–222, 1995.

[7] H. M. Hamed and R. A. Hema. The Application of the 4-D Movie Theater System in Egyptian Museums for Enhancing Cultural Tourism. *Journal of Tourism*, X(1):37–53, 2009.

[8] A. Israr and I. Poupyrev. Tactile Brush : Drawing on Skin with a Tactile Grid Display. In *CHI*, 2011.

[9] L. Itti, C. Koch, and E. Niebur. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, Nov. 1998.

[10] L. a. Jones and K. Ray. Localization and Pattern Recognition with Tactile Displays. In *2008 Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 33–39. Ieee, Mar. 2008.

[11] Y. Kim, J. Cha, J. Ryu, and I. Oakley. A Tactile Glove Design and Authoring System for Immersive Multimedia. *IEEE Multimedia*, 17(3):34–45, 2010.

[12] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry., Jan. 1985.

[13] S. Lee, G. J. Kim, and S. Choi. Real-time tracking of visually attended objects in virtual environments and its application to LOD. *IEEE transactions on visualization and computer graphics*, 15(1):6–19, 2009.

[14] P. Lemmens, F. Crompvoets, D. Brokken, J. V. D. Eerenbeemd, and G.-j. D. Vries. A body conforming tactile jacket to enrich movie viewing. In *Third Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 7–12, 2009.

[15] J. Morrell and M. Ieee. Design and evaluation of a vibrotactile seat to improve spatial awareness while driving. In *Symposium A Quarterly Journal In Modern Foreign Literatures*, pages 281–288, 2010.

[16] E. Oh, M. Lee, and S. Lee. How 4D Effects cause different types of Presence experience ? In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pages 375–378, 2011.

[17] D. Parkhurst, K. Law, and E. Niebur. Modeling the role of salience in the allocation of overt visual attention. *Vision research*, 42(1):107–23, Jan. 2002.

[18] Treisman, Anne and G. Gelade. A feature-integrationtheory of attention. *Cognitive Psychology*, 12(1):97–136, 1980.

[19] D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch. Attentional Selection for Object Recognition A gentle Way. In *Biologically Motivated Computer Vision*, pages 472–479, 2002.

[20] K. Yoon, B. Choi, E.-s. Lee, and T.-b. Lim. 4-D Broadcasting with MPEG-V. In *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop*, number Part 2, pages 257–262, 2010.

감 사 의 글

제가 처음 포항에 온 날이 아직도 생생히 생각이 납니다. 2009년 여름에 대학원 시험을 보고 교수님께서 감사하게도 저를 받아주셔서 햅틱스 및 가상현실 연구실에 입학하게 된 후 앞으로 같이 지내게 될 연구실 선배님들께 잘 보이려고 정장을 입고 인사를 드렸던 점이 엊그제 같은데 벌써 입학 후 2년 6개월이란 시간이 흘렀습니다. 저는 지난 30개월동안 포항에서 생활을 하며 정말 많은 일들이 생겼습니다. 먼저 제 삶에서 가장 큰 일을 겪어 제 일에 지장이 생겼을때에도 그점을 십분 이해해 주신 교수님, 여러 선배님들, 동기분들 그리고 후배님들 정말 죄송하고 감사합니다. 그리고 제가 연구를 하면서 헤매고 있어도 이해해주시고 올바른 길로 다시 갈 수 있도록 지도 해주신 교수님 감사합니다. VT팀의 맏형 인욱이형, 제 직속 선배님으로 항상 고민이나 문제가 있을 때 같이 고민해주셔서 감사합니다. 108호에서 노래부르기 그리울것 같습니다. 지금은 중국에서 베이징오리와 딤섬을 먹고 있을 건혁이형, 항상 1% 부족한 제게 귀찮아하지 않고 좋은 지도를 해주셔서 감사합니다. 줄곧 내 옆자리 종만아, 너도 조만간 멋진 그녀가 너 앞에 나타날테니 너무 걱정하지 말고 지금처럼 당당하게 지내. 용재야, 집에서 떠나 포항에서 공부하는 여러 연구실사람들 건강챙겨줘서 고맙고 앞으로도 부탁할게. 최근 다시 연구실로 들어온 재봉이형, 형의 지치지 않는 체력과 끊임없는 자기동기조절 저도 배우고 싶습니다. FF팀의 성훈이형, 연구실 가장 맏형으로써 꼼꼼한 성격으로 연구실 사람들을 지도해주셔서 감사합니다. 1년전에 비해 훨씬 잘생겨진 인이형, 자기일도 곧잘 미루고 저를 포함한 우리 후배들 항상 도와주셔서 감사합니다. 연구실에 영어를 생활화 시켜준 Reza, 사회 경험을 바탕으로 우리 연구실에 도움을 주어서 감사합니다. 신문에도 나온 공부도사 성환아, 그동안 여러가지로 도와주어서 고맙고 나중에 공부잘하는 방법좀 알려주렴. 연구실 툴툴이 겸 대학원 동기 호진아, 항상 너와 그 번뜩이는 아이디어를

