Vis2Hap: Vision-based Haptic Rendering by Cross-modal Generation

Guanqun Cao[1], Jiaqi Jiang[2], Ningtao Mao[3], Danushka Bollegala[1], Min Li[4], and Shan Luo[2]

***Abstract*— To assist robots in teleoperation tasks, haptic rendering which allows human operators access a virtual touch feeling has been developed in recent years. Most previous haptic rendering methods strongly rely on data collected by tactile sensors. However, tactile data is not widely available for robots due to their limited reachable space and the restrictions of tactile sensors. To eliminate the need for tactile data, we propose a novel method named as *Vis2Hap* to generate haptic rendering from visual inputs that can be obtained from a distance without physical interaction. We take the surface texture of objects as key cues to be conveyed to the human operator. To this end, a generative model is designed to simulate the roughness and slipperiness of the object's surface. To embed haptic cues in Vis2Hap, we use height maps from tactile sensors and spectrograms from friction coefficients as the intermediate outputs of the generative model. Once Vis2Hap is trained, it can be used to generate height maps and spectrograms of new surface textures, from which a friction image can be obtained with our proposed haptic rendering algorithm and displayed on a haptic display.**

I. INTRODUCTION

Haptic rendering has been developed to assist robots in teleoperation tasks recently, which allows human operators access to a virtual touch feeling. Most previous haptic rendering methods are limited to reproducing the properties of an object's surface by using tactile signals. To collect tactile signals, the teleoperation robot is required to contact the target object physically. However, a robot's attempts to reach every object are inefficient and time-consuming. Therefore, it is desirable to provide human operators with haptic rendering without requiring tactile signals.

To address the above problem, we propose a haptic rendering framework *Vis2Hap* based on a cross-modal generation model that uses visual images, which can be obtained from a distance, to generate the signals for haptic rendering. The touch feeling of the surface texture largely depends on two aspects, i.e., ***roughness*** and ***slipperiness*** [1]. Roughness refers to the difference in the height of high-frequency variation on an object's surface while slipperiness is related to surface resistance. In our Vis2Hap framework, the height maps of object's surface, which demonstrate the roughness of the object's surfaces, and spectrograms of dynamic friction

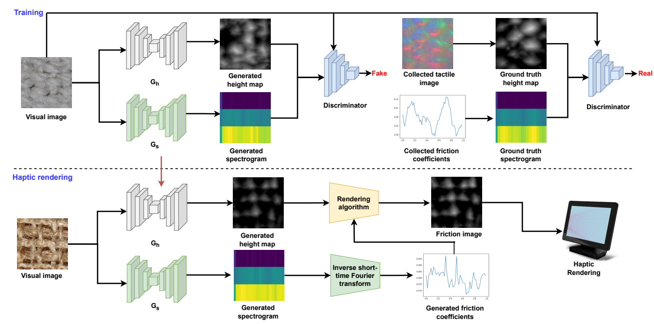

Fig. 1: **The proposed *Vis2Hap* haptic rendering framework.**

coefficients, which can assess the slipperiness, are generated from vision as the intermediate outputs. Then, we combine the generated signals to obtain friction images as the input to the haptic display for haptic rendering.

II. METHODOLOGIES

As shown in Fig. 1, we develop our Vis2Hap framework based on a conditional Generative Adversarial Network (cGAN) that generates tactile signals of friction coefficients and height maps of the object's surface from visual images as intermediate outputs. The obtained signals are then utilised to create friction images for haptic rendering.

Height maps and friction coefficients generation: Our generative model consists of two generators G_s and G_h as well as a discriminator D, as shown in the training phase of Fig. 1. The visual images $\{x_i\}_{i=1}^N$, tactile images $\{t_i\}_{i=1}^N$ and friction coefficients data $\{f_i\}_{i=1}^N$ are used to train the generative model. Specifically, friction coefficients data over different locations, considered as time series, is converted into 2D spectrograms $\{f_i\}_{i=1}^N \rightarrow \{s_i\}_{i=1}^N$, which can illustrate the pattern of coefficients change in time-frequency domain effectively, using Short-Time Fourier Transform (STFT). The height maps of the object's surface are reconstructed from tactile images $\{t_i\}_{i=1}^N \rightarrow \{h_i\}_{i=1}^N$ by using a photometric stereo algorithm [2] to provide the ground truth for training the generator.

For the training of the generative model, we optimise the generators and discriminators iteratively. Concretely, the generators are trained to generate synthetic height maps and spectrograms to fool the discriminator by minimising:

$$\mathcal{L}_G(G_s, G_h) = -\mathbb{E}_x[\log(D(x, G_s(x), G_h(x)))]. \quad (1)$$

At the same time, D is trained by minimising:

$$\begin{aligned}\mathcal{L}_D(D) = &- \mathbb{E}_{x,s,h}[\log D(x, s, h)] \\ &- \mathbb{E}_x[\log(1 - D(x, G_s(x), G_h(x)))].\end{aligned} \quad (2)$$

[1]G. Cao and D. Bollegala are with the Department of Computer Science, University of Liverpool, Liverpool L69 3BX, United Kingdom. Emails: {g.cao, danushka}@liverpool.ac.uk.

[2]J. Jiang and S. Luo are with the Department of Engineering, King's College London, London WC2R 2LS, United Kingdom. E-mails: {jiaqi.1.jiang, shan.luo}@kcl.ac.uk.

[3]N. Mao is with School of Design, University of Leeds, LS2 9JT, United Kingdom. E-mail: n.mao@leeds.ac.uk.

[4]M. Li is with School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China. E-mail: min.li@mail.xjtu.edu.cn.

