

Saliency-Driven Real-Time Video-to-Tactile Translation

Myongchan Kim, Sungkil Lee, and Seungmoon Choi, *Member, IEEE*

Abstract—Tactile feedback coordinated with visual stimuli has proven its worth in mediating immersive multimodal experiences, yet its authoring has relied on content artists. This article presents a fully automated framework of generating tactile cues from streaming images to provide synchronized visuotactile stimuli in real time. The spatiotemporal features of video images are analyzed on the basis of visual saliency and then mapped into the tactile cues that are rendered on tactors installed on a chair. We also conducted two user experiments for performance evaluation. The first experiment investigated the effects of visuotactile rendering against visual-only rendering, demonstrating that the visuotactile rendering improved the movie watching experience to be more interesting, immersive, and understandable. The second experiment was performed to compare the effectiveness of authoring methods and found that the automated authoring approach, used with care, can produce plausible tactile effects similar in quality to manual authoring.

Index Terms—Tactile Effect, Authoring, Visual Saliency, 4D Film, Multimedia

1 INTRODUCTION AND BACKGROUND

TACTILE feedback has proven its worth as an effective communicative source in diverse applications [1]. While the sole use of tactile display has identified its value in many applications, tactile signals in association with visual and auditory channels can facilitate improved physical experiences in terms not only of immersion and entertainment, but also of better content delivery. In this regard, new-generation media such as 4D films are one of the successful attempts, which extends the conventional audiovisual interaction to encompass vibration, breeze, smell, mist, or tickler [2, 3].

Creating a haptic film requires the haptic content that is synchronized with the semantics of the audiovisual content; otherwise, it can cause confusion in understanding the director's intention. For this reason, the vast majority of haptic content authoring still relies on content designers, requiring considerable effort for the coordination with other modalities. For instance, as early as the 1970s, a black-and-white picture was converted to tactile cues for a chair-mounted tactile simulator [4]. Recently, Kim et al. developed a manual line-drawing interface that aids in designing tactile motion segments in a video for their tactile

glove system equipped with a number of vibrotactile actuators [5]. Intuitive GUIs are helpful for pre-encoding haptic stimuli for an array of tactors [5, 6].

Manual authoring may lead to the best quality, but it is laborious and time-consuming. If there exist computational models that can *automatically* create haptic content out of audio/visual signals, the authoring of haptic content becomes easier and more efficient to a great extent. However, most of real-world content is directly captured without recognizing its source, and this poses a great challenge in new haptic media creation. Nonetheless, some recent algorithms have been successful for automatic audio-to-tactile conversion, e.g., for music [7, 8, 9] and games [10].

Since sound is also included in movies or TV content, audio-to-tactile translation is also a promising approach for films [10]. However, video delivers a much more amount of information with different kinds, and the information contained in video and sound is not always in agreement. Therefore, automatic authoring of tactile effects based on visual analysis is indispensable for films.

Early attempts of visual-to-tactile translation have relied on the semi-automatic mapping between visual objects/events and haptic effects [11]. However, difficulties in automatic scene analysis led them to use domain-specific manual annotation (e.g., football games [12, 13]) or explicit transformation of 2D vector graphics [14]. This inspired us to explore a *fully automated* approach for haptic film authoring in accordance with visual media. Our work is distinguished from the previous ones for its direct video-to-tactile translation without offline annotation or analysis, which allows us to realize *real-time* haptic feedback systems for *general* domains.

Our approach derives tactile cues from *visually at-*

- M. Kim and S. Choi are with Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea.
E-mail: {billkim, choism}@postech.ac.kr
- S. Lee is with Sungkyunkwan University, Suwon, Republic of Korea.
E-mail: sungkil@skku.edu

Manuscript received 29 Mar. 2013; revised 28 Oct. 2013; accepted 10 Nov. 2013; published online XX Nov. 2013.

Recommended for acceptance by A.B. CCC.

For information on obtaining reprints of this article, please send e-mail to: toh@computer.org, and reference IEEECS Log Number TH-2013-03-0032.

Digital Object Identifier no. 10.1109/TOH.2013.X.

tentive spots to achieve positive synergy from the synchronized visuotactile rendering. This is motivated by the previous studies that demonstrated a potential of tactile feedback in inducing visual attention. Spatially-directed tactile cues can lead to a rapid shift of visual attention without the interference of other modalities, owing to the existence of robust crossmodal links between vision and touch [15]. Research has shown that spatially-distributed tactile stimulation is effective at notifying the location of important visual events [16]. This principle is instrumental in designing haptic warning signals for visually-overloaded driving situations [17, 18, 19]. Our hypothesis underlying this work is that visual attentive spots are likely to include important events for the movie and tactile stimuli that emphasize the attentive spots can amplify the user's immersion to the events, thereby improving the general experience of movie watching.

Technically, the generation of tactile cues from attentive visual objects is still challenging, without being aware of the semantics and spatiotemporal structure of a scene. Attentional allocation of vision generally involves both bottom-up (feature-driven) saliency and top-down volitional factors [20]. However, the top-down information is unavailable in most cases of visual media. Hence, in the absence of particular contextual information, a key aspect to extract tactile stimuli from visual information can be *visual saliency*, relying solely on visual features and their spatiotemporal structure [21]. For instance, inhomogeneous structures of color, brightness, and edge were reported to be significant in visual saliency perception [22]. In particular, the computational autonomy has been enabled by the *feature integration theory*, one of the most influential theories on bottom-up visual perception [23, 24]. Based on this theory, the streaming media can be processed spatiotemporally, and the corresponding saliency map can be generated. This is the foundation of our work.

This article presents our research on algorithms to produce synchronized visuotactile effects by transforming streaming visual signals to tactile cues using visual saliency and a real-time tactile display built upon an array of tactors installed on a chair. In our previous work [25], we have developed algorithms of extracting visual saliency from streaming video images, mapping the visual saliency to tactile signals, and rendering the tactile signals, and we also reported a user experiment performed to evaluate the usability of our system. In the present work, we extend the previous approach with a new adaptive strategy to better transform visually important spots to the tactile display. Also, we report a new user experiment performed to compare the different visuotactile authoring methods including manual authoring.

In addition, our study uses low-cost vibration motors for tactile rendering. This choice greatly improves the practicability of our system, but it comes with

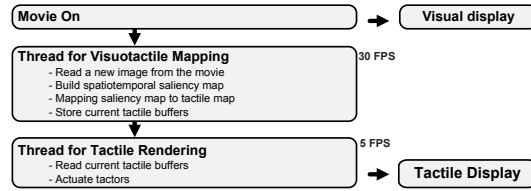


Fig. 1: Flow of our system. Two threads for visual and tactile rendering run simultaneously but at different frame rates.

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.

