

Incorporating the Perception of Visual Roughness into the Design of Mid-Air Haptic Textures

David Beattie
Ultraleap Ltd.
david.beattie@ultraleap.com

William Frier
Ultraleap Ltd.
william.frier@ultraleap.com

Orestis Georgiou
Ultraleap Ltd.
orestis.georgiou@ultraleap.com

Benjamin Long
Ultraleap Ltd.
ben.long@ultraleap.com

Damien Ablart
SCHI Lab, University of Sussex
da292@sussex.ac.uk

ABSTRACT

Ultrasonic mid-air haptic feedback enables the tactile exploration of virtual objects in digital environments. However, an object's shape and texture is perceived multimodally, commencing before tactile contact is made. Visual cues, such as the spatial distribution of surface elements, play a critical first step in forming an expectation of how a texture should feel. When rendering surface texture virtually, its verisimilitude is dependent on whether these visually inferred prior expectations are experienced during tactile exploration. To that end, our work proposes a method where the visual perception of roughness is integrated into the rendering algorithm of mid-air haptic texture feedback. We develop a machine learning model trained on crowd-sourced visual roughness ratings of texture images from the Penn Haptic Texture Toolkit (*HaTT*). We establish tactile roughness ratings for different mid-air haptic stimuli and match these ratings to our model's output, creating an end-to-end automated visuo-haptic rendering algorithm. We validate our approach by conducting a user study to examine the utility of the mid-air haptic feedback. This work can be used to automatically create tactile virtual surfaces where the visual perception of texture roughness guides the design of mid-air haptic feedback.

KEYWORDS

Haptics, Mid-air Haptic Devices, Visuo-Haptic Feedback, Image Processing, Texture Perception, Ultrasound, Machine Learning

ACM Reference Format:

David Beattie, William Frier, Orestis Georgiou, Benjamin Long, and Damien Ablart. 2020. Incorporating the Perception of Visual Roughness into the Design of Mid-Air Haptic Textures. In *Proceedings of ACM Conference (SAP'20)*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Mid-air haptics is a growing field of research that enables users to interactively touch virtual objects and surfaces without the need to use any specialised controllers. This is most often achieved by using ultrasonic phased arrays that electronically focus waves to generate controllable tactile sensations onto a user's hands [Carter et al. 2013; Hoshi et al. 2010]. Coupled with holographic displays,

virtual reality headsets, or traditional desktop displays, mid-air haptic feedback is a powerful tool to create immersive spatial interactions. For instance one can feel a heart beating [Frish et al. 2019], hold the Earth [Kervegant et al. 2017], or remotely shake the hand of a distant collaborator [Makino et al. 2016]. These interactions are intrinsically bi-modal, and while “seeing is believing, feeling is the truth” [Fuller 1732]. Consequently, graphical representations and haptic feedback are equally important in spatial interactions and must be holistically considered when designing multi-modal interfaces [Kortum 2008]; if visual and tactile modalities convey discrepant information, user immersion undeniably breaks.

Displaying shape and texture information on a screen or in augmented and virtual reality (AR/VR) has been an active topic of research and development for many years [Cremers et al. 2007], more so with recent proliferation of machine learning and artificial intelligence (AI) approaches to graphics synthesis [Kulkarni et al. 2015]. Methods for the haptic and audio rendering of texture information are also well studied fields [DiFilippo and Pai 2005; Ren et al. 2010], at least for wearable [Pacchierotti et al. 2017], surface [Meyer et al. 2014], and grounded [Culbertson et al. 2014] haptic interfaces. This has not however been the case with mid-air haptic technology since many other challenges first had to be addressed (see recent survey [Rakkolainen et al. 2019]). Authentic rendering of both the visual and tactile properties of a virtual object are dependent on accurate representation of its geometry (*shape*) and its material/surface properties (*texture*). To that end, research in mid-air haptics is on-going with regards to mid-air haptic shape rendering algorithms [Hajas et al. 2020; Howard et al. 2019; Long et al. 2014; Rutten et al. 2019], and still in its infancy with regards to texture rendering with only two prior published studies [Ablart et al. 2019; Freeman et al. 2017].

