Proceedings Paper:

Cao, G., Jiang, J., Mao, N. orcid.org/0000-0003-1203-9773 et al. (3 more authors) (2023) Vis2Hap: Vision-based Haptic Rendering by Cross-modal Generation. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). 2023 IEEE International Conference on Robotics and Automation (ICRA), 29 May - 02 Jun 2023, London, UK. IEEE , pp. 12443-12449. ISBN 9798350323658

https://doi.org/10.1109/icra48891.2023.10160373

Connecting Physical and Virtual Touch: Haptic Rendering of Virtual Textures from Visual Pictures and Conditioned on Tactile Images

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

***Abstract*— For humans, touch is a fundamental source of information for learning and interacting with the physical world. With the development of teleoperation, haptic rendering is an essential technique for human operators to touch objects remotely and gain a comprehensive understanding of their surroundings. Previous haptic rendering methods are limited to using the recorded tactile signals from tactile sensor for haptic rendering. However, the collection of tactile data is very expensive and time-consuming due to the complex exploration. In this paper, we propose a haptic rendering method based on the generative model that generates the signals for rendering from vision and combines the characteristics of roughness and smoothness of an object's surface to provide a vivid haptic rendering. The evaluation from users demonstrates that our proposed method enables people a realistic haptic feeling and the participants can match the haptic rendering with corresponding physical objects correctly for 6.3 times over 10 trials on average. The improved haptic rendering can be used to enhance the realism and immersion of teleoperation and virtual reality.**

I. INTRODUCTION

The sense of touch is a fundamental information source for humans to learn and interact with the physical world. It allows humans to perceive exclusive physical properties, such as textures, roughness, smoothness, etc., which are difficult to be obtained through vision or any other sensations, thus helping humans understand the properties of contacting objects and explore the surrounding environment.

With the development of teleoperation, great demands exist in providing humans with haptic feedback to touch an object remotely. The act of touching things to prove their existence is a biological need to human nature, and haptic feedback enables human operators a comprehensive cognition of the remote environment from touch sensation.

Various kinds of hardware devices have been developed to provide humans haptic feedback based on different working principles, such as vibrotactile feedback [1], electrovibration feedabck [2] and thermal feedback [3]. As one kind of haptic-rendering device, electrovibration-based haptic display, which enables the frictional force changes on user's fingers, is gaining popularity in simulating surface characteristics for different objects, e.g., frictional information, roughness, and textures [4]–[6].

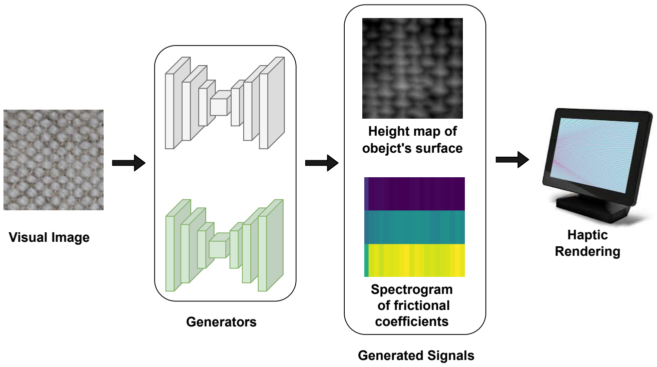
Visual Image
Generators
Height map of obejct's surface
Spectrogram of frictional coefficients
Generated Signals
Haptic Rendering

Fig. 1: **Haptic rendering framework.** A generative model is used to generate signals for rendering from visual images, then the generated signals are used for haptic rendering through a haptic display.

However, it is still difficult to design the signals that are rendered on the haptic display to provide a vivid haptic feeling. The haptic feeling of an object's surface largely depends on two aspects, i.e., ***roughness*** and ***smoothness*** [7]. Height disparities of high-frequency range changes on an object's surface are referred to as roughness. The textures and the height of surface particles differ on different types of materials. The smoothness is related to surface slipperiness which can be assessed by static or dynamic coefficients. Even with the same texture, the smoothness might vary according to their different material characteristics. Both aspects pose great challenges to designing signals in a unified way for the haptic rendering for different materials. Although it is possible to perform haptic rendering utilising the tactile signals captured from a tactile sensor that to be rendered on a haptic display directly, the process of data collection is very expensive and time-consuming. It requires a large number of physical contacts between the tactile sensor and the target objects, and the sensor is easy to be damaged.

To address the above problem in the haptic rendering, as shown in Fig. 1, we propose a cross-modal generation model to translate the visual images to height information and frictional information of object's surface. Subsequently, we integrate the height information, which indicates the roughness, and frictional information, which measures the smoothness of the surface, for haptic rendering. The evaluation results from users demonstrate that haptic rendering on the haptic display has a high similarity with the touch feeling

This work was funded in part by the EPSRC project "ViTac: Visual-Tactile Synergy for Handling Flexible Materials" (EP/T033517/1).

[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.

of physical object's surface using our proposed method.

The contributions of this paper are as follows:

1) We develop a generative model that generates height and frictional information from visual images and a haptic rendering algorithm that uses the generated signals to provide people with haptic rendering;
2) The generated height maps and frictional coefficients, demonstrating the roughness and smoothness of object's surfaces respectively, are combined together for haptic rendering, for the first time;
3) A set of experiments demonstrate our proposed method improves the realism of haptic rendering, which is promising to enhance the immersion of teleoperation and virtual reality (VR) in the future.

