symmetry

MDPI

Article

Design of Desktop Audiovisual Entertainment System with Deep Learning and Haptic Sensations

Chien-Hsing Chou [1], Yu-Sheng Su [2,*], Che-Ju Hsu [1], Kong-Chang Lee [1] and Ping-Hsuan Han [3]

[1] Department of Electrical Engineering, Tamkang University, New Taipei City 25137, Taiwan; chchou@mail.tku.edu.tw (C.-H.C.); 607450193@s07.mail.tiki.edu.tw (C.-J.H.); 608450069@s07.mail.tiki.edu.tw (K.-C.L.)

[2] Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung 20224, Taiwan

[3] Department of Interaction Design, National Taipei University of Technology, Taipei 10608, Taiwan; han@ntut.edu.tw

* Correspondence: ntoucsiesu@mail.ntou.edu.tw

Received: 21 September 2020; Accepted: 16 October 2020; Published: 19 October 2020

Abstract: In this study, we designed a four-dimensional (4D) audiovisual entertainment system called Sense. This system comprises a scene recognition system and hardware modules that provide haptic sensations for users when they watch movies and animations at home. In the scene recognition system, we used Google Cloud Vision to detect common scene elements in a video, such as fire, explosions, wind, and rain, and further determine whether the scene depicts hot weather, rain, or snow. Additionally, for animated videos, we applied deep learning with a single shot multibox detector to detect whether the animated video contained scenes of fire-related objects. The hardware module was designed to provide six types of haptic sensations set as line-symmetry to provide a better user experience. After the system considers the results of object detection via the scene recognition system, the system generates corresponding haptic sensations. The system integrates deep learning, auditory signals, and haptic sensations to provide an enhanced viewing experience.

Keywords: audiovisual entertainment system; deep learning; four-dimensional (4D); haptics; scene recognition; audiovisual entertainment system

1. Introduction

The increasing prevalence of technology has led to the emergence of multimedia products for producing three-dimensional (3D) and four-dimensional (4D) films. A 4D film experience combines physical effects with a 3D film when the audience watches it. Effects simulated in a 4D film may include rain, mist, explosions, wind, temperature changes, strobe lights, scents, vibrations, and a feeling of motion. This type of film can enhance audience experiences by adding elements in addition to visual and auditory sensations. The equipment required to present a 4D film in a theater is expensive, and the ticket price for a 4D film is more expensive than usual. Thus, to date, few 4D films have been released in theaters.

The research of Gonçalves et al. evaluated how multisensory stimuli (e.g., the use of wind, passive haptics, vibration, and scent) affect the senses of virtual reality (VR) users [1]. Their research indicated that the use of multisensory stimuli with VR and multimedia techniques creates new and engaging user experiences. Some researchers have presented VR applications that integrate multiple tactile sensations to enhance immersion experience through kinesthesia and cutaneous sensations. The studies of Han et al. presented two VR applications (called the SoEs (Sword of Elements) [2] and BoEs (Bits of Elements) [3]) demonstrated the potential of using haptic feedback, specifically heat,

air, vibration, and reaction force, for VR controllers. The VR application AoEs (Area of Elements) [4] provided multiple tactile sensations to enhance the teleportation experience in two virtual environments. Inside a room-scale physical space, users can stay or walk through a portal between desert and snow scenes. Günther et al. presented a device called the Therminator that provides warm and cold on-body sensations in VR through heat conduction of flowing liquids with various temperatures [5]. Their experiment results indicated that thermal stimuli could override visual stimuli to control how the audience perceives temperature. Sasaki et al. proposed Virtual Super-Leaping to enhance the VR jumping experience [6]. Their haptic device was constructed with propeller units and a rod to generate various kinesthetic and airflow haptic sensations in midair.

In this study, we integrated multiple tactile sensations to enhance the user immersion experience. In our developed system, users are able to feel haptic sensations during movie scenes they are watching. The first step is to detect specific objects (e.g., fire) in movie scenes; then, our system controls the appropriate haptic modules to simulate immersive environments. Many studies and techniques of object detection have been proposed for a variety of applications. The TensorFlow object detection API [7] and Google Cloud Vision [8] are well-known object detection techniques. The researchers also integrated TensorFlow and Faster R-CNN [9] algorithms to construct a powerful real-time object detection API. Bergman's research incorporated a symmetries concept into the machine learning model. Their method could reduce complexity, training time, and over-fitting in the training process [10]. YOLO (You Only Look Once) is a notably famous object detection technique [11–13], with which numerous object detection systems have been developed [14,15]. Xu et al. proposed an OpenCL-based high-throughput FPGA accelerator for the YOLOv2 object detection algorithm [14]. Simulation results illustrated that their proposed method could execute at speeds of 35 and 71 frames per second (fps) for YOLOv2 inference computation and tiny YOLOv2, respectively. Maher et al. proposed a deep-patch orientation network method for multi-ground target tracking; their system learned the target's orientation based on structural information [15]. YOLO and *Faster R-CNN* have been applied to demonstrate that detection accuracy can be improved at the same processing speed. Although YOLO is a reasonable method for detecting objects in a video, this study selected the MXNet framework [16] and the single-shot multibox detector (SSD) [17] deep learning algorithm to train an object detection system for animated videos. Because the SSD algorithm eliminates proposal generation and subsequent pixels in a single network, and it is also relatively simple and easy to train relative to most YOLO techniques. In our field trials, the SSD model achieved comparable detection accuracy to YOLO.

