International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2020)
F. Argelaguet, M. Sugimoto, and R. McMahan (Editors)

FrictGAN: Frictional Signal Generation from Fabric Texture Images using Generative Adversarial Network

Shaoyu Cai[1,3] Yuki Ban[2] Takuji Narumi[3] and Kening Zhu[†1]

[1]School of Creative Media, City University of Hong Kong, Hong Kong
[2]Graduate School of Frontier Sciences, The University of Tokyo, Japan
[3]Graduate School of Information Science and Technology, The University of Tokyo, Japan

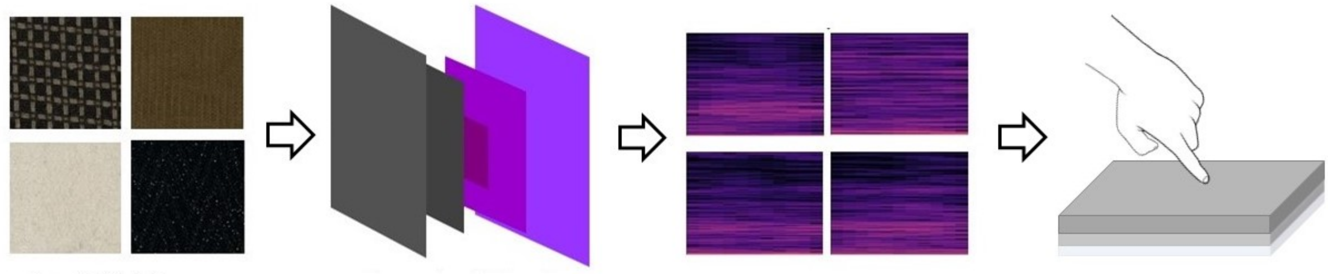

Figure 1: *The main concept of frictional signals generation using the GAN-based method. Users could input the RGB fabric images to the Generator Network, and it will synthesize the frictional signals that could be applied on the electrostatic tactile display to simulate the haptic texture surfaces of the input fabrics.*

Abstract

The electrostatic tactile display could render the tactile feeling of different haptic texture surfaces by generating the frictional force through voltage modulation when a finger is sliding on the display surface. However, it is challenging to prepare and fine-tune the appropriate frictional signals for haptic design and texture simulation. We present FrictGAN, a deep-learning-based framework to synthesize frictional signals for electrostatic tactile displays from fabric texture images. Leveraging GANs (Generative Adversarial Networks), FrictGAN could generate the displacement-series data of frictional coefficients for the electrostatic tactile display to simulate the tactile feedback of fabric material. Our preliminary experimental results showed that FrictGAN could achieve considerable performance on frictional signal generation based on the input images of fabric textures.

CCS Concepts

• ***Computing methodologies*** → ***Generative adversarial network;*** • ***Human-centered computing*** → ***Virtual reality;***

1. Introduction

Recently, due to the development of haptics, various types of haptic feedback are rendered on different haptic devices, such as electrostatic feedback [BPIH10], vibrotactile feedback [Gal12], and thermal feedback [ZPC*19] [NKZ20]. These devices could be integrated with Virtual Reality (VR) to improve the immersion and realness in virtual environments [ZCHW19] [CKNZ20] [CWZ18].

† Corresponding author: keninzhu@cityu.edu.hk

Electrostatic tactile displays [BPIH10] could provide frictional tactile feedback on the fingertips of users, to simulate different surface properties, such as the roughness of fabric textures. This is usually achieved by modulating the lateral force on the touch surface. This technique could be potentially applied for texture simulation and rendering in VR [ZLY*20].

However, the process of designing the appropriate signal for the frictional force between the finger and electrostatic tactile surface could be challenging due to two main reasons. Firstly, the lateral

DOI: 10.2312/egve.20201254

forces are usually dependent upon the frictional properties on the real physical texture surface, such as the frictional coefficient, the normal pressure force, and the displacements. However, many texture surfaces show irregular changes in lateral force during the finger sliding. This makes it difficult to build a general model for the frictional simulation and prediction for a textured surface. Secondly, while it could be feasible to use the recorded data to modulate tactile feedback on the electrostatic tactile display, it is often time-consuming for data collection.