II. Related Works

Vision and touch are two important modalities for humans to perceive the surrounding environment in different dimensions. In this section, we will first review works on visual-tactile matching and tactile signals generation from vision, followed by a discussion of tactile signals rendering on the haptic display.

A. *Matching visual images with tactile signals*

There has been extensive research about cross-modal retrievals where the information from different modalities can be matched with each other [8]–[10]. To match the visual images with corresponding tactile information, Yuan *et al.* [11] apply both visual data and tactile data to train a CNN jointly to project the data to a shared subspace, and use the embedded vectors to determine if the visual image and tactile data are from one same object by a distance metric. Liu *et al.* [12] propose a dictionary learning model for active visual-tactile cross-modal matching where the visual images are retrieved based on the query tactile samples. Zheng *et al.* [13] propose a low-rank similarity learning method with adaptive margin to evaluate the similarity between vision and touch for retrievals.

B. *Cross-modal visual-tactile generation*

With the development of generative models, it is possible for us to translate the data from one accessible domain to another inaccessible domain. In the visual-tactile generation, Lee *et al.* [14] proposed a cross-modal data generation framework based on cGAN to generate pseudo tactile textures from visual images, using the data collected from fabrics. Cai *et al.* [15] come up with a residue-fusion module based on the generative model to do the cross-modal generation between visual images and accelerometer signals. Li *et al.* [16] adapt the generative model to perform two prediction tasks: generating tactile signals from visual video; reconstructing a visual scene that indicates which object is touched from tactile input. Moreover, Zhang *et al.* [17] propose a generative partial visual-tactile fused framework for clustering where the generated data are used to mitigate the missing data. However, these works only generate tactile data from visual data, but the generated tactile signals are not applied for haptic rendering in a further step.

C. *Tactile signals rendering on electrovibration haptic display*

By using a haptic display, an object's surface texture, roughness, temperature, and shape can be reproduced. Among various kinds of devices, electrovibration-based haptic displays are capable of providing vivid haptic feedback as they can change the friction of different locations between the screen surface and bare fingertips by changing electrostatic force.

Several studies have been conducted about providing haptic rendering from visual information, e.g., using shadings, shapes, and gradients of visual textures. İşleyen *et al.* [5] investigate how the roughness experience changes corresponding to different spatial periods and normal force according to the shape of virtual gatings on an electrovibration haptic display. Wang *et al.* [18] develop a tactile-rendering method to obtain the height information by implementing shapes from shading with Gaussian bump. Wu *et al.* [19] propose a mapping model to get frequency and amplitude based on the gradients on image textures, which is able to demonstrate the hardness and granularity on the electrovibration-based haptic display. However, these methods only provide a limited tactile feeling based on properties from visual images directly.

Another popular method is to employ the tactile sensor to record the tactile data of the contacting surface and reproduce the haptic feeling using the recorded tactile data. Jiao *et al.* [6] measure the frictional coefficients from the recorded frictional and normal forces and replay it on the haptic display by controlling the voltage to the display. Ilkhani *et al.* [20] propose a texture rendering algorithm to reproduce the acceleration signal on the haptic display, and a comparison is conducted between simulated feeling and real objects. Zhao *et al.* [21] combines the acceleration signals and friction properties to improve the realism of the haptic rendering.

To eliminate complex steps of tactile data collection, Cai *et al.* [22] use a generative model to synthesise the frictional signals from visual images, then the synthesised frictional signals are rendered on the haptic display instead of the real signal. However, the touch feeling of the object's surface relates to the roughness and slipperiness, and the above work [22] only considers frictional coefficients in a straight line but ignores the height disparities over the object's surface. In our proposed method, we use visual information and generative models to generate height map and frictional coefficients data, and combine them together for haptic rendering, for the first time.

III. Methodologies

In our framework, we aim to generate tactile signals of frictional coefficients on different locations and height maps of the object's surface from visual images, and render these generated signals on the haptic display to provide a realistic touch feeling. Since the frictional coefficients data over different locations can be treated as temporal data and represented by 2D by converting it to spectrograms which