This study designed a four-dimensional (4D) audiovisual entertainment system called Sense. Some processing techniques were based on ideas from our previously published research [18]. This system comprises a scene recognition system and hardware modules that provide haptic sensations to users when they watch movies and animations at home. Figure 1 illustrates the system framework. First, we capture an image from a video and then input it into the scene recognition system to detect objects. For general videos, we used Google Cloud Vision to detect objects in the image; after the objects (e.g., fire, wind, or rain) had been detected, the 4D sensation device triggered a corresponding haptic sensation module (e.g., a heat lamp or fan). Because Google Cloud Vision cannot detect objects in animated videos, this study applied SSD to develop a system that can detect fire-related objects in animated videos. The proposed system can label fire-related objects in an animated video, enabling users to use Sense to obtain the corresponding haptic experience when watching the animated video. The following sections introduce the design methods of the 4D sensation device and the scene recognition subsystem of the proposed system.

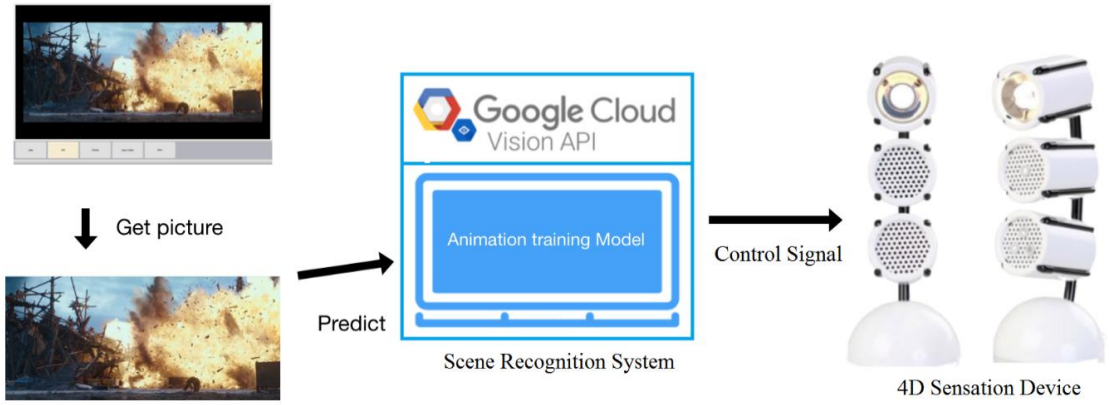


Figure 1. The system framework of Sense.

2. Design of 4D Sensation Device

This study designed a desktop 4D sensation device that gives users a viewing experience similar to that of a 4D cinema when watching a video. Figure 2a presents the appearance of this device. The device allows users to receive haptic sensations of cold, heat, and vibration corresponding to the scene of the video they are watching at home. The red dot at the center of Figure 2b is the optimal experience point of this haptic device. To provide the most effective experience, after actual field trials, we took the head of the user in a sitting position as the center (default distance is 102 cm from the ground) and designed two ellipsoids with radii of 75 cm and 87 cm, respectively; the physical framework of the hardware device has been designed to conform to the radian of the ellipsoids. This design allows for each module to point to the center and superimpose various haptic sensations in the most efficient manner. For actual application, the designed concept of the 4D sensation device is portable, requires low power consumption, and is low cost. The length, width, and height of one single device are 20 cm, 20 cm, and 45 cm; and its weight is 2 Kg. This means that the user could easily set up and move this device. The electric power of one single device requires only 5 A current at 110 V, the user could use it as a home appliance. Besides, its construction cost is US $100.

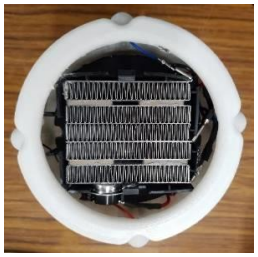
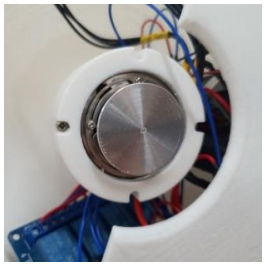
(**a**) (**b**)

Figure 2. (**a**) The appearance of the 4D sensation device; (**b**) the experience area of the 4D sensation device.

2.1. Haptic Modules of 4D Sensation Device

In the 4D sensation device, we used six types of haptic modules set as line-symmetry to provide a better user experience (as listed in Table 1). These modules can be processed individually or together to provide various haptic sensation experiences. The six modules are as follows.

Table 1. Six types of haptic modules.

Heat Module (Infrared Heat Lamp)		Hot Air Module (Mini Space Heater)	
Wet Module (Ultrasonic Humidifier Generator)		Wind Module (Fan)	
Scent Module (Ultrasonic Humidifier Mist Generator)		Vibration Module (Vibration Speaker)	

A. Heat Module

The heat module uses an infrared heat lamp to simulate noncontact heat sources (100 W, 110 V) such as fire, explosion, and burning scenes. In this study, we installed two heat modules, one or both of which are used depending on the scenes to provide users various levels of heat.