In this paper we advance mid-air haptic texture rendering by presenting a novel, easy to generalise, and automated approach that renders mid-air haptic feedback with a tactile roughness equivalent to that of the associated visual perception of roughness in different image textures. The approach described in this work prevents the creation of virtual textures with discrepant stimuli between the visual and tactile modalities. To summarise, this work provides the following contributions: 1) a model that predicts visual roughness perception from image texture data, 2) a method to translate visual roughness to mid-air haptic input parameters that convey predictable and complimentary tactile roughness, 3) a user validated end-to-end algorithm capable of rendering mid-air haptic feedback from textured images.

2 BACKGROUND AND RELATED WORK

General usage of the term ‘texture’ is most often associated with how an object feels when touched. However, humans often utilise visual observation to infer different material properties that comprise the surface of a particular object, most commonly referred to as ‘visual texture’ [Djonov and Van Leeuwen 2011]. In fact, variations in visual texture heavily inform our interpretation of the world, and provide us with cues to understand the shape and orientation of a surface. Just as pattern variations in surface structure can lead to perceptually different tactile sensations, the human visual system also creates its own complimentary interpretation of these patterns [Rosenholtz 2014].

Humans require the use of adjectives to interpret the various qualities found in texture in a natural and ubiquitous way, such as: *roughness*, *directionality*, *contrast*, and *regularity* [Chamorro-Martínez et al. 2016]. In particular, Tamura et al. has shown that *coarseness*, *contrast*, and *directionality* play a fundamental role in the visual interpretation of surface textures [Tamura et al. 1978]. However, the use of vague language to describe variations in visual texture may be interpreted very differently between individuals, affecting the consistency with which such subjective interpretations may be measured. To that end, attempts have been made to associate human perceptual dimensions to quantifiable image attributes, such as *roughness* being related to the spatial size and density of texture primitives, known as *texels* [Haralick and Shapiro 1991]. Furthermore, an alternative approach has been to vary pixel grey-level intensity throughout different local patterns within an image to elicit perceptually noticeable variations in texture dimensions [Russ and Neal 2015]. Considering the dimension of roughness specifically, four visual cues are commonly utilised. These are: the proportion of image in shadow, the variability in luminance of pixels outside of shadow, the mean luminance of pixels outside of shadow, and the texture contrast [Pont and Koenderink 2005]. Feature extraction from images has been a fastidiously researched area of computer vision, with numerous methods having been exploited (e.g., statistical, geometric, model-based) [Tuceryan and Jain 1998].

One such approach are grey-level co-occurrence matrices (GLCM) which offer a simple and easy to interpret method of image texture analysis [Löfstedt et al. 2019]. GLCM’s compile both spatial and statistical information, and enables the computation of second order statistics known as Haralick features [Haralick and Shapiro 1991]. This approach has been widely adopted across various fields, such as, medical image analysis [Korchyine et al. 2014], object recognition from satellite imagery [Zhang 1999], and image classification tasks [de Almeida et al. 2010]. More specifically, GLCM’s express the joint distribution probabilities for neighbouring pixels that occur within an image across relative polar co-ordinates (d, θ) . For an image I with size $(N \times M)$ and p grey-levels, the $(i, j)^{\text{th}}$ value within the resulting GLCM will express the number of times the i^{th} and j^{th} pixel values occur in an image when evaluated at the offset (d, θ) . A non-normalised GLCM can be computed as:

$$GLCM_{\Delta x, \Delta y}(i, j) = \sum_{x=1}^N \sum_{y=1}^M \varphi(x, y) \quad (1)$$

where

$$\varphi(x, y) = \begin{cases} 1, & \text{if } I(x, y) = i \text{ and } I(x + \Delta x, y + \Delta y) = j, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where x and y are the co-ordinate positions in image I , and $I(x, y)$ indicates the pixel value at the relevant co-ordinate position. Selection of appropriate values for d can be difficult to infer, and misinterpretation of this value can lead to an incorrectly calculated matrix that does not capture the underlying structure of an image texture [Zucker and Terzopoulos 1980]. However, Zucker and Terzopoulos document a strategy to overcome this by comparing matrices created over multiple spatial relationships and calculating the associated χ^2 value of each matrix. Higher χ^2 values reflect a value of d that more accurately captures the underlying structure of an image texture. An interesting caveat of this method is its robustness to image magnification.