Our system is aimed at a real-time interaction system unlike the previous research, geared toward the great benefit of distributing haptic content without manual pre-encoding. To our knowledge, this is the first attempt for an automatic, real-time tactile effect authoring system making use of movies. In addition, it has a potential advantage for human-aided design. Initial content can be prototyped rapidly by our system, and then designers can take it over and enhance the tactile scenes, lowering the production cost to a great extent. This is a viable alternative considering no semantics is taken into account in our system.

2 OVERVIEW OF FRAMEWORK

In this section, we provide a brief perspective on our system. Fig. 1 illustrates the rendering pipeline of our framework. In the system, visual and tactile rendering are asynchronously executed using two different threads. For every frame, the thread for visual rendering runs as usual (e.g., at 30 Hz). In the meanwhile, the thread also builds the saliency map that spatiotemporally abstracts perceptual importance in a visual scene. The resulting saliency map is translated and mapped into tactile buffers, which have the resolution 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 the tactors. In particular, the tactile commands are issued ahead of the visual commands to compensate for the actuator latency. Each step is detailed in the following sections.

3 SALIENCY MAP ESTIMATION

In this section, we briefly review the neuroscientific background on visual saliency and its implementation and then present our saliency extraction algorithm in further details.

3.1 Review on Visual Saliency

It is well known that allocation of visual attention involves the reflexive involuntary capture of visual

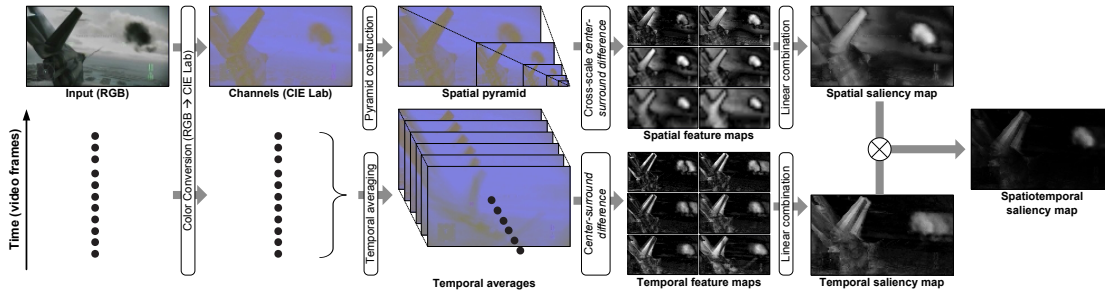


Fig. 2: An example of constructing a spatiotemporal saliency map, given the sequential images of input video frames.

stimuli (bottom-up), in the absence of the user's volitional shifts (top-down factors) [20, 21]. Humans are generally efficient in searching visual information from complex scenes, but this does not necessarily mean that everything is perceived simultaneously. Salient preattentive primitives such as color and lightness are first prioritized and unconsciously detected in parallel. Their *relative contrasts* against surrounding neighborhood (*visual features*) are encoded into different feature maps. Then, a slow conjunctive search is followed to serially integrate the feature maps into a master *saliency map*. The saliency map topographically encodes local conspicuity prioritized for visual search [23, 24], and is believed to be located in the primary visual cortex (V1) but also in the posterior parietal cortex [26, 27].

The neurological mechanism underlying bottom-up attention, called *center-surround antagonism* or *lateral inhibition* [28], gives us an important insight for detecting salient areas. In the receptive fields of photoreceptor cells in the retina, two types of photoreceptors surround each other and compete together with antagonistic responses to light. The photoreceptors in the *excitatory area* are fired in the presence of light, while the others in the *inhibitory area* are fired in the absence of light [24, 29]. When the spot of light becomes large enough to cover the inhibitory area, stimulation of the inhibitory surround counteracts the center's excitatory response, causing a decrease in the neurons' firing rate. Thus, the neurons respond the best to a spot of light that only covers the excitatory areas and does not cover the inhibitory areas. Likewise, visual stimuli where center signals strongly pop out of its local surrounds are likely to be well visible.

Computational models to extract visual saliency from an image typically simulates the center-surround antagonism. The computational model proposed by Itti et al. [30] is the most common choice, which has been distinguished for its effectiveness and plausible outcome in analyzing gaze behaviors. Their key idea is finding salient regions by subtracting a pair of images from each other, each spatially convolved over different kernel sizes. The image blurred with a smaller kernel, preserving finer structures, represents the center, while the image blurred with a larger kernel represents the average intensity of the surround.

Therefore, their image difference (the *center-surround difference*) mimics the lateral inhibition, which effectively captures spatially salient areas.

3.2 Construction of Spatiotemporal Saliency Map

Since the computational saliency map was initially designed for analyzing gaze behaviors in static images, it usually deals with spatial dimension. Many visual features, including color, lightness, orientation, and motion of dynamic objects, have been reported to be preattentive, resulting in high visual saliency [23, 31]. Thus, when directly applying the previous spatial definition to streaming images, overwhelming visual signals would exceed the limited capacity of haptic sensation, resulting in rather distracting haptic stimulation. To deliver effective tactile sensation without excessive tactile signals, it is more plausible to focus on dynamic objects among spatial salient candidates.

To this end, we define the visual saliency both in spatial and temporal dimensions. Fig. 2 illustrates the overall algorithm of building the *spatiotemporal saliency map*. We first separately compute spatial and temporal saliency maps, and multiply them together to yield the final spatiotemporal saliency map. As a result, the final spatiotemporal saliency map is likely to capture dynamically salient spots. In what follows, we describe the algorithms for each saliency map.

3.2.1 Spatial Saliency Map

The common approach for constructing *spatial* saliency map, proposed by Itti et al. [30], is as follows. An input image is first decomposed into a set of distinct *channels* (i.e., preattentive primitives). The channels are typically defined by luminance, hues, and orientations [30]. Then, the image pyramid for each channel is built by successively downsampling the channel image to the quarter 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). Since the level of image pyramid controls the amount of blur, a set of lower levels defines the center (c), and a set of higher levels the surround (s). For each image pyramid, we define six pairs of center and surround levels, $c \in \{2, 3, 4\}$

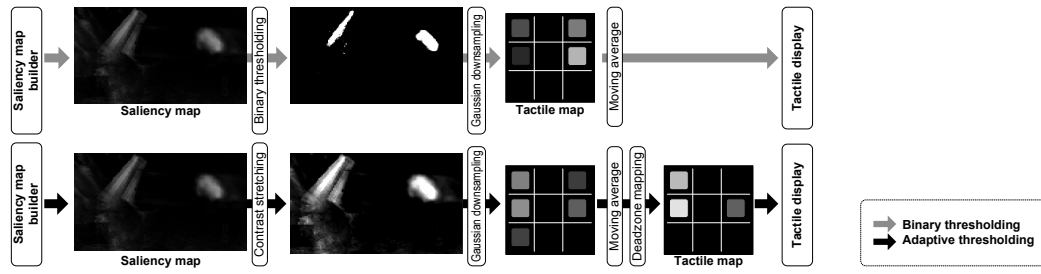


Fig. 3: The two pipelines of visuotactile mapping with binary thresholding and adaptive thresholding.

and $s = c + \delta$, $\delta \in \{3, 4\}$, which is the common configuration from the previous studies [30]. For all (c, s) pairs, the cross-scale image differences (i.e., center-surround differences) are computed by upscaling a coarser image to the finer image and subtracting each other. The difference images are called the *feature maps*. The feature maps for each channel are linearly combined to a *conspicuity map* that encodes saliency for a single channel. Finally, the conspicuity maps for all the channels are linearly integrated again with appropriate weights to yield the spatial saliency map.