B. Hot Air Module

The hot air module was modified with a desktop heater (250 W, 110 V) to simulate contact-type heat sources and provide strong heat sensations. This is used to simulate explosions and intense fires.

C. Wet Module

Our wet module uses ultrasonic humidifiers to emit water vapor (5 W, 5 V), simulating watery scenes such as depictions of rain and hurricane. The module contains a detachable water tank that users can refill with water.

D. Wind Module

Our wind module uses a 12-V fan with adjustable speed and pulse-width modulation (PWM) to control the wind speed. It simulates scenes of wind or air rushing past a speeding object, such as a typhoon, car racing, and flying airplane scenes.

E. Scent Module

Similar to the wet module, our scent module uses ultrasonic humidifiers to emit water vapor (5 W, 5 V), but this module emits light mist. No wetness is felt when the user comes into contact with this mist. Various essential oils can be added so that users can smell these odors during specific scenes. For example, tree smells are generated for forest scenes. This module also has a detachable water tank that the user can refill with water.

F. Vibration Module

A vibration speaker installed at the bottom of the 4D sensation device functions as the vibration module. This device produces sound through desktop resonance, and the tabletop vibrations give users the feeling of vibrations. When an explosion or collision scene appears in a movie, it is accompanied by shaking sound effects to enable the tabletop to generate obvious vibrations.

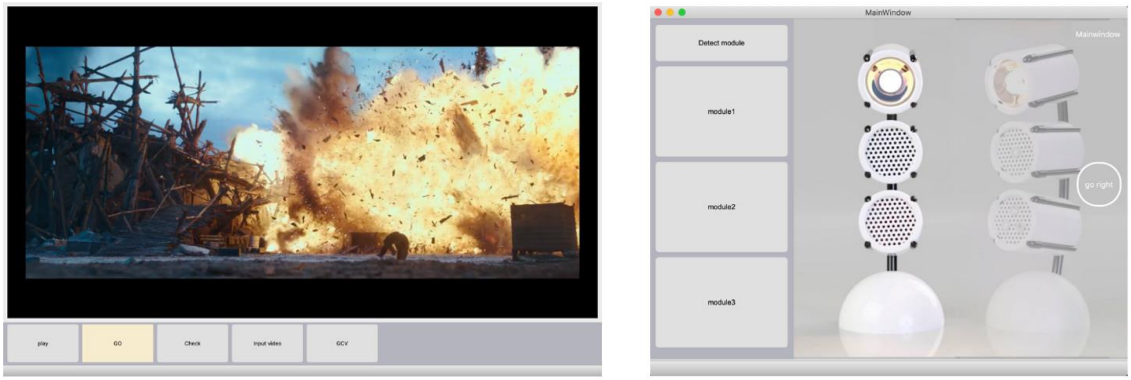
2.2. Hardware Core of 4D Sensation Device

We used Fusion360 [19] to design the external appearance of the device and used a 3D printer to produce the packaging. Additionally, for this system, we used an Arduino Pro Mini as the computing core for receiving computer signals, and the modules were run or stopped based on its control signals. As illustrated in Figure 3, each haptic module can be assembled freely in any position depending on the requirements of the user. Each module position has four DC power jacks of 2.1 mm, two of which are used to provide 110 V AC power. Unlike the operation of general power jacks, the inner tips of two power jacks provide power, and the sleeves of the power jacks protect users from electric shocks due to accidental contact; therefore, they do not provide any power. Furthermore, we placed the two power-providing power jacks diagonally opposite to each other. Such protective measures were intended to protect the device from being damaged by any short circuit that might occur if the module were installed incorrectly. We connected resistors of various values to the two remaining power jacks of each module, and in testing, the system was able to recognize the function of each module according to the resistance value.

Figure 3. Module installation.

3. Scene Recognition System

To use Sense for watching a video, we developed an operating software for this system, as illustrated in Figure 4a. In testing, first, the system read a 30 fps video, and then the system captured the images of the video every second. To use Sense for watching a video, first, the Sense operating system (Figure 4a) read a 30 fps video, and then the system captured images from the video in every second. If the video was a general live video, then these images underwent object detection according to Google Cloud Vision. Each captured image from live video was uploaded to the Google cloud server one by one for computation. The process spent 2–5 s for one image, depending on the network quality. For a one-minute live video, the system usually requires more than 2–5 min to record the detection results of those uploaded images. If the video was an animated video, then the animation scene recognition system trained with the SSD model conducted detection. The process spent 0.2 s for detecting fire-related objects in an animation image by using the MacBook Pro notebook. For a one-minute animation video, the Sense operating system requires only about 15–20 s to generate and record the detection result. When the user watches a general live video or an animation video, the corresponding record of the detection results is loaded in beginning. The 4D sensation device then allows users to receive haptic sensations according to these detection results. Subsequently, based on the detected objects, the system ran the corresponding haptic modules. Figure 4b depicts the control software of the 4D sensation device. Users were able to test the effect of each haptic module in this interface and adjust the operating time and intensity of each module according to their personal preferences. For example, they were able to adjust the operating time of the heat module or the intensity of the vibration module.