To eliminate these gaps, we propose FrictGAN, a signal-generation framework based on Generative Adversarial Networks (GANs) [GPAM*14]. We defined the frictional signals as the frictional coefficients on the finger-contact positions of the electrostatic tactile surface. We converted the frictional coefficients to the displacement-frequency of wave-format, which could be treated as the 2D images. We then trained a conditional-GAN-based deep neural network [IZZE17], to generate the amplitude spectrogram for representing the frictional coefficients according to the finger displacements on the fabric surface. Fig. 1 shows the flow of FrictGAN, which takes the RGB image of the textured surface and generates the amplitude spectrogram of the frictional signals to represent the frictional coefficients of the surface. The series of frictional coefficients further controls the electrostatic tactile display to control the lateral force corresponding to the finger displacement. Our preliminary experimental results showed that FrictGAN could generate a series of considerable amplitude spectrograms in the displacement-frequency domain. Our implementation of source code and data set are available at: `https://github.com/shaoyuca/FrictGAN`.

2. Related Work

With the emergence of electrostatic/electrovibrational tactile surfaces [BPIH10], data-driven texture modeling and rendering has attracted an increasing amount of research interests. Osgouei et al. [OSKC18] proposed an inverse dynamics model for generating the actuation signals to mimic real textures on an electrovibration display. As an extended work, the same group [OKC20] conducted user studies to compare the rendered lateral forces with the real signals, which showed the approach could improve the quality of virtual texture rendering on the electrovibration display. Jiao et al. presented HapTex [JZW*19] to retrieve the friction coefficients from the recording frictional data while a finger sliding on the textured surface. While the aforementioned works could achieve considerable performance on virtual texture rendering with pre-recorded frictional signals, they often required the ready signals. However, the signal-recording procedure could be high-cost and time-consuming.

Recently, generative models showed high performance on high-dimensional data (e.g., images or videos) generation. Several prior research works showed conditional GANs owned effective generation performances on cross-modal contents, such as from images [NHS*19], texts [ZXL*17], and audio [CSDX17] to images. As a preliminary attempt, Ujitoko and Ban [UB18] presented TactGAN, for vibrotactile signal generation from the textured-surface images or label attributes using GANs. Based on the dataset created by Strese et al. [SSIS16], TactGAN defined the acceleration-based vibrotactile signals as the amplitude spectrogram, which is a 2D image and trained a GAN-based pipeline to synthesize the spectrogram from the RGB texture images or the material attributes. Similarly, with the same dataset, Li et al. [LLZS19] proposed a GAN-based method to learn visual-tactile representation by identifying the category labels of input images to guide tactile signals generation. Liu et al. [LGZ*20] built image-to-tactile cross-modal perception based on CycleGAN [ZPIE17] framework and developed a portable device for visually impaired people. Our work shows the feasibility of generating frictional signals from the texture images using GANs.

3. Method and Implementation

3.1. FrictGAN Architectures

As shown in Fig. 2, FrictGAN consists of a generator G and a discriminator D commonly adopted in the pix2pix-based model [IZZE17]. The generator G contains two parts, an encoder network and a decoder network. The encoder network takes the size of $1024 \times 1024 \times 3$ RGB image as input, and output a 128-dimensional latent vector. The decoder could synthesize the size of 257×11 amplitude spectrum from the latent space that includes the concatenation of the latent vector and a random 50-dimensional noise vector. We adopted the structure of conditional GAN [MO14] and passed spectrograms together with the input RGB image data to the discriminator D to distinguish the real and fake/generated amplitude spectrograms based on the conditional input. The final generated spectrogram is converted into the wave format by the Griffin-Lin algorithm [GL84].

We adopted Convolutional Neural Networks (CNN) in both the generator and the discriminator. For the generator G, we implemented the U-net structure [RFB15] to build a series of skip connections between the layer of encoder and decoder networks. For each layer in the generator, we adopted convolution/deconvolution, batch normalization, and ReLU units; three dropout units were added in the last three layers of the encoder. To generate the corresponding output size of spectrograms, we proposed a well-designed encoder-decoder structure through down-/up-sampling manipulation in our architecture. After the encoder and decoder processing, we added a ReLU function in the last layer of generator to generate the final outputs. For the discriminator D, we used the PatchGAN structure [IZZE17], and input a channel-wise concatenated vector with the size of $257 \times 11 \times 4$ from cropped texture image and generated/real amplitude spectrum and output a $30 \times 1 \times 1$ patch. Each layer in the discriminator consists of convolution, batch normalization, and LeakyReLU units. After finishing the training procedure, we removed the discriminator and only used the generator to synthesize the final amplitude spectrogram and converted it to the wave-format frictional signals for rendering on the electrostatic tactile display.