Our algorithm for constructing the saliency map basically follows the same approach, but we use CIE $L^*a^*b^*$ (in short, Lab) color space as channels (see the upper row in Fig. 2). In Lab color space, a Euclidean 3D distance between two colors linearly scales with their perceived color difference [32], which considers the effects of three Lab channels at the same time. We use the Lab color distance as a unique feature, and thus, the feature maps are defined only for a single channel. Hence, our algorithm outputs a single conspicuity map equivalent to the spatial saliency map. This approach allows us to avoid finding the linear combination weights of different channels to take their relative contributions into account, which is one of the challenges in estimating the saliency map.

3.2.2 Temporal Saliency Map

The construction of *temporal* saliency map (the lower row in Fig. 2) is performed similarly, except that we use temporal averaging instead of spatial pyramidal averaging. Note that the temporally averaged images and the spatial saliency map have the same size.

Our strategy to define temporal saliency is similar to the recent approach [33]. Our model directly extends the spatial definition, and the spatiotemporal fusion is simply multiplication of both. Hence, it is less costly and suitable for real-time applications, unlike the recent approach [33] using information theory to define and fuse spatial and temporal saliency maps.

Given the streaming video frames, the images are averaged along time instead of space. A level- n image of the temporal averages takes 2^n previous frames as inputs including the current frame. The final temporal saliency map is computed using the *temporal* center-surround differences, whereas the temporal center

levels $c \in \{0, 1, 2\}$ are used instead of $c \in \{2, 3, 4\}$. This strategy is taken to capture finer dynamic changes (e.g., rapid motion).

4 VISUOTACTILE MAPPING

The next step is relating the spatiotemporal saliency map to tactile cues. While the previous work relied on heuristically driven mapping between visual objects and tactile cues [12, 13], our approach directly translates the visual saliency map to the tactile map.

The tactile hardware system that actuates factors can be abstracted to a 2D tactile map (e.g., 3×3 or 5×3). Since the resolution of the visual saliency map is usually much higher than the tactile map, we need to define a mapping between the saliency map and tactile map. In this section we present our approach for effective visual-to-tactile mapping.

A major challenge in the visuotactile mapping lies on the significant difference between the spatiotemporal perceptual sensitivities of vision and touch. The visual saliency map can still carry excessive visual information, e.g., many salient spots or very frequent changes, in spite of our spatiotemporal selection. Due to the limited perceptual bandwidth of human tactile perception, a low-resolution tactile map might not deliver that large amount of information to users. Hence, we need to suppress less salient (presumably less important) areas, which is more likely to direct the user's visual attention to the "right" spot. To this end, we present two novel techniques based on binary thresholding and adaptive thresholding.

4.1 Binary Thresholding

The visuotactile mapping with binary thresholding is a straightforward approach (the upper row in Fig. 3). We first apply hard thresholding that uses a fixed threshold (e.g., 50%) to the saliency map. The binary thresholding removes weak details and emphasizes stronger details (to the maximum intensity) such that they manifest themselves in the tactile map. Then, we apply Gaussian downsampling down to the resolution of the tactile map. The tactile map is moving-averaged (using the window size of a half second) to avoid abrupt changes in the tactile signals. This is also helpful to weaken unstable signals of short duration, which are mostly irrelevant to the movie context.

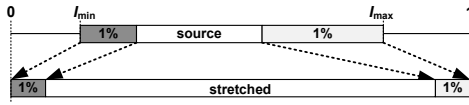


Fig. 4: Contrast stretching in the image histogram.

While this approach works well with simple scenes, it is observed that potentially-meaningful information is often discarded. Also, cell intensities in the down-sampled tactile map can be weakened due to the Gaussian prefiltering. To overcome this weakness, we also propose an adaptive thresholding strategy, better utilizing image characteristics.

4.2 Adaptive Thresholding

The adaptive thresholding approach aims to preserve or emphasize meaningful salient areas, while still suppressing irrelevant areas. Since the saliency values are relative measurements (without precise modeling of the human neuronal system), we take the approach that enhances relative contrasts, consisting of the four steps: 1) image contrast enhancement, 2) Gaussian downsampling, 3) thresholding, and 4) re-emphasis of the salient spots (see the lower row in Fig. 3). The first two steps are applied to the spatiotemporal saliency map, and the last two steps are performed for the tactile map (we call them together *deadzone mapping*).

4.2.1 Contrast Stretching and Downsampling

We first apply *contrast stretching* to the saliency map, which is a standard technique used to enhance low-contrast images in image processing [34]. The saliency map takes differences among similar image pairs resulting from the image pyramid, and its gray-scale contrast is relatively low. By stretching the range of the intensity spectrum, the contrast of the image can be redistributed to widen the gap between salient and non-salient regions. This allows us to better capture important details than the simple hard thresholding.

The contrast stretching is applied by analyzing the histogram of the saliency map (see Fig. 4). Each outer end trail of the histogram (e.g., a single percentile; I_{\min} and I_{\max} in the figure) is converted into the maximum and minimum pixel intensities in the saliency map. The rest of the histogram are stretched and remapped. This procedure makes the intensities of salient regions stronger, while the rest are weakened. Once the contrast of the saliency map is enhanced, it is Gaussian-downsampled to the tactile map.

4.2.2 Deadzone Mapping

The *deadzone mapping* is the combination of thresholding and linear remapping of the tactile signals (see Fig. 5). The deadzone mapping rejects weak tactile signals via thresholding and then stretch cell intensities higher than the cutoff threshold (t_{cutoff}) to the fully saturated cell intensity to re-emphasize the

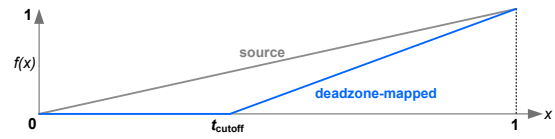


Fig. 5: Deadzone mapping function.

salient tactile cells. Different cutoff values can result in different tactile feelings; a higher cutoff threshold leads to crispier tactile feedback of a shorter duration. Meanwhile, we may encounter an entirely empty tactile map that. Such a tactile map issues no commands in the tactile rendering stage, and contributes to the selective emphasis of visually important events.