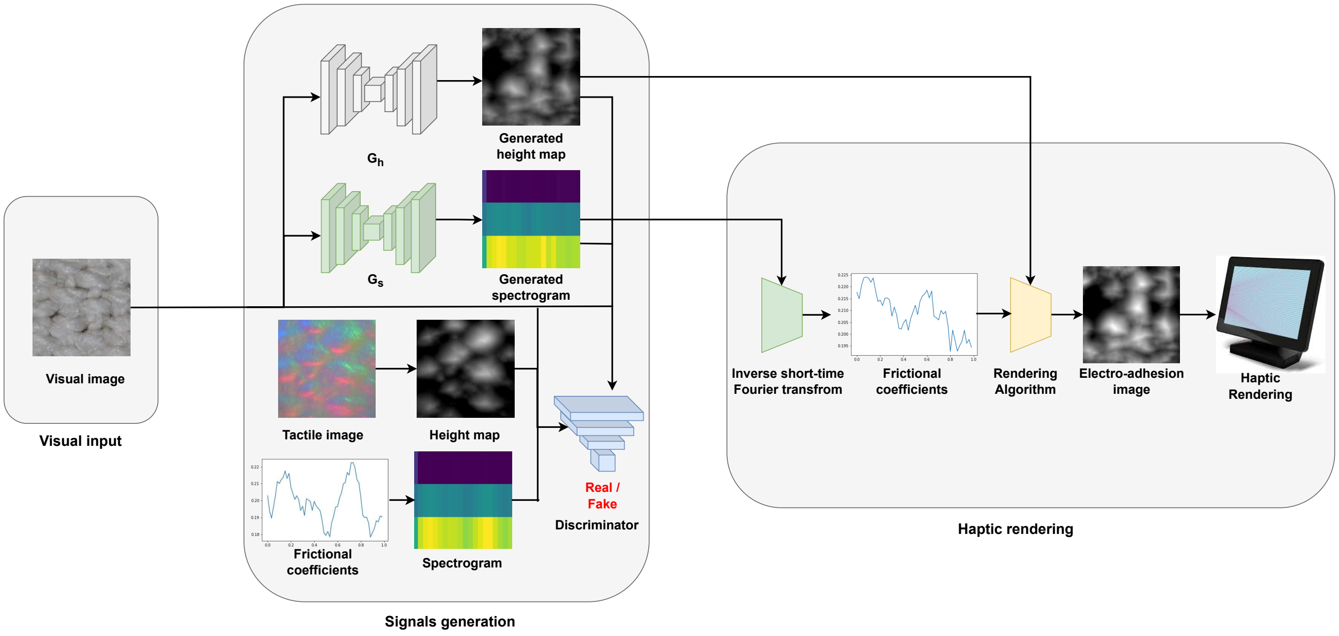

Fig. 2: **Diagram of proposed haptic rendering framework.** Two generators G_h and G_s are implemented to generate the height maps of object's surface and spectrogram of frictional coefficients respectively. The discriminator is used to identify the generated tactile signals from real tactile signals. After training, the generators are capable to generate realistic tactile signals from corresponding visual images. Then, the spectrogram are transformed to the waveform using inverse short-time Fourier transform algorithm and combined with generated height map for haptic rendering.

can illustrate the pattern of coefficients change in time-frequency domain effectively, as shown in Fig. 2, an image-to-image translation method based on cGAN [23] is proposed to generate both spectrograms of frictional-coefficients and height maps. The spectrogram and height map can demonstrate the smoothness and roughness of the object's surface respectively. By rendering the generated signals of these two key characteristics, the haptic display is able to provide a realistic haptic feeling to humans.

A. Generation of height maps and frictional coefficients from vision

As illustrated in Fig. 2, our generative model consists of two generators G_s and G_h as well as a discriminator D. The generators G_s and G_h take visual images to generate corresponding spectrograms of frictional coefficients and height maps of object's surfaces respectively. The discriminator D uses the input visual image x as auxiliary information, along with the generated results and data from real distribution, to train the model to identify whether the input to the D is from real distribution or generated.

During the training process, we optimise the generators and discriminators iteratively. Concretely, the discriminator D is trained by minimising:

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

where s and h represent the spectrogram of frictional coefficients and height maps respectively. At the same time, the generators are trained to generate indistinguishable signals to fool the discriminator by minimising:

$$\mathcal{L}_G(G_s, G_h) = -\mathbb{E}_x[\log(D(x, G_s(x), G_h(x))]. \tag{2}$$

Through the competition, the generators are capable to generate realistic spectrograms and height maps for haptic rendering. Moreover, we minimise the L1 distance between the generated data and real data for less blurring [24]:

$$\begin{aligned}\mathcal{L}_{L1}(G_s, G_m) = \mathbb{E}_{x,s}&\left[\|s - G_s(x)\|_1\right] \\ &+ \mathbb{E}_{x,m}\left[\|m - G_m(x)\|_1\right].\end{aligned} \tag{3}$$

The final objective is:

$$\begin{aligned}G_s^*, G_h^* = \arg\min_{G_s,G_h}\max_D &\,\mathbb{E}_{x,s,h}[\log D(x,s,h)] \\ &+ \mathbb{E}_x[\log(1 - D(x, G_s(x), G_h(x))] \\ &+ \mathbb{E}_{x,s}\left[\|s - G_s(x)\|_1\right] \\ &+ \mathbb{E}_{x,h}\left[\|h - G_h(x)\|_1\right].\end{aligned} \tag{4}$$

B. Haptic display

A TanvasTouch Desktop Development Kit is used for haptic rendering. The Tanvas haptic display, based on electrovibration mechanism, is able to provide software-defined haptics through the SDK. The Tanvas haptic display has a 10.1-inch screen with a resolution of 1280 × 800 pixels. The haptics are mapped 1:1 to the input electro-adhesion image. The value of pixels of the electro-adhesion 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. The device will output the required interaction as soon as the finger is over a location where an electro-adhesion image has been added.

C. Haptic rendering algorithm

In the physical world, when we use a finger to slide on the object's surface, the surface of the finger is inserted into the textures of the object due to the pressure. As a result, the locations with higher heights prevent the finger from moving and the locations with lower heights provide less friction [25]. However, even the surfaces with the same textures, the tactile feelings may vary due to the different frictional coefficients. To this end, we can use the average frictional coefficients to scale the value of the height map as the electro-adhesion image for haptic rendering.