(**a**) (**b**)

Figure 4. (**a**) Sense operating software; (**b**) control software of the 4D sensation device.

According to our field trials, the Google Cloud Vision is a superior object-detection system. It can provide a high accuracy rate for detecting a variety of objects in an image, and some of these objects are difficult to define and detect (e.g., rain, winter, or snow). But two limitations have occurred when we integrated Google Cloud Vision with the 4D sensation device. First, our system needs to upload the captured image to the Google cloud server for computation, and then receives the detection results. The process spent 2–5 s for one image, depending on the network quality and traffic of the Google cloud server. Second, Google Cloud Vision cannot detect objects in animated videos. To overcome these limitations, we trained the fire-related object detection system with the SSD model. According to the field test for the general live videos, however, the accuracy rate obtained by using Google Cloud Vision outperforms the one using the SSD-based model. In our opinion, it is a better choice to use Google Cloud Vision to detect objects from general live video.

3.1. Scene Recognition System Based on Google Cloud Vision

For general live videos, we used Google Cloud Vision on each captured image for object detection; for each detected object, the corresponding haptic module was run. A video of our system based on Google Could Vison is available at Reference [20]. We set up 35 types of objects in our study, some of which are listed in Table 2 along with the corresponding haptic modules. As displayed in Figure 5, the objects 'wildfire', 'explosion', 'waterfall', 'snow', 'wind', or 'forest' could be detected by Google Cloud Vision. If the scene recognition detected the object 'explosion' in an image, then the 4D haptic sensation device ran two heat modules, with the hot air module giving users an enhanced viewing experience. If the scene recognition system detected the object 'snow' in an image, then the 4D haptic sensation device ran the wet module and wind module. In this study, we added pine needle oil into the scent module; the users were able to smell the pine needle scent when the system detected 'forest' or 'tree' in a video. Because the vibration strength was determined by the sound volume in the video, the vibration module is not included in Table 2.

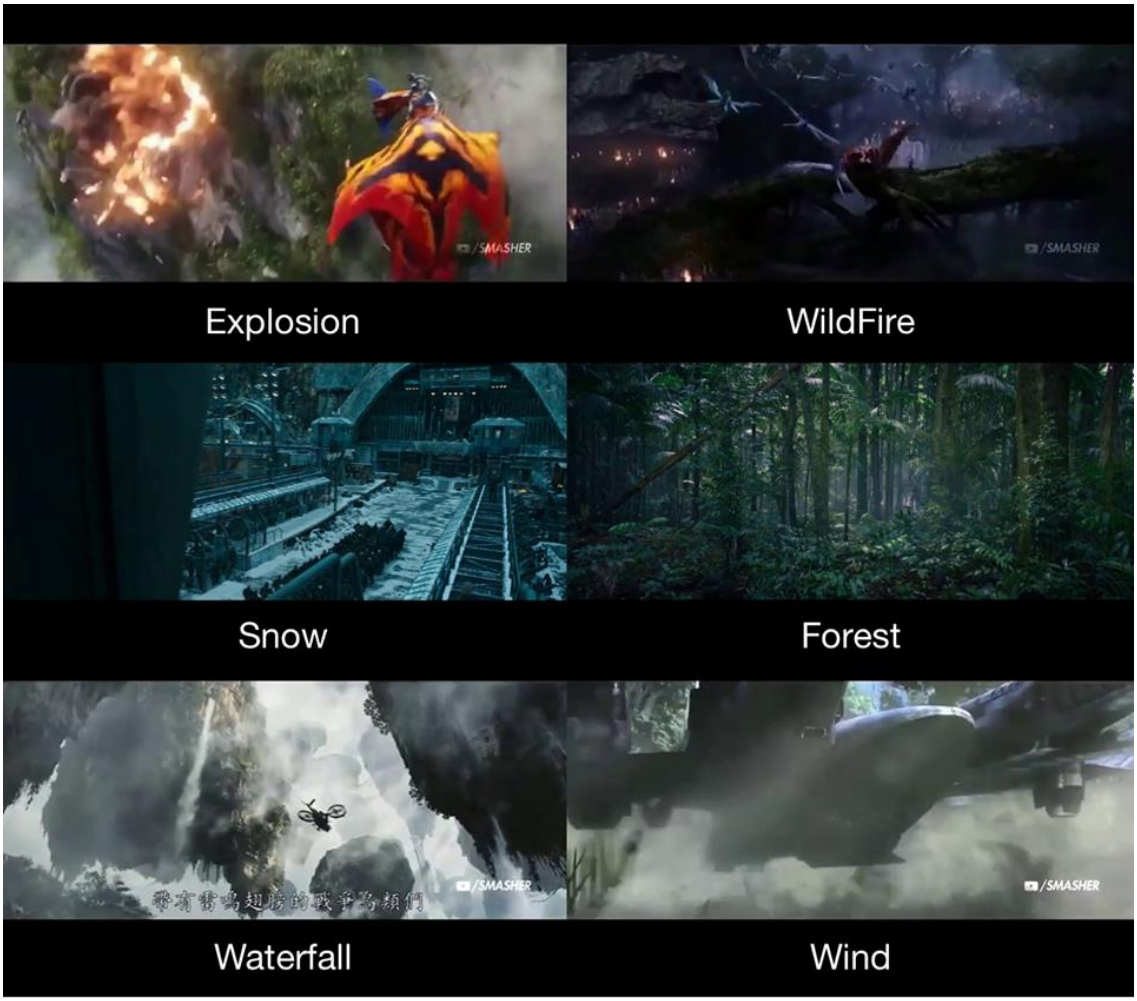


Figure 5. Examples of detected results after applying Google Cloud Vision.