3.2. Objective Functions

We adopted a conditional GAN to learn the mapping from the image data x and random noise vector z to the amplitude spectrogram y. The objective mapping of the model is as below:

$$G : x, z \rightarrow y. \quad (1)$$

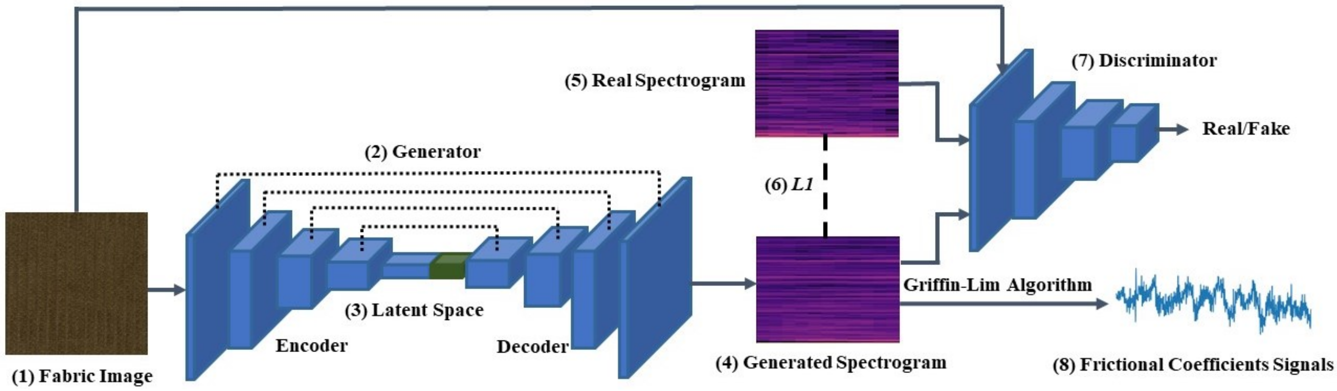

Figure 2: *The Network structure of FrictGAN. The left part (1) is a sample of training fabric texture images passed into the Generator (2) composed by the Encoder and Decoder parts. The Generator (2) is a U-net-based structure with skip connections between the Encoder and Decoder layers. We added a 50-dimensional random noise vector (green) in the latent space (3) to increase variance for the Generator training. We concatenated the input RGB fabric image (1) and the spectrogram data (4) and (5) to the Discriminator (7), which will guide the Generator (2) to synthesis the final generated spectrum (4). We also included the $L1$ loss (6) between the real and generated spectrum. After the training stage, we removed the Discriminator (7) and only fed the image data to the Generator (2) to generate the spectrogram and used Griffin-Lim algorithm [GL84] to convert the spectrogram data to the waveform frictional coefficients signals (8) for electrostatic tactile displays rendering.*

In the original GAN [GPAM*14], the proposed objective function may cause gradient vanishing [AB17]. To solve the problem, we implemented the Wasserstein GAN (WGAN) [ACB17] for more stable generator G and discriminator D training, so we replaced the original GAN loss as the Wasserstein GAN loss L_{WGAN}. The objective function of WGAN is:

$$L_{WGAN} = -\mathbb{E}_{x\sim p_{(x)},y\sim p_{(y)}}[D(y|x)] + \mathbb{E}_{x\sim p_{(x)},\tilde{y}\sim p_{(\tilde{y})}}[D(\tilde{y}|x)] \quad (2)$$

We also included the Manhattan distance, which is $L1$ distance, as our pixel-wised loss, so the loss function becomes:

$$L = L_{WGAN} + \lambda L_{L1} \quad (3)$$

In this equation, L is our final objective loss, and L_{L1} is the pixel-wised $L1$ loss between the real and generated spectrograms. λ is the hyperparameter, which was set to 100 in our preliminary experiments.