4.3 Computational Performance

We report the computational performance of our algorithm. Our system was implemented on an Intel i5 2.66 GHz with OpenCV library. For most movie clips, up to the resolution of 1000×1000 , our system performed faster than 30 frames per second (FPS), proving sufficient real-time performance required for interactive movie playing systems. For high-definition resolutions such as 1080p, a GPU mipmapping technique can be used, as was done in [35].

5 TACTILE RENDERING

To test our saliency-based algorithm for visuotactile mapping, although it is independent of particular hardware platforms, we have built a test platform for tactile display. This section describes the design of our hardware and tactile rendering algorithm.

5.1 Tactile Rendering Hardware

The tactile display is designed to provide vibrotactile stimuli onto the lower back of a user sitting on a chair. The tactors used coin-type eccentric-mass vibration motors, one of the most popular and inexpensive actuators. The maximal voltage to the tactors is 3 V, which can provide the vibration intensity up to 49 G at a 77-Hz frequency.

An array of the tactors is installed on the chair (see Fig. 6). We used 3×3 and 5×3 resolutions for the first and second user experiments, respectively. Each tactor is independently connected to a customized control circuit. The tactors are placed 6 cm apart from each other for reliable discrimination on the back; the spatial discrimination threshold on the back was reported to be roughly 4 cm [36].

For installation, the tactors are wrapped in a cushion cover and inserted into the housing of a chair cushion, which allows for a comfortable contact without propagation of the vibrations to neighbor tactors.

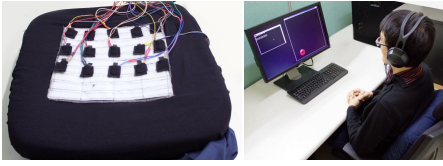


Fig. 6: Our test platform for tactile display.

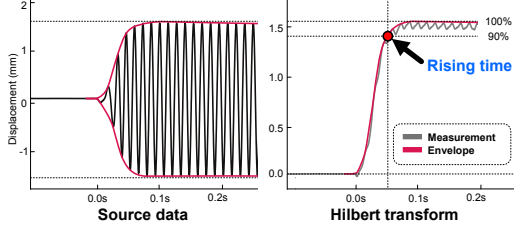


Fig. 7: Vibration intensity response for the definition of rising time.

5.2 Tactile Rendering Algorithm

At the tactile rendering stage, the mapping between the tactile map and the commands to the tactor is straightforward; the tactor array has the same dimension as that of the tactile map. The intensity of each cell in the tactile map is interpreted as the stimulation level of the vibration motor in the corresponding position. The actuation is performed in a continuous form, since the tactile map is also streaming along with the source video.

The procedure of tactile rendering is as follows. The tactile map reads the tactile buffers that store the translations of the video thread. One issue that deserves attention is the mechanical latency of triggering vibration motors, which is generally longer than the video update period. A simple remedy for this latency problem is pre-issuing tactile commands a few frames earlier than the corresponding visual signals.

To this end, we measured the latency values for a number of starting and target vibration voltages. The vibration intensity was noninvasively measured by looking at the tactor's vibration displacement using a laser vibrometer (SICK, model: AOD5-N1). During the measurement, the vibration motor was fixed on a flat sponge. 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. 7). We defined the rising time of actuation as the time period required to reach the 90 percent of the steady-state vibration amplitude. The rising time is an appropriate basis for compensation to achieve vibrotactile stimulation of the desired intensity determined from the given degree of visual saliency.

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

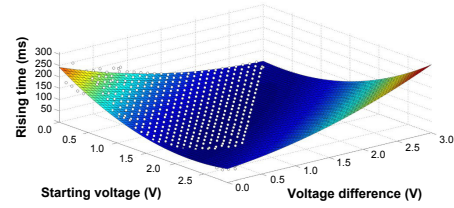


Fig. 8: Rising time measured using a laser vibrometer and its fitted function.

(Fig. 8) to a quadratic form ($R^2=0.98$), such that:

$$f_r(V_s, V_d) = 241.9 - 175.7V_d - 117.7V_s + 38.9V_d^2 + 49.6V_dV_s + 18.1V_s^2, \quad (1)$$

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, we do not have to consider all the target voltages. Weak voltage commands less than a certain threshold (e.g., 0.5 V) can be discarded, since they provide virtually no perceptible actuation. 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. A single data block for tactile signals is filled out of the six preceding video frames; the update rates for visual and tactile rendering are 30 and 5 Hz, respectively. For instance, suppose that the next tactile buffer (six frames after the current frame) is read and 66 ms is found as the rising time. The first four frames of the data block are written with the current input voltage and the remaining two frames with the target voltage at the next data block. This is realized by pre-issuing a tactor command ahead of two video frames.

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



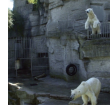
$$f_f(V_d, V_s) = 4.1 + 9.3V_d - 13.5V_s - 3.6V_d^2 - 0.9V_dV_s - 2.5V_s^2, \quad (2)$$

where $R^2 = 0.82$. Most falling times fall within 33 ms (the time required for processing a single image), and thus, we decided not to compensate for the delay.

Lastly, the intensity of each cell in the tactile map is linearly scaled to the voltage range of the vibration motor. The perceived intensity of a vibration motor increases monotonically with its input voltage, albeit the coupled frequency and amplitude of its vibration output [37, 38]. Even though the functional relation is likely to be nonlinear, the simple linear scaling is sufficient for our purpose, also considering the possible nonlinearity in visual saliency estimation and our subsequent transformation procedure.

A concern about force discontinuity might arise here in issuing discrete force commands at 5 Hz within the tactile data block. However, it does not manifest itself, since the low-bandwidth dynamics of the actuator filters out the abrupt changes of the

TABLE 1: Summary of the movies used in Experiment I.

| Movie | M1A | M1B | M1C |
|------------|---|---|---|
| Genre | synthetic | synthetic | real world |
| Motion | static | dynamic | dynamic |
| Complexity | simple | simple | complex |
| Resolution | 1000×1000 | 1000×1000 | 1000×1000 |
| Length | 32 s | 32 s | 60 s |
| Excerpt |  |  |  |

vibration intensity. Our tactile rendering algorithm, tailored to the use of vibration motors, can effectively compensate for the motor delay of vibrotactile rendering to synchronize the visual and tactile stimulations, while avoiding force discontinuity.

6 EXPERIMENT I: COMPARISON OF VISUAL AND VISUOTACTILE RENDERINGS

In this section, we report a user experiment conducted to subjectively assess the usability of our system based on the binary thresholding. Six usability items comparing visual-only and visuotactile presentations for three different types of movies were collected and analyzed via questionnaire.

6.1 Methods

6.1.1 Participants and Apparatus

Twelve paid undergraduate students (6 males and 6 females; 19–30 years old with average 22.3) participated in the experiment. All of them had no problems in detecting vibrations in their backs.