2.1 Tactile Texture Perception and Rendering

In reality, texture is experienced as an entirely integrated sensation, where aspects of both visual and tactile modalities influence one’s response to a texture. Previous work has identified that both these modalities operate in parallel, and display consistency in interpretations of three textural dimensions: *roughness*, *hardness*, and *slipperiness* [Ballesteros et al. 2005]. With that being said, early work by Binns [Binns 1936], showed that humans are capable of similar performance during texture classification whether using vision only or both vision and tactile modalities. With regard to singular dimensions, Lederman and Abbot demonstrated that texture roughness is perceived equivalently whether using vision, haptic or visuo-haptic modalities [Lederman and Abbott 1981].

While an object’s texture and shape can be assessed visually, there is no way to obtain a complete understanding of the intricacies of its physicality without exploring it in a tactile manner. This can be achieved through static (pressing down) or dynamic (sliding) physical interactions, which reveal tactile qualities like surface friction, roughness, and temperature, with each being discovered differently depending on the type of interaction used. These tactile interactions stimulate each mechanoreceptor differently, sending information to the central nervous system to inform us of how explore, grasp, and manipulate our environment [Phillips and Johnson 1981]. Okamoto et al. suggests there are three fundamental perceptual tactile dimensions: roughness (rough/smooth), hardness (hard/soft), warmth (cold/warm) [Okamoto et al. 2013]. Roughness can be further broken down into micro-roughness (fine), and macro-roughness (bumpiness) with previous work finding that surfaces with grating wavelengths above 1 mm (macro) are perceived in a different fashion in contrast to surfaces with wavelengths below 1 mm (micro) [Lederman and Klatzky 2009], therefore introducing a multi-scale element into roughness perception. Further to roughness, hardness, and warmth, resides an additional frictional dimension relating to both moist/dry and sticky/slippery surface qualities.

Roughness features can be classified in two levels (macro/micro) due to the different mechanoreceptors activated following either spatial or temporal stimulation during surface exploration. For coarse surfaces with many macro-scale roughness features, neurophysiology studies have shown that the spatial distribution of SA1

(Merkel) receptor cells contribute to the perception of such roughness, but the temporal information due to skin vibration during dynamic exploration of a surface does not [Weber et al. 2013]. Conversely, for fine (micro) surface textures, motion is a necessary part of the haptic perception. Specifically, FA1 (Meissner) and FA2 (Pacini) receptor cells are related to the perception of fine roughness, and require dynamic stimulation to perceive such features. For example, seminal work by Katz, surmised that surface roughness cannot be estimated without the lateral motion between the object and the skin [Katz 1925].

Rendering haptic virtual texture information has been a key focus in the haptic community and has been explored using many apparatus, such as: force-feedback devices, pin-arrays, vibrotactile actuators, and ultrasonic plates to name a few [Ikei and Shiratori 2002; Strohmeier and Hornbæk 2017; Wiertlewski et al. 2014]. Researchers have explored a multitude of different parameters to vary the perceived texture of tactile stimuli. Among these parameters, frequency and waveform have shown greater influence on the perception of tactile texture [MacLean and Enriquez 2003]. However, less work has explored the application of ultrasonic mid-air haptic feedback devices to appropriately render texture.