Table 2. Some detected objects and their corresponding haptic modules.

Objects	4D Haptic Feedback Device					
	Heat1	**Heat2**	**Hot Air**	**Wet**	**Wind**	**Scent**
Heat	√					
Wildfire	√					
Flame	√	√				
Fire	√	√				
Explosion	√	√	√			
Waterfall				√		
Rain				√		
Snow				√	√	
winter				√	√	
Wind					√	
Wing					√	
Forest						√
Tree						√

3.2. Scene Recognition System Based on SSD

This study used the MXNet framework and the SSD deep learning algorithm to train an object detection system for animated videos. A video of our system based on the SSD deep learning algorithm is available at Reference [21]. We collected 1038 and 100 frames consisting of fire-related objects from animations as the training data set and testing data set of this study, respectively. Figure 6 displays some training data used in this study. Next, we manually marked these fire-related objects within the images with red boxes, as illustrated in Figure 6. Among these image data, we found that the colors of some fire objects in the animation were substantially different from the colors of fire in the real world. Some examples also show that the ways fires burn in animations are notably diverse (e.g., burning in concentric circles). Some stories may have irrational burning scenarios, such as a fire that burns only the face of the character but does not ignite or injure the rest of the body.

Figure 6. Some animation images among the dataset used in this study.

In this study, we applied the MobileNet-SSD model network [22,23] to perform object detection, and the image size was 512 × 512. According to the related research results in Reference [22],

the integration of MobileNet into an SSD framework can provide high-accuracy results with high detection speeds; the trained network utilized multiple layers to lower input resolution and generated predictions at various scales. The loss function of the location offset was SmoothL1, and the loss function of the classification output was cross-entropy. The initial learning rate was set to 0.001. We used a personal computer with a GTX 1080Ti graphics card in the training and testing process. The total computation time for training the MobileNet-SSD model network was approximately 30 min. After the network was trained, the animation test images were analyzed by the SSD-based object detection system. We compared the detected image areas with the ground truth to calculate the Intersection over Union (IoU) to evaluate the performance of the SSD-based object detection system. Specific details are given in the next section.

4. Experiments and Field Trial

This section describes the experimental evaluation of the design of the hardware devices and the performance of the scene recognition system.

4.1. User Experience and Field Trials of 4D Haptic Sensation Device

In this study, a total of 10 people participated in the testing of the haptic experience provided by the hardware device. The entire test process follows the ethical guidelines [24]. Figure 7 presents the actual trial scenarios. The first trial tested the intensity of the haptic sensations; all testers used the same module individually, and we then asked the testers to rate the intensity of feeling. The intensity was classified into five levels (from no feeling to intense feeling). The statistical results are listed in Table 3; most testers could clearly feel the haptic sensations provided by each device; only two testers did not clearly feel hot air or scent module sensations. Additionally, we also tested running two modules simultaneously. The tests were wind in addition to water (to create a cold feeling) and hot wind with vibration (to create an explosive feeling). According to the statistical results in Table 3, each tester clearly felt the haptic sensations provided by the device.

Figure 7. The actual trial scenarios.

Table 3. The statistical results of experience testing with various haptic modules.

Module Type	Intensity of Feeling No Feeling (1) to Intense Feeling (5)				
	1	**2**	**3**	**4**	**5**
Heat Module	0	0	1	2	7
Hot Air Module	0	2	1	4	3
Wind Module	0	0	2	5	3
Wet Module	0	0	3	1	6
Scent Module	0	2	3	2	3
Vibration Module	0	0	1	2	7
Wind & Wet Modules	0	0	1	1	8
Hot Air & Vibration Module	0	0	0	5	5

The second trial tested whether the modules were triggered at appropriate times. Testers used Sense to watch two videos, and we used our scene recognition system to analyze the results and trigger the corresponding haptic modules. The testers were asked whether the modules were triggered at appropriate times. The appropriateness was classified into five levels (irrational to very rational). The statistical results are listed in Table 4. Overall, most testers agreed that the times at which the modules were triggered appropriately. However, some users felt that the timing of the scent module was not rational.

Table 4. The test of rational time for running haptic modules.