An LCD display of 22 inches presented movie clips to participants, and our in-house tactile display system (Fig. 6) presented the tactile stimuli. The participants were asked to wear a thin T-shirt, to lean back on the chair, and to wear earplugs and a headset to isolate them from factor noises. This experiment used a tactor array of 3×3 as a tactile display system.

6.1.2 Stimuli

Three types of movies, two synthetic and one real (M1A, M1B, and M1C), were used in the experiment (see Table 1 for each). One synthetic movie showed static motions in which one or two balls repeatedly appeared at various locations and stayed more than 1 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 movements of two bears in a zoo. Refer to the accompanying video clips.

In this experiment, we used the binary thresholding algorithm for visuotactile mapping to examine the effects of a straightforward algorithm on the user experience; in the second experiment, we report the

difference responses for the two visuotactile mapping methods and one with manual authoring.

6.1.3 Design and Procedure

The experiment used a one-factor within-subject design. The factor was the provision of tactile cues while playing a movie. By combination with the three movies, 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 so that they can take a rest and fill out the questionnaire. The questionnaire consisted of six questions (see Table 2). Four questions (Q1–Q4) were common to all the sessions, and the other two (Q5 and Q6) were given only in the conditions where tactile cues were presented. Each question used a 100-point scale, where 100 represents the strong agreement to the question, 50 a neutral opinion, and 1 the strongest disagreement. An additional survey asking to freely evaluate the overall system was followed after the six questions.

6.2 Results and Discussion

The subjective ratings of the participants obtained in the experiment are plotted with standard errors in Fig. 9. Overall, the presence of tactile stimulation elicited much positive responses than the visual-only stimulations in all the aspects. The tactile cue supports significantly enhanced the immersion and content delivery as well (Q1 and Q4). The participant preferred the tactile-enabled movies to the original movies (Q2), and also, they found that the tactile-enabled movies are more interesting (Q3).

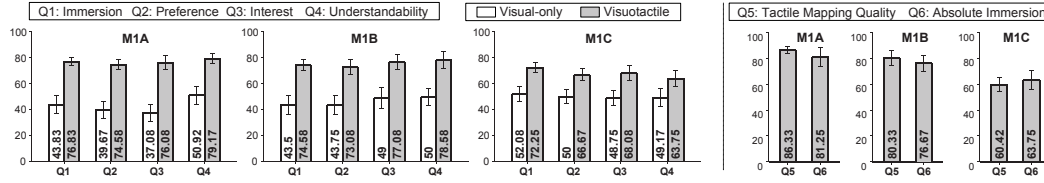
We applied analysis of variance (ANOVA) to see the statistical significances of these differences between the visual-only and visuotactile presentations. Statistical significances were found for all the four questions; all the p values were less than 0.001.





The tactile-specific question, Q5, examined the quality of visuotactile translation. As expected, the tactile stimulations elicited positive responses for all the three movies; note that 50 indicates a neutral response. As for the types of movies, M1A was evaluated as the most well matched, while M1C had the weakest performance. The question on the measurement of absolute immersion, Q6, showed that M1C was the least favored over the other two movies.

The experimental results indicate that saliency-based visuotactile rendering has potential for enhancing movie watching experience, but the results cannot be generalized without care. This results from a potential *novelty bias*, favoring a condition using a new technology (here, haptics), and we do not know how large the bias is. An alternative experimental design to mitigate such a bias effect would be to base

TABLE 2: Subjective questionnaire used in Experiments I and II.

| No. | Type | Questions | Scale |
|-----|--------------|---|-------|
| Q1 | All | How much were you immersed in the movie? | 1–100 |
| Q2 | All | How much did you like the whole system? | 1–100 |
| Q3 | All | How interesting did you find the system? | 1–100 |
| Q4 | All | How easily did you understand the contents of the movie? | 1–100 |
| Q5 | Tactile-only | How well were the <i>vibrations</i> matched with the movie? | 1–100 |
| Q6 | Tactile-only | How much did the <i>vibrations</i> improve the immersion? | 1–100 |

**Fig. 9:** Average subjective ratings measured for Q1, Q2, Q3, Q4, Q5, and Q6 in Experiment I.**TABLE 3:** Summary of the four movies used in Experiment II.

| Movie | M2A | M2B | M2C | M2D |
|-------------------|---|---|--|---|
| Genre | synthetic motion | racing | documentary | flight simulation |
| Motion | static + dynamic | dynamic | dynamic | dynamic |
| Complexity | simple | simple | complex | complex |
| Resolution/Length | 1000×1000/42s | 1280×720/44s | 1300×800/44s | 1280×720/47s |
| Excerpt |  |  |  |  |

the experiment on another condition with tactile cues uncorrelated with the movies.

It was also found that the real-world movie was less favored over the other two. This may have resulted from that the binary thresholding often provides inadequate translation for the more complex real-world movies. In Experiment II, we provide more detailed analysis on the visuotactile mapping algorithms and movie types.

7 EXPERIMENT II: COMPARISON OF TACTILE AUTHORIZING METHODS

This section reports our second experiment performed to subjectively assess the usability of three different tactile authoring methods, the binary thresholding, the adaptive thresholding, and manual authoring. The main objective is to reveal the utility of fully automatic approaches over manual authoring. We note that comparison with visual-only stimulation was excluded in this experiment. This is based on our observation that, in the presence of visual-only cues, subjects are likely to only distinguish the effects of the visual and tactile effects regardless of the differences in the tactile authoring methods. Our goal was to see the sole differences among the tactile authoring methods.

7.1 Methods

Since most of the experimental configuration is the same as that of Experiment I, we highlight only differences in what follows.

7.1.1 Participants and Apparatus

Twenty four paid healthy undergraduate students (12 males and 12 females; 18–27 years old with an average of 22.0) participated in the experiment. While Experiment I used the tactor array of 3×3, Experiment II used a tactor array of 5×3 to better match with horizontally-wider movies.