Specifically, we use the trained generators G_h and G_s to generate the height maps $h' = G_h(x')$ and spectrograms $s' = G_s(x')$ of test objects respectively, where x' are the visual images of test objects. Then, the spectrograms are converted to the frictional coefficient waveform $f = istft(s')$ by using the inverse short-time Fourier transform algorithm [26]. Consequently, the electro-adhesion image can be denoted as:

$$m = f_{avg} * h', \tag{5}$$

where f_{avg} denotes the average value of frictional coefficients over different locations. Finally, we map the electro-adhesion images of test materials into the range of the input values of Tanvas haptic display:

$$m^k_{norm_{i,j}} = 255 * \frac{m^k_{i,j} - \min(m)}{\max(m) - \min(m)}, \tag{6}$$

where k represents the index of the test materials, and i, j denotes the location of pixels.

IV. Data Collection and Experiment Setup

In order to train the generative model, we collect a novel weakly-paired dataset, which includes visual images, height maps, and spectrograms of frictional coefficients from 15 different kinds of fabrics. Some examples are shown in Fig. 4.

A. A set of physical fabrics

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, e.g., tarlatan cotton, loomstate, zeddana silk, etc. The selected fabrics have different height distributions and frictional coefficients on their surfaces. Compared with other objects, fabrics have finer textures and irregular surfaces. As a result, the proposed method can be generalised to other objects if the haptic rendering has a significant similarity to the haptic feeling of physical fabrics.

B. Visual images of the physical fabrics

The visual images of fabrics are collected by a digital camera Canon 150D. Fabrics are placed on a flat plane with the image plane approximately parallel to them. For each piece of fabric, 5 colour images are taken under different in-plane rotations. Moreover, data augmentation is performed such as by using random rotation, flip and Gaussian noise to extend our dataset. As a result, there are 3375 colour images of fabrics in total in our data set.

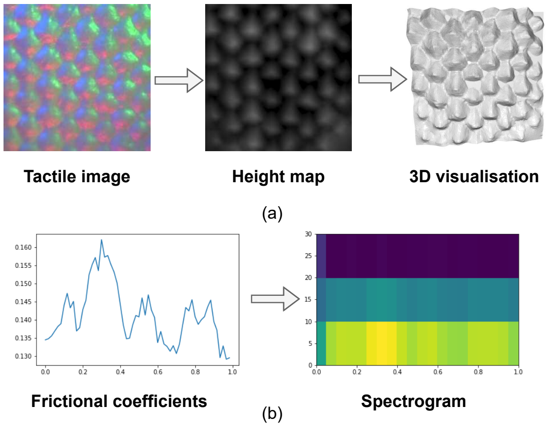

Fig. 3: **Data collection.** (a) a GelSight sensor is controlled to press against fabrics to collect tactile images. Then, height maps can be obtained from tactile images by using photometric stereo algorithm. (b) a force/torque sensor is used to slide over fabrics at a constant speed to collect frictional coefficients on a straight line. Then, the recorded frictional coefficients are transformed into spectrograms by using a short-term Fourier transform algorithm.

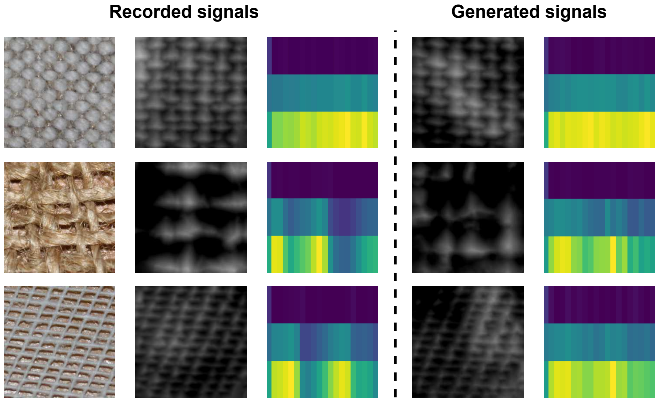

Fig. 4: **Recorded signals and generated signals.** Left three columns: visual images; height maps obtained from tactile images; spectrograms of frictional coefficients. Right two columns: generated height maps and spectrograms respectively.

C. Height maps obtained from pressing the GelSight sensor against the fabrics

A GelSight sensor [27] is used to collect the height maps of the fabric's surface textures. A GelSight sensor mainly consists of an elastomer, a webcam, a supporting plate as well as RGB LEDs. The elastomer is deformed when it contacts a fabric. The surface texture of the fabric is mapped to this deformation which is recorded by the webcam under the RGB lights.

Firstly, the GelSight sensor is mounted on the UR5 robot arm to press against the flat fabrics by a constant force (20N) to collect tactile images. Specifically, the GelSight sensor, which has a perception field around $1.5cm \times 1.1cm$,

TABLE I: The used input modalities of different haptic rendering methods. (Vis, Tac, H, and Fric represent visual images. tactile images, height maps, and frictional coefficients. H_G and $Fric_G$ represent generated height maps and generated frictional coefficients.)

Methods	Vis	Tac	H	Fric	H_G	$Fric_G$
Vis2Haptic	✓					
Tac2Haptic		✓				
H2Haptic			✓			
Fric2Haptic				✓		
H&Fric2Haptic			✓	✓		
H_G2Haptic					✓	
$Fric_G$2Haptic						✓
H_G&$Fric_G$2Haptic					✓	✓

is controlled by moving along the warp directions with a $0.2cm$ step length after each press, and this process is repeated by changing the weft location by $0.2cm$ step length as well to collect the tactile data from different locations. After collecting tactile images, following [27], we use the photometric stereo algorithm to reconstruct the height map that demonstrates the vertical displacement on the elastomer (as shown in Fig. 3 (a)). Consequently, there are 3375 height maps of the surface textures in the data set.