Module Type	Rational Time for Running Haptic Module Irrational (1) to Very Rational (5)				
	1	**2**	**3**	**4**	**5**
Heat Module	0	0	1	3	6
Hot Air Module	0	0	1	3	6
Wind Module	0	0	0	4	6
Wet Module	0	0	0	4	6
Scent Module	0	1	4	3	2
Vibration Module	0	0	1	2	7
Overall Evaluation	0	0	0	5	5

After the previous two tests, we interviewed each user and asked them to rate their satisfaction with the system. The satisfaction was classified into five levels (unsatisfactory to very satisfactory); the average user satisfaction score was 4.7 (full score being 5) [25]. These test results verified that the system developed in this study provides users an enhanced experience. This system can provide users an excellent immersive experience and it really meets our design goal. Here, we compiled the following user opinions.

1. Two testers could not clearly smell the odor emitted by the scent module; therefore, we increased the concentration of essential oil in the scent module. The testers could clearly smell the odor in the retest.
2. Simultaneously running two modules provided a highly intense haptic experience. Based on the combinations of various modules, this system could create multiple levels of haptic experiences.
3. In two tests, one tester felt that the timing of the scent module seemed irrational and inappropriate. This was possibly because when the scent module was run, the background scene was a forest or a garden, but the user was focused on the story plot and therefore was not paying attention to the change in scenery. Additionally, because forest scenes appeared frequently throughout the video, causing the scent module to be run frequently, and owing to olfactory fatigue, the user was unable to clearly detect scent in the end.

4. When scenes presented low frequencies and loud sounds (e.g., explosions), all users could feel the vibrating feeling generated by the vibration module. All testers agreed that the timing of the vibration module was the most accurate and that the experience was exhilarating.

4.2. Simulation Performance of the SSD-Based Animation Detection System

To evaluate the performance of the SSD-based animation detection system, a total of 100 animation images were used for testing, and 142 fire-related objects were marked in these images. Then, these animation images were analyzed using the SSD-based object detection system. We compared the detected image areas with the ground truth to calculate IoU, and the threshold of IoU was set as 0.5 [26]. The detection result was considered as truth-positive if the value of IoU was larger than 0.5; otherwise, it was regarded as false positive. Subsequently, we obtained the area under the curve (AUC) according to the recall and precision rates, Figure 8 illustrates the AUC results for some fire-related objects. The *x*-axis indicates the false positive rate, and the *y*-axis represents the true positive rate. Finally, the AP was 0.802 for 142 fire-related objects; the simulation results indicated that the proposed method could accurately detect fire-related objects in animation videos. The SSD-based system could accurately detect fire-related objects to illustrate that this is a possible way to detect objects in animation videos. However, one point should be emphasized. The color or diverse objects in the animation were varied and substantially different from the objects in the real world. Besides, there are no open datasets of animation objects for training. The result is that collecting representative and sufficient training patterns is a challenging task, and it is also the reason why we only trained the proposed SSD-based system with fire-related objects.

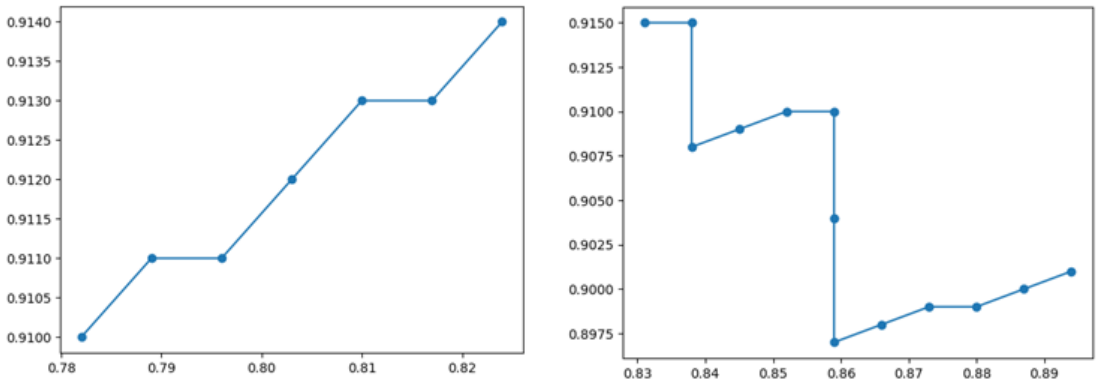
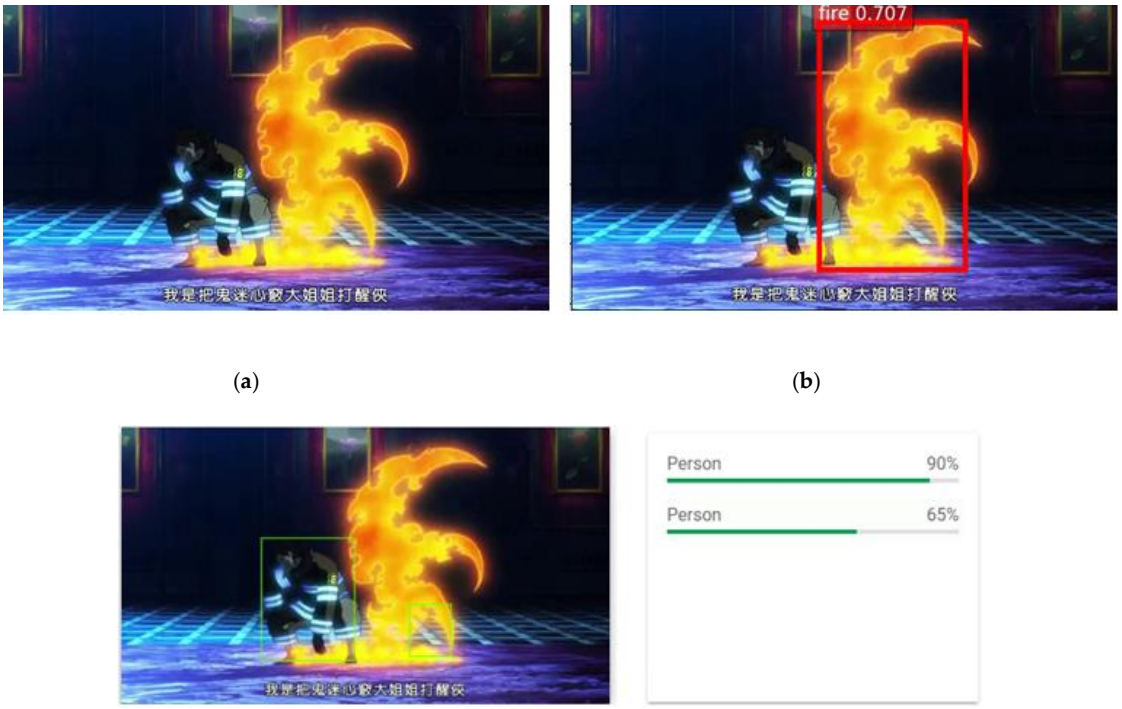