Moreover, we minimise the L1 distance between the generated data and real data for less blurring:

$$\mathcal{L}_{L1}(G_s, G_h) = \mathbb{E}_{x,s}\left[\|s - G_s(x)\|_1\right] + \mathbb{E}_{x,h}\left[\|h - G_h(x)\|_1\right]. \tag{3}$$

Haptic rendering algorithm: As shown in the test phase of Fig. 1, the trained generators G_h and G_s are used to generate the height maps $h' = G_h(x')$ and spectrograms $s' = G_s(x')$ of test objects, respectively, where x' is the visual images of test objects. Then, the spectrograms are converted to the wave-format friction coefficients $f' = \mathrm{istft}(s')$ by using the inverse short-time Fourier transform algorithm. Consequently, the scaled height maps can be denoted as:

$$m' = f'_{avg} * h', \tag{4}$$

where f'_{avg} denotes the average value of friction coefficients over different locations. Finally, we map the scaled height maps m' to the friction images m by a linear transformation for haptic rendering:

$$m = \frac{m' - \min(m')}{\max(m') - \min(m')} * (p_{max} - p_{min}) + p_{min}, \tag{5}$$

where p_{max} and p_{min} are the max and min values of the haptic display, and in our case p_{max} and p_{min} are 255 and 0, respectively.

III. Experimental Setup and Results

A total of 15 kinds of fabrics are selected in our experiments, which are made of different materials and manufactured using different weaving or knitting techniques. A weakly paired dataset is collected by sampling from them, including 3,375 visual images collected by a digital camera, 3,375 tactile images collected by a GelSight sensor [2], and friction coefficients by a force/torque sensor Nano17. We randomly split the whole dataset with a ratio of 8:1:1 for training, validation and testing, respectively.

A TanvasTouch Desktop Development Kit[1] is used for haptic rendering in this work. The Tanvas haptic display, based on electrovibration mechanism, is able to provide software-defined haptics. The values of the input friction image ranges from 0 to 255, where 0 represents the friction that naturally exists on the surface of the haptic display, and 255 represents the highest amount of friction that the device is capable of producing. To evaluate the effectiveness of our proposed method, a number of baseline methods that employ different input signals are used for comparison. The different inputs are listed below:

1) Grey-scale visual images (denoted as v_{grey}) that are obtained by the weighted mean of RGB channels of colour visual images;
2) Shape from shading (denoted as v_{shape}) using visual images [3];
3) Generated friction coefficients (denoted as f_g);
4) Generated height maps (denoted as h_g);

[1]https://tanvas.co/products/tanvastouch-dev-kit

TABLE I: The participants will be asked the following questions to measure the similarity between haptic rendering and physical fabrics.

Q1	Which of the showed three fabric pieces does the haptic rendering match with?
Q2	Does the haptic rendering have the same slipperiness as the physical fabric?
Q3	Does the haptic rendering have the same texture as the physical fabric?
Q4	How much realism of haptic rendering do you feel compared with physical fabric?

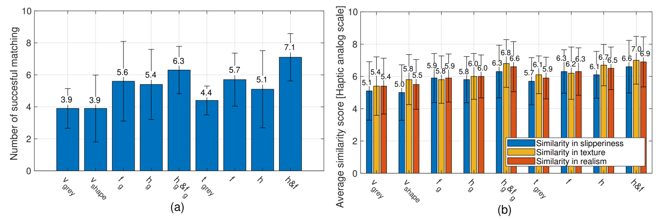

Fig. 2: **Evaluation results from user study. (a)** Average number of successful matching for each haptic rendering by different methods; **(b)** Similarities of slipperiness, texture and realism by different methods.

5) Generated height maps h_g & generated friction coefficients f_g (input of proposed method);
6) Grey-scale tactile images (denoted as t_{grey});
7) Ground truth friction coefficients (denoted as f);
8) Ground truth height maps (denoted as h);
9) Ground truth height maps h & ground truth friction coefficients f.

We recruit 10 volunteers (8 males and 2 females) and the age of participants ranges from 24 to 31. To reduce the time consumption of the testing, 7 pieces of fabrics are selected randomly in our user study. The participants are blinded to touch the physical fabrics and the haptic display and then respond to a series of designed questions as described in Table I. For Q1, the participants are asked to match the haptic rendering with the most similar physical fabrics through haptic feeling. For Q2-Q4, a rating of 0 indicates that the haptic rendering and the haptic feeling of physical fabric are unrelated, and a rating of 10 indicates that the haptic rendering is the same as the properties of physical fabrics. As shown in Fig. 2, our proposed method achieves the second-highest evaluation results, which are only slightly outperformed by the method using collected tactile signals $h\&f$. Specifically, the number of successful matching is 0.8 times less than the method using $h\&f$ and the average scores of similarities in slipperiness, textures, and realism are only 0.3, 0.2, and 0.3 less respectively, which demonstrates the effectiveness of our proposed method.

IV. Conclusions

In this paper, we propose a *Vis2Hap* haptic rendering framework to generate the height maps and friction coefficients of the object's surface from visual data, which are then applied for haptic rendering. Vis2Hap provides realistic haptic feedback of surface textures without requiring tactile data, reducing the workload of tactile data collection. Our work has a potential application in online shopping. For example, it is possible for customers to feel the haptic properties of clothes online at home without going to shopping malls.

References

[1] N. Mao, Y. Wang, and J. Qu, "Smoothness and roughness: Characteristics of fabric-to-fabric self-friction properties," in *The Proceedings of 90th Textile Institute World Conference*, The Textile Institute, 2016.

[2] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.

[3] T. Wang and X. Sun, "Electrostatic tactile rendering of image based on shape from shading," in *2014 International Conference on Audio, Language and Image Processing*, pp. 775–779, IEEE, 2014.