4. Experiments and Results

We conducted several experiments to measure and evaluate the FrictGAN model properties for frictional signals outputs from fabric texture images. The experimental results showed that our FrictGAN model could generate frictional coefficients data closed to the pre-recorded data from the real fabric surfaces in the displacement-frequency tactile spectrum.

We used HapTex dataset [JZW*19], which includes the visual data and the frictional data for 120 types of fabrics. The frictional data were recorded while a user sliding his/her finger on the real fabric textures. The dataset presents the relationships between the frictional coefficients and the displacements with wave formats for each fabric surface. To this end, we computed the size of 257 × 11 amplitude spectrogram data from wave-format frictional coefficients signals using Short-time Fourier transform (STFT) with a 512-hamming window and a 128-hop size and took the RGB fabric texture images as the visual domain for the input of the FrictGAN.

In our preliminary experiments, five types of fabrics (see Fig. 3) were selected for training to verify the feasibility of FrictGAN. We adopted the following data-augmentation strategies on our image and spectrogram data. Firstly, we used a size of 1024 × 1024 sliding window to move horizontally and crop the 2362 × 2362 original fabric image with the offset of two pixels, so finally we acquired 669 RGB images for each fabric. Referring to the corresponding information between image pixels and frictional coefficients data in [JZW*19], we also implemented a similar manipulation on frictional coefficients data according to the sliding-window movements on RGB images to build one-to-one corresponding relations between images and amplitude spectrograms. Considering the displacement range of frictional coefficients data during the data collection stage, we chose the previous 600 pairs of RGB images and frictional coefficients signals. We then converted the frictional coefficients signals to amplitude spectrograms through STFT for each type of fabric material. We then randomly selected 100 pairs of RGB images and spectrograms for the validation set and the same amount of data for the testing set. Thus, we totally acquired 2000, 500, 500 pairs of RGB images and spectrograms for training, validation, and testing databases, separately.

We implemented our networks with the Tensorflow framework on an Nvidia Geforce GTX 2080 Ti GPU. We used the RMSprop optimizer with a batch size of 2, and both of the learning rates for the generator and discriminator were 5e−5 in our experiment. Fig. 3 (a) shows 5 kinds of input fabric texture RGB images, Fig. 3 (b) shows generated amplitude spectrograms, and Fig. 3 (c) shows

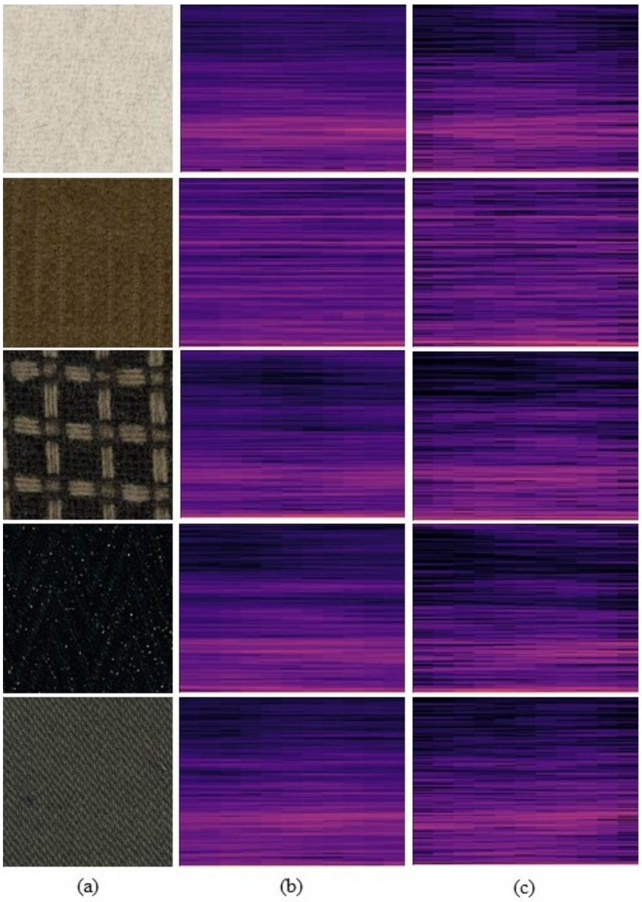

Figure 3: *The preliminary results of FrictGAN. (a) The input texture images of fabrics; (b) Generated spectrograms based on FrictGAN from input fabric texture images; (c) The ground truth (real) spectrograms*

the ground truth spectrum. We then used the Griffin-Lim algorithm [GL84] to convert the frequency-domain spectrograms to the wave-format signals for the lateral forces rendering on the electrostatic tactile display. We also computed Mean Squared Error (MSE) values with reference frictional coefficients for each type of fabric individually. The MSE values of these 5 kinds of fabrics (from top to down in Fig. 3) were 0.006, 0.005, 0.014, 0.005 and 0.005, separately, averagely 0.035.