Ultrasound mid-air haptic devices are realised through phased arrays comprising of hundreds of 40kHz ultrasonic transducers capable of creating many focused points of high sound pressure level (see recent survey [Rakkolainen et al. 2019]). When in contact with the skin, these focal points induce a localised acoustic radiation force, typically in the order of several millinewtons, causing an indentation in the skin of the order of a few micrometers. The diameter of these focal points (~8.6mm) is determined by the wavelength of the ultrasound carrier, thus defining the smallest possible haptic ‘pixel’ that can be rendered by a localised mid-air tactile stimulus. Non-localised tactile stimuli can produce complex skin surface vibrations of higher spatial resolutions [Reardon et al. 2019] thus enabling the tactile perception of features that are finer than the carrier wavelength. [Chilles et al. 2019].

To induce a perceptible tactile effect however, an ultrasonic focal point must be modulated in space or in time, or both, at a frequency within the frequency range relevant to touch (5 - 500 Hz). Further, in order to convey shape information, several focal points can be positioned along the shape perimeter and their amplitude modulated at a frequency of 200 Hz [Long et al. 2014], or a single focal point can be slowly [Rocchesso et al. 2019] or rapidly moved round the shape’s perimeter with or without short pauses [Hajas et al. 2020; Martinez et al. 2019] at some angular speed, or draw frequency (the rate at which the pattern is repeated).

Despite the numerous works on ultrasonic mid-air haptic modulation techniques and shape rendering algorithms, to the best of our knowledge, to date there have only been two attempts at understanding tactile textures in mid-air [Ablart et al. 2019; Freeman et al. 2017]. The latter used an algorithm that maps 3D surface tessellation heights to ultrasound intensity [Freeman et al. 2017]. Higher surface points were rendered with higher intensity, and lower points with a lower intensity. As a user passed their hand through a 3D surface made from tessellated pyramids, the mid-air haptic device would attempt to render a point cloud of high pressure focal points with the appropriate intensity levels therefore inducing a sensation of *bumpiness*. However, the mid-air haptic

device is asked to render several contact points all at once, thereby splitting its available output-power over large intersection areas. Additionally, the rapid high pressure changes in the acoustic field may cause a crackling audible noise as an unwanted side effect [Hoshi 2017]. Finally, while different amplitude modulation waveforms and tessellation heights and resolution were studied, authors neither provide guidance on how to set these parameters, nor report any user testing of the proposed algorithms.

Ablart et al. attempted to address these challenges by studying how the draw frequency and size of a simple tactile shape affects the perception of texture and the user’s emotional response [Ablart et al. 2019]. More specifically, a single focal point (thus being more power efficient) was rapidly moved around circles of different diameters (5, 10, and 20 cm) at different draw frequencies (5-100 Hz) while participants rated the projected haptic sensation in terms of intensity, roughness, regularity, roundness, and valence using a Likert scale. Their work showed a clear linear relationship between perceived roughness and mid-air haptic spatio-temporal draw frequency, where lower frequencies (~25 Hz) elicited a significantly rougher sensation than higher frequencies (~75 Hz).

The work described in this paper draws on knowledge from computer vision and image processing to assess statistical components of image texture and find their relationship with visually perceived image roughness. This relationship is then used to train a machine learning model that can predict visually perceived roughness from new images. We then apply this prediction to the application of mid-air haptic feedback. This paper uses Ablart et al.’s findings to create a matching pipeline between the perceived tactile roughness of a mid-air stimulus set, and the visual perception of roughness in different image textures [Ablart et al. 2019].

3 PERCEPTUAL VISUAL ROUGHNESS DATA

An assumption of this work is that image data can be used to produce a mid-air haptic sensation that produces a level of tactile roughness that appropriately matches the visual perception of roughness contained within a texture image. In order to explore this assumption, a visual texture database was required that had been subjectively assessed for the textural dimension of roughness [Okamoto et al. 2013]. A suitable image database had to meet a number of prerequisite criterion, which were: 1. The data set must contain surfaces textures, 2. textures must contain a single homogeneous, or near homogeneous texture, 3. images must have been taken from a constant viewpoint, 4. images must be constantly illuminated, 5. must have been acquired from real surfaces, 6. images must contain a high enough resolution from which to capture exact detail, 7. the data set must be sufficiently large (~100). Numerous image data sets were assessed, such as Brodatz [Brodatz 1966], MIT Vision Texture (VistTex) database, PerTex database [Halley 2012], Drexel database [Oxholm et al. 2012], and Penn Haptic Texture Toolkit (HaTT) [Culbertson et al. 2014]. Only the HaTT image data set appropriately met each requirement. However, 2 of the 100 images, (“*playing card square.bmp*” and “*candle square.bmp*”), were removed from the HaTT during this stage because these images violated our criteria.