D. Spectrograms of frictional signals collected from sliding a force sensor over fabrics

Apart from the height maps, we collect frictional coefficients on a straight line of fabrics to measure the smoothness of each fabric. Specifically, the UR5 robot arm is equipped with a force/torque sensor Nano17, with a sampling rate of around 60Hz, to move over the fabrics. The sensor is controlled to slide along the fabric for 4 cm at a steady speed of 5 mm/s after being pressed against it with a force of roughly 15 N. By using the recorded friction and the normal pressure force, we can calculate the coefficient of friction over a straight line for each fabric. Then, we apply the short-term Fourier transform (STFT) [28] to convert the frictional coefficients into a spectrogram because the spectrogram, as a time-frequency analysis, can show the pattern of force changes effectively (as shown in Fig. 3 (b)) compared to the waveform. Finally, we have 3375 spectrograms after subsampling on recorded frictional coefficients.

E. Baselines of haptic rendering of virtual textures

Our proposed method enables humans to have haptic feelings on the Tanvas haptic display. To evaluate the effectiveness of our proposed method, a number of baseline methods that employ different input signals are used for comparison. The comparison can be divided into three groups: (1) using the visual input for haptic rendering; (2) using the generated signals from visual images for haptic rendering; (3) using the recorded signals from tactile sensor for haptic rendering. Table I details the baseline methods with different input signals.

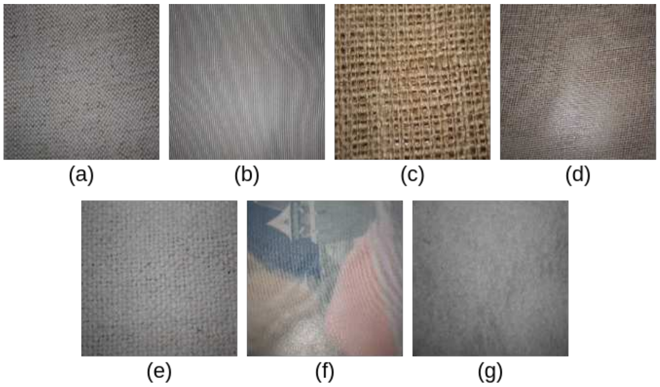

Fig. 5: Test fabrics. (a) loomstate; (b) viscose/cotton rib; (c) jute hessian (d) tarlatan cotton; (e) zeddana silk; (f) crepe fine polyester; (g) wool/cotton felt

F. Experimental setup for user study

In our experiment, we investigate if the haptic rendering based on our proposed methods keeps a high similarity with the physical fabrics. We recruit 10 volunteers (8 males and 2 females) from the University of Liverpool. The age of participants ranges from 24 to 31. None of them have experience with haptic displays. To reduce the time consumption of the testing, 7 pieces of fabrics are selected in our experiment (as shown in Fig 5). As illustrated in Fig 6, the participants will be blinded to touch the physical fabrics and haptic display respectively, and then be asked to respond to a series of questions as described in Table II. Before the experiments, the haptic rendering of random fabrics will be given on the haptic displays and let participants have a mock-up test and be familiar with the device.

For question Q1, the participants will be given one haptic rendering on the haptic display and three physical fabrics, and the physical fabric corresponding to haptic rendering is among these three physical fabrics and the other two are randomly selected. The participants will be asked to match the haptic rendering with the most similar physical fabrics

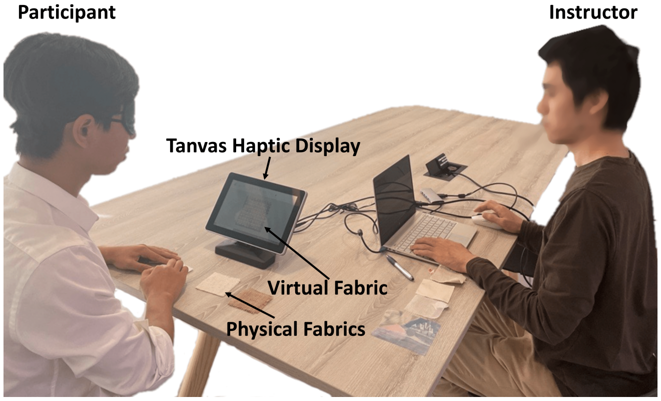

Fig. 6: **Experimental setup.** The participant is blinded with an eye mask and let to touch the physical fabrics on the table and the virtual fabrics on the haptic display. The instructor will change the physical and virtual fabrics and record the reaction from the participant.

TABLE II: The participants will be answer 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 smoothness 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?

through haptic feeling.

After testing question Q1 for all testing fabrics, questions Q2-Q4 will be asked for each fabric. Specifically, a haptic analog scale (HAS) based on a visual analog scale (VAS) rating is proposed to measure the degree of similarity between physical fabrics and haptic rendering. Each physical fabric and its corresponding haptic rendering will be shown to participants one by one. In the experiment, participants are required to memorise a nine-sectioned line segment before the test and grade similarity on this analog scale during the testing. 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 very similar to the properties of physical fabrics. To help participants understand the scale, an example is provided: the haptics of a piece of sandpaper and a piece of silk are unrelated, which receives a score of 0; the haptics from two same fabrics are totally the same, which receives a score of 10.