Figure 8. AUC results of some fire-related objects.

4.3. Comparison SSD with Google Cloud Vision

In this study, a 64-s animated video was used as the test video to compare Google Cloud Vision with the proposed SSD-based animation detection system. Sixty-four animation images were captured as test images from the animation video during each second. Twenty-five images consisted of the fire-like objects, and 39 images had no fire-like objects. Figure 9 presents some test animation images. This experiment examined whether two systems could accurately detect fire-like objects in the animation images. Figure 10 gives an example of a testing result. Figure 10a is the test animation image. Figure 10b shows that the proposed system could accurately detect the fire-like object. As can be seen from the testing result of Google Cloud Vision in Figure 10c, none of fire-like objects could be detected, but two 'person' objects were misdetected. In the experimental results of the entire animation images, both systems will not detect any fire-like object if the animation images did not really include any fire-like objects. However, for 25 animation images consisting of fire-like objects, Google Cloud Vision was unable to detect any fire-like objects, but the proposed SSD-based system was able to detect fire-like objects in most animation images (accuracy rate: 84%). These experiment results illustrate that our system provides a better user experience with animation than using Google Cloud Vision.

Figure 9. Sixty-four test animation images.



(**a**) (**b**)

(**c**)

Figure 10. (**a**) The test animation image; (**b**) the testing result of the proposed SSD-based system; (**c**) the testing result of Google Cloud Vision.

5. Conclusions

In this study, we designed a 4D audiovisual entertainment system to provide haptic sensations for users when they watch movies and animations at home. The 4D sensation device contains six types of haptic modules. These modules can be processed individually or together to provide various haptic sensations experiences. Most testers could clearly feel the haptic sensations provided by each device. Based on the combinations of various modules, this system could create multiple levels of haptic experiences. Additionally, this study used the MXNet framework and the SSD deep learning algorithm to train an object detection system for animated videos. The experiment results illustrate that the SSD-based system provides a better user experience with animation than using Google Cloud Vision.

Although the proposed 4D sensation device can provide great haptic sensations for users when they watch movies and animations at home, there are some limitations of this device. Because the

designed concept is portable and requires low power consumption and low cost, the optimal experience area of this haptic device might be limited. To serve more users, we could use more sensation devices to enlarge the effective experience area. Due to the characteristics (portable, low power consumption, and low cost) of the proposed 4D sensation device, this is an intuitive and easy way of serving more users. Besides, if we use the vibration speaker as the vibration module, the vibration experience area will be limited on the tabletop. This represents that some vibration feelings (e.g., an earthquake) could not be simulated. To provide a better sensation experience, the kinesthetic device will be studied and developed in our future work.

Author Contributions: All authors contributed to the paper. C.-H.C., Y.-S.S., C.-J.H., K.-C.L., and P.-H.H. collected and organized data, and wrote the manuscript. Finally, Y.-S.S. acted as a corresponding author. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Science and Technology (MOST), Taiwan, R.O.C., under MOST 108-2221-E-032-029-MY2, MOST 109-2511-H-019-001, MOST 109-2511-H-019-004 -MY2.

Conflicts of Interest: Authors declare no conflict of interest.

Ethics Committee Approval: All participants agreed to participate in this research. Ethical approval was not required for the study on them by the local legislation and institutional requirements. The experimental data they provided was anonymous and would not be of any commercial use.

