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

Shaoyu Cai\textsuperscript{1,3} Yuki Ban\textsuperscript{2} Takuji Narumi\textsuperscript{3} and Kening Zhu\textsuperscript{1,\textdagger}

\textsuperscript{1}School of Creative Media, City University of Hong Kong, Hong Kong
\textsuperscript{2}Graduate School of Frontier Sciences, The University of Tokyo, Japan
\textsuperscript{3}Graduate School of Information Science and Technology, The University of Tokyo, Japan

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
\begin{itemize}
  \item Computing methodologies → Generative adversarial network;
  \item Human-centered computing → Virtual reality;
\end{itemize}

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\textsuperscript{*}19] [NKZ20]. These devices could be integrated with Virtual Reality (VR) to improve the immersion and realism in virtual environments [ZCHW19] [CKNZ20] [CWZ18].

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\textsuperscript{*}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
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 electrovibrational 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 real 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 [CSDX'17] 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 Stree 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 $257 \times 11$ 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-Lim 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:

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

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

(2)

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

$$L = L_{WGAN} + \lambda L_{L1}$$

(3)

In this equation, $L$ is our final objective loss, and $L_{L1}$ is the pixel-wise $L1$ loss between the real and generated spectrograms. $\lambda$ 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 $257 \times 11$ 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 $1024 \times 1024$ sliding window to move horizontally and crop the $2362 \times 2362$ 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


© 2020 The Author(s)
Eurographics Proceedings © 2020 The Eurographics Association