3.1 Crowdsourcing Data Collection

Accurate perceptual assessments for a given data set are impinged on the collection of data from a sufficient number of observers. Therefore we incorporated a crowd-sourced approach by utilising Amazon's Mechanical Turk (AMT) [Buhrmester et al. 2011]. The benefit of this approach was that a much larger user group could be obtained for a small monetary reward. 187 participants were recruited through AMT. Participants were first given a consent page to complete along with a description of what was required of them during the task. They were then presented with each of the 98 images consecutively in a randomised order. Their task was to rate each image across the textural dimension of roughness, as per Okamoto et al.'s description [Okamoto et al. 2013]. Assessment of all 98 images was considered the entire Human Intelligence Task (HIT). Participants were given a maximum time limit of 3 hours to complete the task, as it was expected that the process may have taken an extended period of time, therefore a substantial time period was allowed, with the expectation that participants may have required breaks between roughness assessments. The mean time taken by users was 50mins. Participants were required to be AMT *Masters*, have a HIT approval rating of >95%, and have over 1000 HITs approved. These caveats helped to minimise the risk of poor quality data being collected. In addition, a unique randomised ID was given at the beginning of the study that was to be entered upon completion of the study. This step acted as a validation method to ensure participants completed the HIT in its entirety. In return for their time, participants were rewarded with \$7.40/hr. An absolute magnitude estimation procedure was applied during the subjective roughness assessment. No reference stimuli was first provided, instead participants were presented with a slider positioned below each image, with the adjectives "rougher" and "smoother" at each end point. No value range was provided, other than these adjectives. The goal was to establish individual participant ranges during their image roughness assessments.

3.2 Results

In order to ensure participants provided answers for each image, a prompt was displayed on screen before moving on to subsequent image assessments. Participant data was excluded from further analysis if any images contained missing roughness values. Individual values were assessed to ensure participant responses were approximately even with respect to other responses. From the 187 initial responses, 114 were retained. Data for each participant was re-scaled to a range of 0 - 100, so their individual range for roughness was retained, but distributed evenly for all users. Note that data for each image were not normally distributed, so median roughness values are reported and used throughout this work.

4 VISUAL ROUGHNESS PREDICTION MODEL

Having obtained a collection of perceptual data for the 98 images from HaTT data set, our ensuing task was to design and implement a prediction model that could successfully approximate a value of visually perceived roughness ranging from 0 (for very smooth) to 100 (very rough) for any two-dimensional image texture passed to it. While feeding an image's raw pixel values directly into a convolutional neural network (CNN) is a commonly adopted image

processing method, particularly for classification tasks [Schmidhuber 2015], we computed a series of additional features based on the computation of a grey-level co-occurrence matrix (GLCM). Our reasoning for this was that we wanted our network to learn associations on the underlying structure contained within the entirety of the image texture. To that end, our model takes as input several features collected through this processing step, in addition to the matrix itself and the pixel data from the image. The following sub-sections describe in detail each of our feature sets.

4.1 Feature Encoding

Image Feature Data. Texture images from the HaTT image data base are encoded as 24bpp 1024×1024 bitmaps. We resized each image to 256×256 pixels, using a *constant* scale factor. Images were converted to grey-scale and downsampled to 8 bpp with anti-aliasing applied (i.e. 256 grey levels). This step reduced file size and enabled a GLCM to be computed with size $(2^8)^2$. This information was passed into our CNN as a 2D matrix with shape 256×256 , i.e. single-level grey-scale images of height and width of 256.