7.1.2 Stimuli

Four types of movies (M2A, M2B, M2C, and M2D; our synthetic movie clip, excerpts from 2011 FIA Formula One World Championship, the film *Cloverfield*, and the game footage of *Ace Combat Assault Horizon*) were used in this experiment. See Table 3 for summary and the accompanying electronic materials for the videos. M2A presents sudden appearances of two balls or their simple linear/circular motions, combining the two synthetic movies used in Experiment I. Since the balls are very salient, finding the director's intention is easy for the audience. M2B shows dramatic scenes of two cars racing each other. Although a few salient spots appear besides the racing cars, generally the scenes are not very complicated. M2C is a documentary film seen from the bird's eye perspective. The camera usually stays at a fixed point while it rotates either horizontally or vertically. The scenes were taken late at night, so there are only a few salient scenes. One of the salient scenes is found when an explosion begins, giving audiences a clear idea that the explosion moves from left to right. M2D is footage

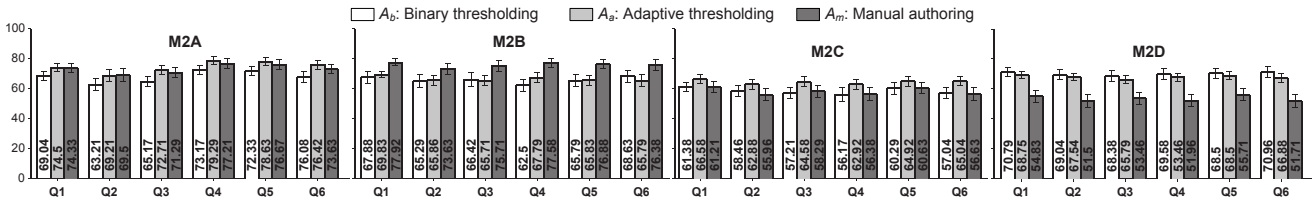


Fig. 10: Summary of the average subjective ratings with standard error bars measured in Experiment II. The results compare the three tactile authoring methods, A_b , A_a , and A_m for the six questionnaire items (Q1–Q6).

from a flight simulation game where a user controls the fighter to protect the city from enemies. The movie contains mixed scenes with slow flying scenes and complex scenes of chasing and destroying enemies.

Three kinds of tactile authoring methods were chosen to present tactile cues: (1) the binary thresholding (A_b), (2) the adaptive thresholding (A_a), and (3) manual authoring (A_m). A_b used 0.5 as the hard threshold, which was carefully chosen as the best in our pilot test among the three values (0.25, 0.5, 0.75 over the maximum of 1). The largest difference between A_b and A_a was the presence of *selective emphasis*, controlled by t_{cutoff} of the deadzone mapping of A_a . We chose $t_{cutoff} = 0.25$ as the best value based on our pilot test that compared the effect of different cutoff values (0.25, 0.5, 0.75). Although the differences were not very noticeable, $t_{cutoff} = 0.25$ elicited the most positive responses.

In the manual authoring, tactile cues were generated by a tactile authoring expert. Since the quality of manual authoring heavily relies on the expertise of designers, we recruited a tactile effect designer who had industrial experience of tactile feedback authoring for more than two years. The designer was asked to carefully design tactile cues according to the context of the movies. For this, the designer watched the movies a couple of times before making tactile effects so that he could obtain understandings in a longer context. The designer then tried to optimize tactile effects in a way that emphasizes the most exciting events over the whole story while ignoring small, minor, or meaningless events. It should be stressed that this strategy was determined by the designer himself based on his experiences; we did not provide any input about desired tactile effects.

To facilitate the manual design process, the designer used the authoring tool developed by Kim et al. [5], which supports a frame-by-frame graphical authoring of tactile signals for an array of tactors. The program was initially made for tactile gloves, but we adapted it to work with our tactile chair system.

7.1.3 Design and Procedure

The experiment used a one-factor within-subject design. The factor was the type of tactile authoring method, which had three levels, A_b , A_a , and A_m . Combining with the four movies, each participant went

through the total of 12 successive experimental sessions (3 authoring methods \times 4 movies). Their presentation order was balanced using Latin squares. The six questionnaire items (Q1–Q6) identical to those used in Experiment I were also used in this experiment. Note that, unlike Experiment I, the participants were asked to fill all of the six questions, as all the conditions provided tactile cues in this experiment.

7.2 Results and Discussion

The subjective ratings obtained in the experiment are plotted with standard errors in Fig. 10. Overall, as expected, all the three methods elicited positive responses ($\gg 50$ of a neutral response), indicating that the synchronized tactile display system improved the multimedia presentation; the participants were more engaged and immersed in the movie watching experience. For all the questions (Q1–Q6), similar responses were observed with slight variations.

However, the three authoring methods showed quite different scores depending on the movie type. In what follows, we analyze their differences with respect to the movie type. We again applied one-way within-subject ANOVA to see the statistical significances in the tactile authoring methods (see Table 4).

In M2A, A_a and A_m were rated similar, implying that the tactile feedback can be well generated with A_a to the quality similar to A_m . This would result from the simple motions of the balls in the movie. However, A_b was rated a bit lower than the other two. The tactile feedback generated with A_b tend to be rather exaggerated with many factor actuation. Nonetheless, statistically significant differences were not found in any of the six questions.

In M2B, it is clear that the manual authoring (A_m) worked best, and the differences against the other two were statistically significant. The other two algorithms (A_b and A_a) showed similar responses, as confirmed by Tukey's multiple comparison test ($p > 0.2$ for all the A_b – A_a pairs for the six questions). This trend seems to result from the characteristics of the racing scene; the periphery background is rapidly passing, but the screen locations of the cars are not moving (refer to the accompanying video). While the tactile designer could easily reflect these patterns to the tactile feedback, the other two methods often failed to capture such characteristics. For A_b , too many tactors were actuated, because the binary thresholding regards even

TABLE 4: Source table of the 1-way within-subject ANOVA for the effect of tactile rendering method. Note $*p < .05$.

| | M2A | | M2B | | M2C | | M2D | |
|----|------------|-------|------------|-------|------------|--------|------------|-------|
| | $F_{2,23}$ | p | $F_{2,23}$ | p | $F_{2,23}$ | p | $F_{2,23}$ | p |
| Q1 | 2.40 | .1023 | 5.23* | .0090 | 1.25 | 0.2951 | 9.43* | .0004 |
| Q2 | 2.73 | .0760 | 3.40* | .0420 | 1.35 | 0.2694 | 11.80* | .0001 |
| Q3 | 3.00 | .0611 | 4.55* | .0158 | 1.26 | 0.2945 | 7.54* | .0015 |
| Q4 | 2.64 | .0824 | 10.77* | .0001 | 1.23 | 0.3011 | 12.78* | .0001 |
| Q5 | 2.49 | .0941 | 7.23* | .0019 | 0.62 | 0.5414 | 7.59* | .0014 |
| Q6 | 3.13 | .0533 | 5.16* | .0095 | 2.21 | 0.1213 | 10.46* | .0002 |

small movements of the background as salient. For A_a , visual saliency was captured for the background movements instead of car motions. To improve this types of motions, it would be necessary to take the relative motion against background into account.

In M2C, no statistically significant differences were found in any of the six questions, but A_a had a marginal win over the other two methods. The scene is very dark and complex, which does not allow the system to detect clear salient spots. A_a provided adequate tactile feedback to moderate extent, well following the tension of the scene, as well as the explosion. In contrast, A_b failed to pinpoint the salient spots, because too many factors were actuated. As for A_m , the tactile designer perceived only the explosion scene to be important, leading to rather insufficient tactile cues for delivering the tension of the scene.