보며 나는 많은 반성을 하고 많이 배운단다. 이제 한달후면 파란색으로 피가 물들 연구실 지킴이 경표형, 그동안 항상 같은 배를 타고 같이 활동하고 도움을 많이 받았는데 정말 감사합니다. 새벽에 문앞에서 마신 다방커피 한잔 잊지 못할겁니다.

그리고 제 대학원 과정내내 제 큰형처럼 바쁘신 와중에도 좋은 지도를 해주시고 많은 영향을 주셨던 성길이형 정말 감사합니다. 또한 제 디펜스에서 좋은 지도를 해주신 이승용교수님 감사합니다. 연구한다고 많이 놀아주지 못해도 툴툴거렸지만 항상 옆에서 이해하려고 해주는 비진아 고맙다. 그밖에 연구실에서 먼저 졸업하신 선배님들, 저를 항상 걱정해주신 강수형 아주머니, 친척분들 감사합니다. 위에 쓴 글을 포함하여 저를 생각해주신 모든분들 덕분에 석사학위디펜스를 치루고 이렇게 석사논문을 인쇄하게 되었습니다. 저를 여기까지 오게끔 도와주신 모든 분들 정말 감사합니다.

끝으로 큰일을 함께 겪고 같이 울고 위로했었던 작은누나 큰누나 그리고 큰 매형, 항상 제가 정신적으로나 외적으로 힘들어할때 도와주시고 같이 일들을 헤쳐왔습니다. 삼남매가 똘똘 뭉쳐서 이 세상 살아가자는 그 말 꼭 지키고 이제는 그동안의 힘든일에서 벗어나 좋은일들만 겪고 웃으며 살아요. 고맙습니다. 마지막으로 그동안 저를 키워주신 아버지, 어머니, 제가 성공해서 호강 시켜드리려 했는데 그러지 못해서 정말 죄송합니다. 지금은 저를 마냥 지켜만보시고 응원해주시겠지만 아직도 눈을 감으면 부모님이 그리워 생각이 납니다. 저를 앞으로도 지켜봐주세요. 그동안 키워주셔서 감사합니다. 사랑합니다 아버지 어머니.

Curriculum Vitae

Name : Myongchan Kim

Education

2005 - 2009 : Bachelor in Information Technology, Carleton University- Interactive Multimedia and Design

2010 - 2012 : M.S. in Computer Science and Engineering, POSTECH
Thesis Title :
영상주목도 기반 진동 효과 자동 저작 (Saliency-driven Real-time Tactile Effects Authoring)
Advisor: Prof. Seungmoon Choi

Publications

International Journals

1. **Myongchan Kim**, Sungkil Lee, and Seungmoon Choi, "Saliency-driven tactile effect authoring for real-time visuotactile feedback," Eurohaptics, Tampere, Finland, 2012.
2. Hyun K. Kim, **Myongchan Kim**, Wongi Hong, Sung H. Han "Touch & Drag: A technique for placing a cursor on small touch screens", IIE Asian 2011, Shanghai, China, pp. 261-266, 2011

Patent

1. 최승문, **김명찬**, 이성길, "촉각 효과의 제공 방법 및 장치", 출원번호 10-2012-0064031, (Method and apparatus for providing of tactile effect)

Presentations

1. **Myongchan Kim**, Sungkil Lee, and Seungmoon Choi, "Smart Haptic Chair: Automatic Haptic Feedback Authoring System Based on Visual Saliency," Demonstrated in the IEEE Haptics Symposium, 2012.
2. **Myongchan Kim**, Sungkil Lee, and Seungmoon Choi, "Generating immersive environment with tactile feedback for learning English," Presented at Korean Haptics Workshop, 2012