5. Limitations and Future Work

We also identify some limitations in our preliminary experiments. Currently, we only tested five kinds of fabrics for our FrictGAN model; the diversity of generated frictional signals is restricted, limiting the generalization and scalability of the FrictGAN model. We will include more types of fabrics in our dataset for training in the future. In addition, for the waveform signals, only calculating the MSE values could not support the acceptable results from the FrictGAN model, so we intend to conduct ablation studies among different generative models by evaluating some parameters, such as Geometry Score [KO18]; or compare with previous frictional force generative models [JWV18], to quantitatively investigate the performance of different methods.

As electrostatic tactile displays need to modulate the driven voltages for rendering the lateral force, as our future work, we will add an additional data-driven regression model to predict the correlation coefficients between the driven voltages and the frictional coefficients. This is to render high-fidelity tactile feedback for the virtual fabric surfaces. Therefore, we will conduct user studies to collect human users' subjective ratings on the realness of the generated frictional signals.

In addition, our method not only reproduces haptic textures from known subjects but also allows users to create new haptic textures from real physical textures or materials. Some related works [HAJ19] [UBH20] proved the capability to produce or render new haptic surfaces using machine learning methods. In our future work, we will explore the feasibility of creating new virtual fabric textures with haptic properties from real-life textures.

Last but not least, VR is one of the most important applications for haptic rendering. In the future, we will integrate the electrostatic tactile display and the FrictGAN model with VR devices to improve the immersive experience of the simulated textures in the virtual environment.

6. Conclusion

In this paper, we present FrictGAN, a deep-learning-based generative method for cross-modal fictional-signal synthesis. FrictGAN was designed based on the GAN structure, taking the RGB fabric images as the input and the distribution of the frictional coefficients of the fabric surface as the output. We also conducted some preliminary experiments to evaluate the effectiveness of our model, and the results showed that FrictGAN could generate visually similar amplitude spectrograms compared to the real spectrum.

7. Acknowledgements

We thank Dr. Yusuke Ujitoko for his helpful discussion in the initial experiment. This research was partially supported by the Young Scientists Scheme of the National Natural Science Foundation of China (Project No. 61907037), the Applied Research Grant (Project No. 9667189), and the Centre for Applied Computing and Interactive Media (ACIM) of School of Creative Media, City University of Hong Kong.

References

[AB17] Arjovsky M., Bottou L.: Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862* (2017). 3

[ACB17] Arjovsky M., Chintala S., Bottou L.: Wasserstein gan. *arXiv preprint arXiv:1701.07875* (2017). 3

[BPIH10] Bau O., Poupyrev I., Israr A., Harrison C.: Teslatouch: electrovibration for touch surfaces. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology* (2010), pp. 283–292. 1, 2

[CKNZ20] CAI S., KE P., NARUMI T., ZHU K.: Thermairglove: A pneumatic glove for thermal perception and material identification in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2020), IEEE, pp. 248–257. 1

[CSDX17] CHEN L., SRIVASTAVA S., DUAN Z., XU C.: Deep cross-modal audio-visual generation. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017* (2017), pp. 349–357. 2

[CWZ18] CHEN T., WU Y.-S., ZHU K.: Investigating different modalities of directional cues for multi-task visual-searching scenario in virtual reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (2018), pp. 1–5. 1

[Gal12] GALAMBOS P.: Vibrotactile feedback for haptics and telemanipulation: Survey, concept and experiment. 1

[GL84] GRIFFIN D., LIM J.: Signal estimation from modified short-time fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing 32*, 2 (1984), 236–243. 2, 3, 4

[GPAM*14] GOODFELLOW I., POUGET-ABADIE J., MIRZA M., XU B., WARDE-FARLEY D., OZAIR S., COURVILLE A., BENGIO Y.: Generative adversarial nets. In *Advances in neural information processing systems* (2014), pp. 2672–2680. 2, 3