In M2D, the manual authoring (A_m) scored the lowest. This was also statistically significant, while A_b and A_a had no differences (confirmed via Tukey's multiple comparison test). It seems that this resulted from that the tactile designer focused solely on the firing shots (the most exciting scene in the movie) to clearly distinguish the firing scene against shots with smoothly flying actions. In contrast, A_a was successful to capture smooth flying motions as well as firing shots, but A_b still actuated too many factors.

We initially expected that the manual authoring would always outperform the two automatic authoring methods, but this was not shown true with M2C and M2D. The manually-authored tactile effects were apt to be less frequent and spatially sparse, only emphasizing important events. If films of a longer running time had been used, such efforts could have been more effective. However, the movie clips were short, and it appears that the participants preferred more frequent vibrations, resulting from the saliency-based authoring methods, to sparse ones. The frequent vibrations may have contributed to stressing the ambience of the scene, similar to the diegetic sound (e.g., mood music) in movies [11]. Nonetheless, the manual authoring in this experiment did not employ such dependency of user preference on the characteristics of the movies such as running time, ambience of the scene, and the point of view [5]; we encourage further investigation on this dependency for future work. We hereby note that additional iteration to better reflect user preference would enhance

the quality of manual authoring and the validity of experimental results as well.

The user-perceived differences between the two automatic authoring methods were not clearly distinguishable, although our impression is that the adaptive thresholding provides better-tuned tactile effects. In terms of the frequency of stimulation and the number of factors simultaneously actuated, the adaptive thresholding algorithm sits in between the binary thresholding algorithm and the manual authoring, but seems to be closer to the binary thresholding.

8 CONCLUSIONS AND GENERAL DISCUSSION

Tactile feedback is one of the promising components in mediating more immersive multimodal experience beyond the traditional audiovisual interaction. In this article, we presented an automated framework for authoring synchronized visuotactile effects in real time. Visual saliency served as a basis for extracting spatiotemporal importance from existing visual media and translating it to the tactile cues that are rendered on factors installed on a chair.

The two user studies were conducted to evaluate the visuotactile effects on user experience. The first user study found that visuotactile rendering was much preferred to visual-only presentation, eliciting more immersion and involvement, as well as better understandings of the content. The second user study showed that our visual saliency-driven automated approach has high potential for creating effective tactile feedback and supporting sophisticated manual authoring to a large extent. Nonetheless, it is not conclusive yet which method is the best, encouraging more research on this direction.

The quality of visuotactile mapping in creating tactile cues is one of the keys in conveying better experience. The fact that our saliency-driven framework does not require the high-level semantics of objects present in the scene is an important advantage, but excessive or spatially inadequate translation can be a natural consequence. The experimental results indicate that the simple configurations (as in the ball movies) were free from this problem, while complex scenes were more difficult to synchronize with the scene semantics. One way to improve the quality of translation is combining automated tools and manual

authoring. Given tactile cues generated from the automated tools, the designer can prune out redundant cues or add semantics overlooked by the tools.

The qualitative evaluations collected in Experiment I via additional user survey revealed: (1) some of the users regarded the system as abnormal or difficult to follow in their initial trials; and (2) the audience often wanted to be immersed into the movie but to less care about the exact spatial information. Such responses led us to limit the occurrences and duration of tactile cues, such that the tactile cues are more memorable without invoking excessive tactile events. In this sense, our approach with the adaptive thresholding was successful to a moderate extent.

Another important finding is the presence of tactile cues—even uninformative and *ambient* tactile effects—often allows the user to feel present in the scene. In Experiment II, sparse feedback focusing solely on the most important shots (in particular for M2D) was less favored. It can be a good research direction how to incorporate *ambient vibration* to match the visual feeling of the entire scene (e.g., weak vibration before the exciting firing scene). Also, finding optimal combination of selective emphasis (used in the adaptive thresholding) with the ambient vibration could be another good research direction.

In addition, we note another tuning point, relative motions. During extracting the temporal motions, we overlooked the relative motions of foreground objects against the rapidly changing background. We envision that, by considering this, the whole impression of the system can be significantly improved. For instance, in the racing movie, we can assign more vibrations on the cars to emphasize their rapid movements.

Our approach has importance for haptic content creators and interaction designers, who strive to create online or offline content inducing spatially-present experience. Haptic cues extracted automatically from the existing media can facilitate the rapid production of tactile movies in the postproduction. However, automatic authoring needs careful consideration on the multimedia context. Our lessons learned from the design and evaluation of our system can be instrumental in guiding such hybrid authoring approaches.

ACKNOWLEDGMENTS

We greatly appreciate Yeongmi Kim and Jeha Ryu for allowing us to use their graphical authoring tool of tactile effects. This work was supported by the Pioneer Research Center, Mid-career, and Global Frontier (on *Human-centered Interaction for Coexistence*) R&D programs through the NRF grants funded by the Korea Government (MSIP) (Nos. 2011-0027994, 2013R1A2A2A01016907, 2012R1A2A2A01045719, 2012M3A6A3055695). Correspondence concerning this article can be addressed to Sungkil Lee.