GLCM Feature Data. GLCMs were computed for each image in the HaTT image data set. Firstly, an array of pixel distances ($d = 1, \dots, 20$) were defined, and matrices were produced at each distance step, across displacement vectors ($\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$) respectively. Taking Zucker and Terzopoulos' [Zucker and Terzopoulos 1980] approach to correctly build a matrix that represents the underlying structure contained within each image, we calculated χ^2 values for each matrix, and selected the matrix that produced the highest value for d . Once an appropriate value for d was established, we generated 4 matrices for each displacement vector. Transposed matrices were also created in order to represent relationships between pixels across the horizontal, vertical, and both diagonal directions. Summation of each matrix for a given value of θ and its transpose, as well as averaging between directions, allowed for the constructed GLCM to be symmetric and semi-direction invariant. Values were then normalised so that the resultant matrix contained the estimated probabilities for each pixel-co-occurrence. For our prediction model, this matrix is two-dimensional in the shape 256×256 , and passed as an input to a separate CNN with the same architecture as our image CNN.

Haralick Feature Data. From the computation of each image's GLCM, a series of second-order statistical measures, known as Haralick features [Haralick and Shapiro 1991], could be calculated. In order to ensure our feature set contained independent variables, we computed features for the separate groups *Contrast*, *Orderliness*, and *Descriptives* [Tomita and Tsuji 1990]. Homogeneity was calculated for the contrast group, and energy for the orderliness group. We also computed mean, standard deviation, and maximum correlation coefficient for the descriptive group, and included cluster shade and prominence to assess symmetry in the GLCMs. As such, a total of seven features were used as inputs to our model as a separate Multi-Layer Perceptron (MLP) network.

4.2 Model Architecture & Learning

In order to process both our GLCM and image, we constructed a network with 3 convolutional layers with ReLU activations [Nair and Hinton 2010], and He normal kernel initialisers [He et al. 2015].

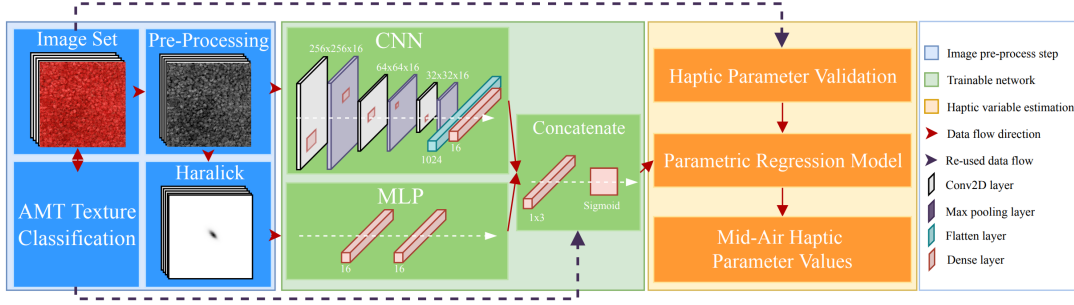


Figure 1: Prediction model pipeline overview. The image pre-processing step (blue section) includes GLCM and Haralick feature calculations. The machine learning step (green section) highlights separate networks trained on different feature sets, and ensembling (indicated by red arrows). A test set of images flows to the haptic parameter prediction section (orange) and the predicted visual roughness is mapped to a haptic parameter value via a linear regression phase (see Sec. 4.2).