References

1. Gonçalves, G.; Melo, M.; Vasconcelos-Raposo, J.; Bessa, M.E. Impact of different sensory stimuli on presence in credible virtual environments. *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 3231–3240. [CrossRef]
2. Chen, Y.-S.; Han, P.-H.; Hsiao, J.-C.; Lee, K.-C.; Hsieh, C.-E.; Lu, K.-Y.; Chou, C.-H.; Hung, Y.-P. SoEs: Attachable Augmented Haptic on Gaming Controller for Immersive Interaction. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, Japan, 16–19 October 2016; pp. 71–72.
3. Han, P.H.; Chen, Y.S.; Yang, K.T.; Chuan, W.S.; Chang, Y.T.; Yang, T.M.; Lin, J.Y.; Lee, K.C.; Hsieh, C.E.; Lee, L.C.; et al. BoEs: Attachable Haptics Bits On Gaming Controller For Designing Interactive Gameplay. In Proceedings of the ACM SIGGRAPH Asia 2017 VR Showcase, Bangkok, Thailand, 27–30 November 2017; Article No. 3. pp. 1–2.
4. Han, P.-H.; Hsieh, C.-E.; Chen, Y.-S.; Hsiao, J.-C.; Lee, K.-C.; Ko, S.F.; Chen, K.W.; Chou, C.-H.; Hung, Y.-P. AoEs: Enhancing teleportation experience in immersive environment with mid-air haptics. In Proceedings of the ACM SIGGRAPH 2017 Emerging Technologies, Los Angeles, CA, USA, 30 July–3 August 2017; Article No. 3. pp. 1–2.
5. Günther, S.; Müller, F.; Schön, D.; Elmoghazy, O.; Mühlhäuser, M.; Schmitz, M. Therminator: Understanding the Interdependency of Visual and On-Body Thermal Feedback in Virtual Reality. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–14.
6. Sasaki, T.; Liu, K.-H.; Hasegawa, T.; Hiyama, A.; Inami, M. Virtual Super-Leaping: Immersive Extreme Jumping in VR. In Proceedings of the 10th Augmented Human International Conference 2019 (AH2019), Reims, France, 11–12 March 2019; Article No. 18. pp. 1–8.
7. TensorFlow Object Detection API. Available online: https://github.com/TensorFlow/models/tree/master/research/object_detection (accessed on 24 November 2019).
8. Google Cloud Vision. Available online: https://cloud.google.com/vision/docs/2018 (accessed on 8 September 2018).
9. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *arXiv* **2015**, arXiv:1506.01497. [CrossRef] [PubMed]
10. Bergman, D.L. Symmetry Constrained Machine Learning. In Proceedings of the 2019 Intelligent Systems and Applications Conference, Las Palmas, Spain, 7–12 January 2019; pp. 501–512.
11. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
12. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
13. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

14. Xu, K.; Wang, X.; Liu, X.; Cao, C.; Li, H.; Peng, H.; Wang, D. A dedicated hardware accelerator for real-time acceleration of YOLOv2. *J. Real Time Image Process.* **2020**. [CrossRef]
15. Maher, A.; Taha, H.; Zhang, B. Realtime multi-aircraft tracking in aerial scene with deep orientation network. *J. Real Time Image Process.* **2018**, *15*, 495–507. [CrossRef]
16. Chen, T.; Li, M.; Li, Y.; Lin, M.; Wang, N.; Wang, M.; Xiao, T.; Xu, B.; Zhang, C.; Zhang, Z. MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems. In Proceedings of the Advances in Neural Information Processing System, Workshop on Machine Learning System, Barcelona, Spain, 5–10 December 2016.
17. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
18. Ko, S.F.; Lin, Y.H.; Han, P.H.; Chang, C.C.; Chou, C.H. Combining Deep Learning Algorithm with Scene Recognition and Haptic Feedback for 4D–VR Cinema. In Proceedings of the ACM SIGGRAPH Asia 2018, Tokyo, Japan, 4–7 December 2018; Article No. 18. pp. 1–2.
19. Fusion 360. Available online: https://www.autodesk.com/products/fusion-360/overview (accessed on 19 October 2017).
20. Operating Video of Our System Based on Google Could Vision. Available online: https://youtu.be/r-iIezN6quI (accessed on 25 August 2020).
21. Operating Video of Our System Based on SSD Deep Learning Algorithm. Available online: https://youtu.be/2-XtgCypyYA (accessed on 25 August 2020).
22. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
23. Su, Y.S.; Chou, C.H.; Chu, Y.L.; Yang, Z.F. A finger-worn device for exploring Chinese printed text with using CNN algorithm on a micro IoT processor. *IEEE ACCESS* **2019**, *7*, 116529–116541. [CrossRef]
24. Su, Y.S.; Lin, C.L.; Chen, S.Y.; Lai, C.F. Bibliometric study of social network analysis literature. *Libr. Hi Tech.* **2019**, *38*, 420–433. [CrossRef]
25. Su, Y.S.; Chen, H.R. Social Facebook with Big Six approaches for improved students' learning performance and behavior: A case study of a project innovation and implementation course. *Front. Psychol.* **2020**, *11*, 1166. [CrossRef]
26. Su, Y.S.; Ni, C.F.; Li, W.C.; Lee, I.H.; Lin, C.P. Applying deep learning algorithms to enhance simulations of large-scale groundwater flow in IoTs. *Appl. Soft Comput.* **2020**, *92*, 106298. [CrossRef]