V. Experimental Results and Analysis

In our proposed framework, the signals including the spectrograms and height maps are generated from the visual images first. Fig. 5 demonstrate a visual comparison between the generated signals and the ground truth signals recorded by the tactile sensor. Concretely, the generated height maps and real height maps are illustrated in the second column and fourth column respectively. The third and fifth columns show the generated spectrograms and real spectrograms respectively. It can be seen that our proposed generative model is capable to generate corresponding height maps and spectrograms from visual images, and the generated results exhibit diversity and a high degree of similarity with the ground truth signals. Then, the results of haptic rendering using generated signals are compared with the baseline methods by a user study.

A. Do the vision-based methods work in haptic rendering?

Firstly, we adopt two baseline methods that use visual information as input signals for haptic rendering: (1) using grey-scale visual images that are obtained by the weighted mean of RGB channels of colour images ; and (2) using shape from shading from visual images [18]. Fig. 7 (a) (b) demonstrate the evaluation results respectively. It can be seen that haptic renderings are successfully matched with the corresponding physical fabrics for 3.9 times on average in 10 trials with the grey-scale images as input. The average similarity scores of smoothness, texture, and overall realism are 5.1, 5.4, and 5.4, respectively. The use of shape from shading, which extracts the height information from visual images, has a minor improvement in the similarities of texture and realism. However, the overall performance is low, as the similarities of realism are around 5.5 out of 10 and the participants usually cannot match the rendering correctly in most cases.

B. Do the height map and frictional coefficients generated from vision improve the realism for haptic rendering?

Our proposed method employs visual images to generate the corresponding height maps and frictional coefficients for haptic rendering. We ablate our method to demonstrate how the generated tactile signals affect the haptic perception of humans. Firstly, we only use the generated frictional coefficients for haptic rendering. Secondly, we create the haptic feedback based on generated height maps. Finally, generated height maps and frictional coefficients are combined in haptic rendering for comparison.

The Fig. 7 (c) (d) (e) illustrate the ablated results respectively. Compared with visual input, the use of generated frictional coefficients improves the haptic feeling significantly. The participants are able to match 5.6 haptic rendering correctly on average over 10 trials, 1.7 times higher than the results of visual input. Moreover, compared to the results of shape of shading, the average similarities of smoothness and realism increase by 0.9 and 0.4 respectively.

The generated height maps, which contain the 2D texture geometry, improve the average similarities in textures and overall realism by 0.2 and 0.1 respectively, compared to the results using frictional coefficients. In a further step, the combination of the generated frictional coefficients and height maps achieve the highest scores in all evaluation metrics, compared with the ablated results. It means that the combination of the generated height maps, which represent the degree of roughness, and the generated frictional coefficients, which measure the smoothness of an object's surface, is able to improve the realism of haptic rendering.

C. What is the difference between using recorded tactile signals and generated tactile signals in haptic rendering?

We further compare our proposed method with the methods which apply the recorded signals as input for haptic rendering without the generation process. The experiments here are four-fold: (1) using grey-scale tactile images; (2) using height maps obtained from tactile images; (3) using recorded frictional coefficients data; (4) combining both height maps and frictional coefficients data for haptic rendering.

As shown in Fig. 7 (f) (g) (h) (i), it is observed that using the grey-scale tactile images achieves the lowest performance as the mapping from colour tactile images to grey-scale images does not provide valid height or friction information for haptic rendering. The other results follow a similar trend to the results of generated tactile signals. The combination of height maps from tactile images and recorded frictional coefficients produces the best results among all experiments.