REFERENCES

- [1] S. Choi and K. J. Kuchenbecker, "Vibrotactile display: Perception, technology, and applications," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2093–2104.
- [2] 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*, vol. 10, no. 1, pp. 37–53, 2009.
- [3] E. Oh, M. Lee, and S. Lee, "How 4D Effects cause different types of Presence experience?" in *Proc. Int. Conf. on Virtual Reality Continuum and Its Applications in Industry*, 2011, pp. 375–378.
- [4] C. Collins, "Tactile Television-Mechanical and Electrical Image Projection," *IEEE Trans. Man-Machine Systems*, vol. 11, no. 1, pp. 65–71, 1970.
- [5] Y. Kim, J. Cha, J. Ryu, and I. Oakley, "A Tactile Glove Design and Authoring System for Immersive Multimedia," *IEEE Multimedia*, vol. 17, no. 3, pp. 34–45, 2010.
- [6] P. Lemmens, F. Cromptvoets, D. Brokken, J. van den Eerenbeemd, and G. De Vries, "A body-conforming tactile jacket to enrich movie viewing," in *Proc. World-haptics*, 2009, pp. 7–12.
- [7] M. Karam, F. A. Russo, and D. I. Fels, "Designing the model human cochlea: An ambient crossmodal audio-tactile display," *IEEE Trans. Haptics*, vol. 2, no. 3, pp. 160–169, 2009.
- [8] Y. Visell, J. R. Cooperstock, B. L. Giordano, K. Franinovic, A. Law, S. McAdams, K. Jathal, and F. Fontana, "A Vibrotactile Device for Display of Virtual Ground Materials in Walking," in *Proc. Eurohaptics'08*. Springer-Verlag, 2008, pp. 420–426.
- [9] I. Hwang, H. Lee, and S. Choi, "Real-time dual-band haptic music player for mobile devices," *IEEE Trans. Haptics*, 2013, to appear.
- [10] J. Lee and S. Choi, "Real-time perception-level translation from audio signals to vibrotactile effects," in *Proc. ACM SIGCHI Conf. on Human Factors in Computing Systems*, 2013, pp. 2567–2576.
- [11] F. Danieau, A. Lécuyer, P. Guillotel, J. Fleureau, N. Mollet, and M. Christie, "Enhancing audiovisual experience with haptic feedback: a survey on hav," *IEEE Trans. Haptics*, vol. 6, no. 2, pp. 193–205, 2013.
- [12] B.-C. Lee, J. Lee, J. Cha, C. Seo, and J. Ryu, "Immersive live sports experience with vibrotactile sensation," *Lecture Notes in Computer Science*, vol. 3585, pp. 1042–1045, 2005.
- [13] S. ur Réhman, J. Sun, L. Liu, and H. Li, "Turn your mobile into the ball: rendering live football game using vibration," *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1022–1033, 2008.
- [14] T. Yamaguchi, A. Akabane, J. Murayama, and M. Sato, "Automatic generation of haptic effect into published 2d graphics," in *Proc. EuroHaptics Conference*, 2006.
- [15] C. Spence and J. Driver, *Crossmodal space and crossmodal attention*. Oxford University Press, USA, 2004.
- [16] H. Z. Tan, R. Gray, J. J. Young, and R. Traylor, "A haptic back display for attentional and directional cueing," *Haptics-e* (<http://www.haptics-e.org>), vol. 3, no. 1, 2003.
- [17] A. Sklar and N. Sarter, "Good vibrations: Tactile feedback in support of attention allocation and human-automation coordination in event-driven domains," *Human Factors*, vol. 41, no. 4, pp. 543–552, 1999.
- [18] N. Sarter, "The need for multisensory interfaces in support of effective attention allocation in highly dynamic event-driven domains: the case of cockpit automation," *The International Journal of Aviation Psychology*, vol. 10, no. 3, pp. 231–245, 2000.
- [19] C. Ho, H. Tan, and C. Spence, "Using spatial vibrotactile cues to direct visual attention in driving scenes,"

Transportation Research Part F: Traffic Psychology and Behaviour, vol. 8, no. 6, pp. 397–412, 2005.

- [20] C. E. Connor, H. E. Egeth, and S. Yantis, "Visual attention: Bottom-up versus top-down," *Current Biology*, vol. 14, no. 19, pp. R850–R852, 2004.
- [21] D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of saliency in the allocation of overt visual attention," *Vision Research*, vol. 42, no. 1, pp. 107–123, 2002.
- [22] D. Hoffman and M. Singh, "Saliency of visual parts," *Cognition*, vol. 63, no. 1, pp. 29–78, 1997.
- [23] A. M. Treisman and G. Gelade, "A Feature-Integration Theory of Attention," *Cognitive Psychology*, vol. 12, pp. 97–136, 1980.
- [24] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annual review of neuroscience*, vol. 18, no. 1, pp. 193–222, 1995.
- [25] M. Kim, S. Lee, and S. Choi, "Saliency-driven tactile effect authoring for real-time visuotactile feedback," *Lecture Notes on Computer Science (Proc. Eurohaptics'12, Part I)*, vol. 7282, pp. 258–269, 2012.
- [26] J. P. Gottlieb, M. Kusunoki, M. E. Goldberg *et al.*, "The representation of visual saliency in monkey parietal cortex," *Nature*, vol. 391, no. 6666, pp. 481–484, 1998.
- [27] Z. Li, "A saliency map in primary visual cortex," *Trends in cognitive sciences*, vol. 6, no. 1, pp. 9–16, 2002.
- [28] S.-i. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological cybernetics*, vol. 27, no. 2, pp. 77–87, 1977.
- [29] J. Allman, F. Miezin, and E. McGuinness, "Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons," *Annual review of neuroscience*, vol. 8, no. 1, pp. 407–430, 1985.
- [30] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [31] K. Nakayama, G. H. Silverman *et al.*, "Serial and parallel processing of visual feature conjunctions," *Nature*, vol. 320, no. 6059, pp. 264–265, 1986.
- [32] C.I.E., "Recommendations on uniform colour spaces, colour difference equations, psychometric colour terms," 1978, supplement No.2 to CIE publication No.15 (E.-1.3.1) 1971/(TC-1.3.).
- [33] C. Liu, P. C. Yuen, and G. Qiu, "Object motion detection using information theoretic spatio-temporal saliency," *Pattern Recognition*, vol. 42, no. 11, pp. 2897–2906, 2009.
- [34] R. C. Gonzalez and R. E. Woods, *Digital image processing*. Prentice Hall Press, 2002.
- [35] S. Lee, G. J. Kim, and S. Choi, "Real-Time Tracking of Visually Attended Objects in Virtual Environments and Its Application to LOD," *IEEE Trans. Visualization and Computer Graphics*, vol. 15, no. 1, pp. 6–19, 2009.
- [36] S. Weinstein, "Intensive and extensive aspects of tactile sensitivity as a function of body part, sex, and laterality," in *Proc. Int. Symp. Skin Senses*, 1968, pp. 195–222.
- [37] J. Ryu and S. Choi, "posVibEditor: Graphical authoring tool of vibrotactile patterns," in *Proc. IEEE Int. Workshop on Haptic, Audio and Visual Environments and Games*, 2008, pp. 120–125.
- [38] J. Ryu, J. Jung, G. Park, and S. Choi, "Psychophysical Model for Vibrotactile Rendering in Mobile," *Presence: Teleoperators and Virtual Environments*, vol. 19, no. 4, pp. 364–387, 2010.



Myongchan Kim received the degree of Bachelor of Information Technology from Carleton University in 2009, and the MS degree in Electrical and Computer Engineering from POSTECH in 2012. His main research interest is visuotactile feedback rendering.



computer interaction, with emphasis on perception-based rendering and GPU algorithms.



Seungmoon Choi is an associate professor of Computer Science and Engineering at Pohang University of Science and Technology (POSTECH). He received the BS and MS degrees in Control and Instrumentation Engineering from Seoul National University in 1995 and 1997, respectively, and the PhD degree in Electrical and Computer Engineering from Purdue University in 2003. He is an associate editor of IEEE Transactions on Haptics and an editorial board member of Virtual Reality. He was a co-chair of the IEEE Technical Committee on Haptics in 2009–2011. He received a 2011 Early Career Award from IEEE Technical Committee on Haptics and several best paper awards from premium international conferences. His research interests lie on haptic rendering and perception, with emphasis on kinesthetic rendering of hardness and texture, tactile rendering, skill modeling and transfer, haptic augmented reality, mobile haptic interface, data haptization, and applied perception. His basic research has been applied to mobile devices, automobiles, virtual prototyping, and motion-based remote controllers. He is a member of the IEEE.