[HAJ19] HASSAN W., ABDULALI A., JEON S.: Authoring new haptic textures based on interpolation of real textures in affective space. *IEEE Transactions on Industrial Electronics 67*, 1 (2019), 667–676. 4

[IZZE17] ISOLA P., ZHU J.-Y., ZHOU T., EFROS A. A.: Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 1125–1134. 2

[JWV18] JANKO M., WIERTLEWSKI M., VISELL Y.: Contact geometry and mechanics predict friction forces during tactile surface exploration. *Scientific reports 8*, 1 (2018), 1–10. 4

[JZW*19] JIAO J., ZHANG Y., WANG D., GUO X., SUN X.: Haptex: A database of fabric textures for surface tactile display. In *2019 IEEE World Haptics Conference (WHC)* (2019), IEEE, pp. 331–336. 2, 3

[KO18] KHRULKOV V., OSELEDETS I.: Geometry score: A method for comparing generative adversarial networks. *arXiv preprint arXiv:1802.02664* (2018). 4

[LGZ*20] LIU H., GUO D., ZHANG X., ZHU W., FANG B., SUN F.: Toward image-to-tactile cross-modal perception for visually impaired people. *IEEE Transactions on Automation Science and Engineering* (2020). 2

[LLZS19] LI X., LIU H., ZHOU J., SUN F.: Learning cross-modal visual-tactile representation using ensembled generative adversarial networks. *Cognitive Computation and Systems 1*, 2 (2019), 40–44. 2

[MO14] MIRZA M., OSINDERO S.: Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014). 2

[NHS*19] NAKANO K., HORITA D., SAKATA N., KIYOKAWA K., YANAI K., NARUMI T.: Deeptaste: Augmented reality gustatory manipulation with gan-based real-time food-to-food translation. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2019), IEEE, pp. 212–223. 2

[NKZ20] NASSER A., KENG K.-N., ZHU K.: Thermalcane: Exploring thermotactile directional cues on cane-grip for non-visual navigation. In *Proceedings of The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)* (2020), pp. 1–12. 1

[OKC20] OSGOUEI R. H., KIM J. R., CHOI S.: Data-driven texture modeling and rendering on electrovibration display. *IEEE Transactions on Haptics 13*, 2 (2020), 298–311. 2

[OSKC18] OSGOUEI R. H., SHIN S., KIM J. R., CHOI S.: An inverse neural network model for data-driven texture rendering on electrovibration display. In *2018 IEEE Haptics Symposium (HAPTICS)* (2018), IEEE, pp. 270–277. 2

[RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (2015), Springer, pp. 234–241. 2

[SSIS16] STRESE M., SCHUWERK C., IEPURE A., STEINBACH E.: Multimodal feature-based surface material classification. *IEEE transactions on haptics 10*, 2 (2016), 226–239. 2

[UB18] UJITOKO Y., BAN Y.: Vibrotactile signal generation from texture images or attributes using generative adversarial network. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications* (2018), Springer, pp. 25–36. 2

[UBH20] UJITOKO Y., BAN Y., HIROTA K.: Gan-based fine-tuning of vibrotactile signals to render material surfaces. *IEEE Access 8* (2020), 16656–16661. 4

[ZCHW19] ZHU K., CHEN T., HAN F., WU Y.-S.: Haptwist: creating interactive haptic proxies in virtual reality using low-cost twistable artefacts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–13. 1

[ZLY*20] ZHAO L., LIU Y., YE D., MA Z., SONG W.: Implementation and evaluation of touch-based interaction using electrovibration haptic feedback in virtual environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2020), pp. 239–247. 1

[ZPC*19] ZHU K., PERRAULT S., CHEN T., CAI S., PEIRIS R. L.: A sense of ice and fire: Exploring thermal feedback with multiple thermoelectric-cooling elements on a smart ring. *International Journal of Human-Computer Studies 130* (2019), 234–247. 1

[ZPIE17] ZHU J.-Y., PARK T., ISOLA P., EFROS A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2223–2232. 2

[ZXL*17] ZHANG H., XU T., LI H., ZHANG S., WANG X., HUANG X., METAXAS D. N.: Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 5907–5915. 2