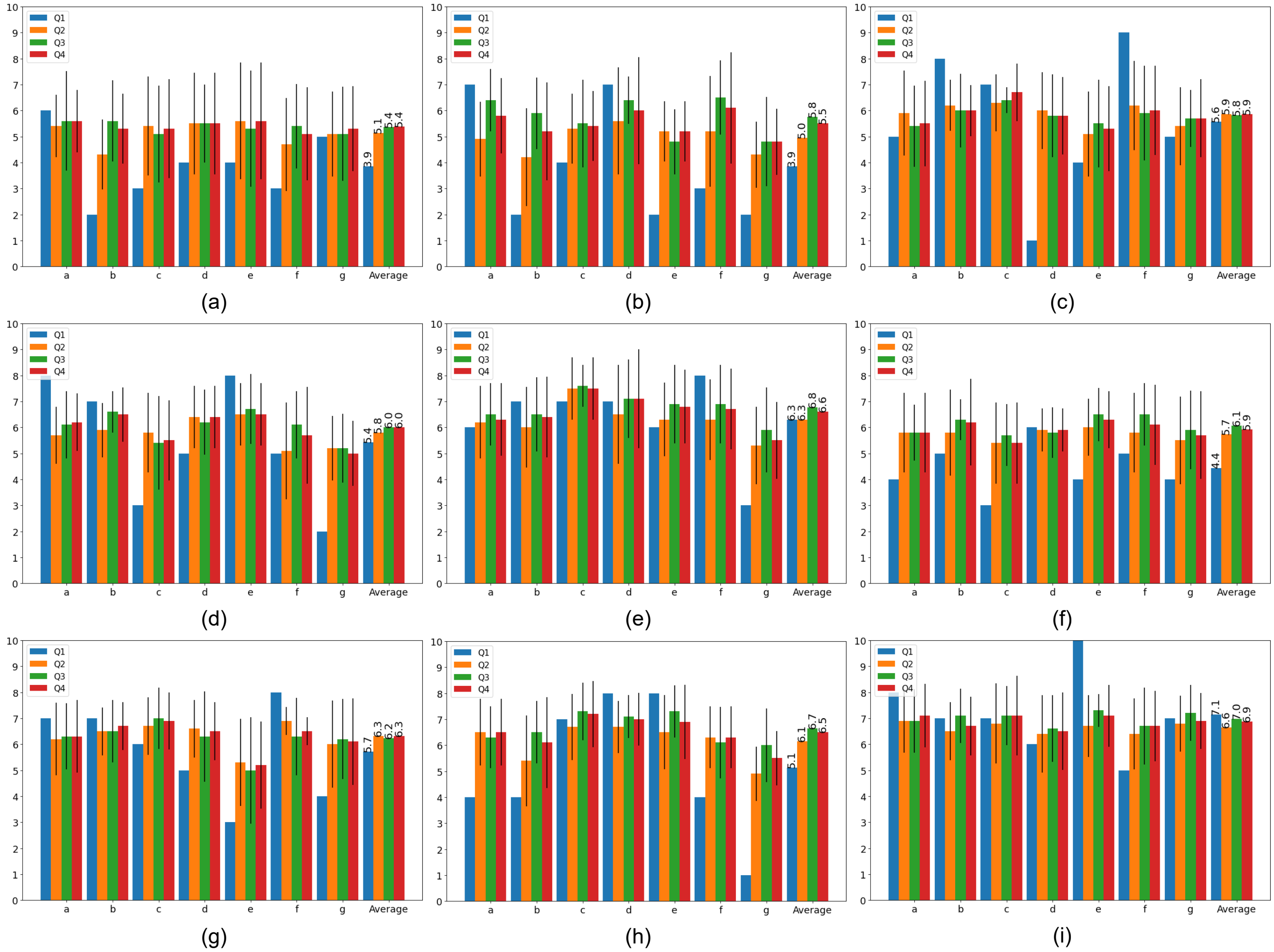

Fig. 7: Results with different input signals (a) grey-scale visual images; (b) shape from shading; (c) generated frictional coefficients; (d) generated height maps; (e) generated height maps and coefficients; (f) grey-scale tactile images; (g) frictional coefficients; (h) height maps from tactile images (i) height maps and frictional coefficients

The participants are capable to match the haptic rendering with physical fabrics for 7.1 times over 10 trials, and the score of overall realism comes to 6.9 out of 10. It is worth noting that the results of our proposed method maintain at the same level compared to the results using both height maps and recorded frictional coefficients (as shown in Fig. 7 (e) (i)). Specifically, the average scores of similarities in smoothness, textures, and realism are only 0.3, 0.2, and 0.3 less respectively, which demonstrates the effectiveness of our proposed method.

D. Comparison of electro-adhesion images with different input

Fig. 8 illustrates how different input signals result in different electro-adhesion images that to be rendered on the haptic display. It is observed from Fig. 8 (a) (c) that the simple mapping from colour image to a grey-scale image does not preserve the height or friction information of the object's surface, which can lead to a blurred haptic rendering. Fig. 8 (b) shows that the height information is reconstructed from vision by using the shape from shading method. However, there are many noisy bright pixels, which indicate a high friction value, that can adversely affect the human perception of haptic rendering. Fig. 8 (d) (g) demonstrate the electro-adhesion images using the frictional coefficients data. The frictional coefficients, however, only represent friction changes along straight lines, so humans can only feel friction changes in one dimension and cannot perceive the 2D texture clearly through the haptic rendering. Concretely, pixels in the electro-adhesion image change in value horizontally but stay the same vertically.

By combining the frictional coefficients and height maps, it can be seen that the intensity of the electro-adhesion images increase (as shown in Fig. 8 (f) (i)) compared to Fig. 8 (e) (g). The increase in intensity is corresponding to the property of object's smoothness, which can enhance the realism of haptic rendering. In addition, the electro-adhesion images in the second and third rows that represent generated and recorded tactile signals, respectively, show a high degree of similarity, which illustrates the effectiveness of our haptic

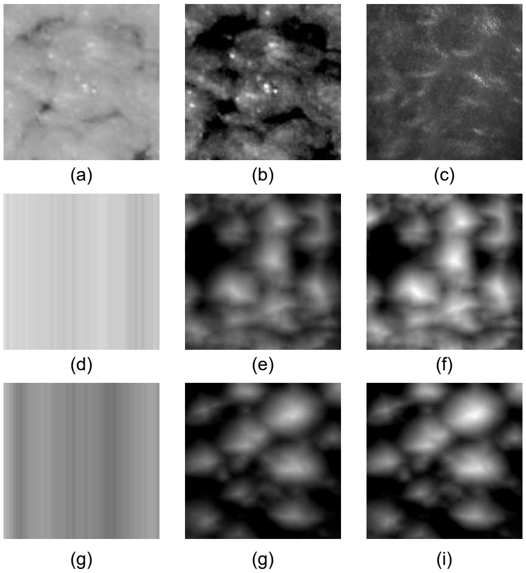

Fig. 8: The electro-adhesion images for rendering with different input signals. (a) grey-scale visual image; (b) shape from shading; (c) grey-scale tactile images;(d) generated frictional coefficients; (e) generated height maps; (f) generated height maps and coefficients; (g) grey-scale tactile images; (h) frictional coefficients; (i) height maps from tactile images (g) height maps and frictional coefficients

rendering method with the generative model.

VI. Conclusions

In this paper, we propose a haptic rendering framework that uses a generative model to generate the height maps and frictional coefficients of the object's surface, which are then combined together for haptic rendering. The realism of the haptic rendering produced by our proposed method is comparable to approaches that render haptics using recorded tactile signals by tactile sensors. The participants are able to identify the haptic rendering and its corresponding physical fabrics for 6.3 times over 10 trials on average. The proposed haptic rendering method can be used in teleoperation and VR to provide a vivid haptic feedback to humans in the future.

References

[1] H. Liu, D. Guo, X. Zhang, W. Zhu, B. Fang, and F. Sun, "Toward image-to-tactile cross-modal perception for visually impaired people," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 2, pp. 521–529, 2020.

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

[3] M. Benali-Khoudjal, M. Hafez, J.-M. Alexandre, J. Benachour, and A. Kheddar, "Thermal feedback model for virtual reality," in *MHS2003. Proceedings of 2003 International Symposium on Micromechatronics and Human Science (IEEE Cat. No. 03TH8717)*, pp. 153–158, IEEE, 2003.

[4] R. L. Klatzky, A. Nayak, I. Stephen, D. Dijour, and H. Z. Tan, "Detection and identification of pattern information on an electrostatic friction display," *IEEE transactions on haptics*, vol. 12, no. 4, pp. 665–670, 2019.

[5] A. İşleyen, Y. Vardar, and C. Basdogan, "Tactile roughness perception of virtual gratings by electrovibration," *IEEE Transactions on Haptics*, vol. 13, no. 3, pp. 562–570, 2019.

[6] J. Jiao, Y. Zhang, D. Wang, Y. Visell, D. Cao, X. Guo, and X. Sun, "Data-driven rendering of fabric textures on electrostatic tactile displays," in *2018 IEEE Haptics Symposium (HAPTICS)*, pp. 169–174, IEEE, 2018.

[7] 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.

[8] F. Feng, R. Li, and X. Wang, "Deep correspondence restricted boltzmann machine for cross-modal retrieval," *Neurocomputing*, vol. 154, pp. 50–60, 2015.

[9] J. Kim, J. Nam, and I. Gurevych, "Learning semantics with deep belief network for cross-language information retrieval," in *Proceedings of COLING 2012: Posters*, pp. 579–588, 2012.

[10] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *International conference on machine learning*, pp. 1247–1255, PMLR, 2013.

[11] W. Yuan, S. Wang, S. Dong, and E. Adelson, "Connecting look and feel: Associating the visual and tactile properties of physical materials," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5580–5588, 2017.

[12] H. Liu, F. Wang, F. Sun, and X. Zhang, "Active visual-tactile cross-modal matching," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 11, no. 2, pp. 176–187, 2018.

[13] W. Zheng, H. Liu, B. Wang, and F. Sun, "Online weakly paired similarity learning for surface material retrieval," *Industrial Robot: the international journal of robotics research and application*, 2019.

[14] J.-T. Lee, D. Bollegala, and S. Luo, "“touching to see" and "seeing to feel": Robotic cross-modal sensory data generation for visual-tactile perception," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 4276–4282, IEEE, 2019.

[15] S. Cai, K. Zhu, Y. Ban, and T. Narumi, "Visual-tactile cross-modal data generation using residue-fusion gan with feature-matching and perceptual losses," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7525–7532, 2021.

[16] Y. Li, J.-Y. Zhu, R. Tedrake, and A. Torralba, "Connecting touch and vision via cross-modal prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10609–10618, 2019.

[17] T. Zhang, Y. Cong, G. Sun, J. Dong, Y. Liu, and Z. Ding, "Generative partial visual-tactile fused object clustering," *arXiv preprint arXiv:2012.14070*, 2020.

[18] 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.

[19] S. Wu, X. Sun, Q. Wang, and J. Chen, "Tactile modeling and rendering image-textures based on electrovibration," *The Visual Computer*, vol. 33, no. 5, pp. 637–646, 2017.

[20] G. Ilkhani, M. Aziziaghdam, and E. Samur, "Data-driven texture rendering on an electrostatic tactile display," *International Journal of Human–Computer Interaction*, vol. 33, no. 9, pp. 756–770, 2017.

[21] L. Zhao, Y. Liu, Z. Ma, and Y. Wang, "Design and evaluation of a texture rendering method for electrostatic tactile display," in *Extended abstracts of the 2019 CHI conference on human factors in computing systems*, pp. 1–6, 2019.

[22] S. Cai, L. Zhao, Y. Ban, T. Narumi, Y. Liu, and K. Zhu, "Gan-based image-to-friction generation for tactile simulation of fabric material," *Computers & Graphics*, vol. 102, pp. 460–473, 2022.

[23] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[24] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.

[25] S. Lafaye, C. Gauthier, and R. Schirrer, "The ploughing friction: analytical model with elastic recovery for a conical tip with a blunted spherical extremity," *Tribology Letters*, vol. 21, no. 2, pp. 95–99, 2006.

[26] B. Yang, "A study of inverse short-time fourier transform," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3541–3544, IEEE, 2008.

[27] 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.

[28] E. Sejdić, I. Djurović, and J. Jiang, "Time–frequency feature representation using energy concentration: An overview of recent advances," *Digital signal processing*, vol. 19, no. 1, pp. 153–183, 2009.