The first convolutional layer applies a series of $16 \times 7 \times 7$ filters to the image and GLCM, which is followed by a 4×4 max pooling layer to reduce the dimensionality of the image and GLCM data. Filter size is then reduced to a series of $16 \times 3 \times 3$ in CNN layers 2 and 3. After CNN layer 2, another 4×4 max pooling layer is applied, with a final 2×2 max pooling layer after CNN layer 3. The subsequent output is then flattened and passed to a fully connected layer of 16-dimensions with ReLU activations, and L2 kernel regularisation set to a value of 0.005 in order to minimise overfitting. This architecture was used for both GLCM and image feature data as separate input channels. Haralick feature data was processed using an MLP with 2 fully connected layers of 16-dimensions, and L2 kernel regularisation applied to the second layer set to a value of 0.01, again to ensure overfitting was minimised. Each 16-dimension fully connected layer from the 3 models was then concatenated in order to return a single tensor passed to a fully connected layer with 3 dimensions and ReLU activations. The output layer used a sigmoid activation function to output a predicted value of visually perceived roughness in the range of 0 - 100. The model was trained using the mean absolute error (MAE) between the predicted values and roughness values from our crowd-sourced data gathering exercise. The Adam optimiser with Nesterov momentum [Dozat 2016], and a learning rate of 0.0005 was implemented, and a batch size of 1 applied. Our model was built using Tensorflow [Abadi et al. 2016] and the Keras API in Python and ran on a Nvidia GTX1070 GPU. The model architecture is displayed in Figure 1.

4.3 Model Performance

Using the scikit-learn API, and the native `train_test_split` function, our image data set (98 images) [Culbertson et al. 2014] was split into separate train (80 images), validation (9 images), and test (9 images) sets. Splitting our data set in this way acted as cross-validation to minimise over-fitting whilst ensuring our validation and test sets contained an appropriate proportion of texture classes that ranged from low to high visual roughness. The split test set was retained and used further throughout the visuo-haptic matching validation study. Our model was then trained over 150 epochs. We examined how accurately our model could predict the median visually perceived roughness values obtained during our crowd-sourcing task. The model produced a mean absolute error (MAE) of

6.46 on our training set, and 9.73 on our validation set. Our model achieved an MAE of 4.246, mean squared error (MSE) of 35.6, and a mean absolute percentage error (MAPE) of 10.45% on the 9 images contained in our test set data. Figure 2 displays plots of the AMT subjective roughness values and the model's prediction for our test images. As can be seen, the test set contains textured images with median roughness ratings that span from 14.44 (quite smooth) to 74.71 (quite rough). Similarly, the CNN predicted values closely follow these and span from 14.66 to 79.77. 92% accuracy ($R^2 = 0.92$) was observed in the model's predictions on the test set data when examined via a linear regression. Better predictions are observed at low and high values of the roughness scale.

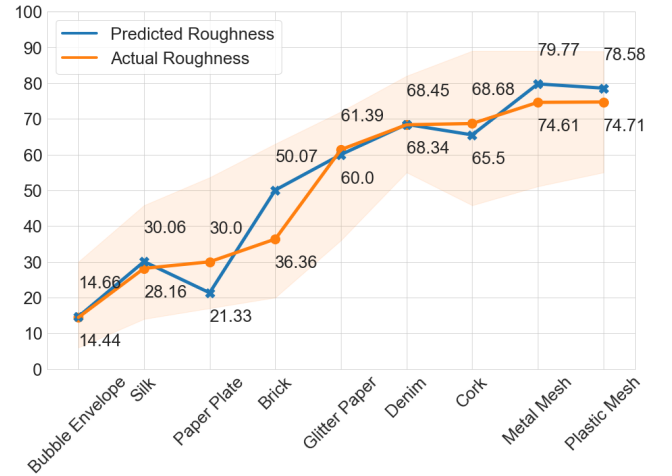


Figure 2: Comparison of visually perceived roughness between model predictions and roughness values from AMT crowd-sourced data. Orange highlighted region displays IQR values for visually perceived roughness ratings.

We conducted non-parametric analysis against our AMT visual roughness values for our test set data. As assessed by visual inspection of boxplots and Shapiro-Wilk tests, crowd-sourced values violated the assumption of normality ($p < 0.05$ for all). A Spearman's rank-order correlation test was run in order to examine the

Table 1: One-sample Wilcoxon test results on comparisons between visual roughness (Vis. R.) ratings, and model predicted roughness (Pred. R.).

Tex Name	Vis. R.	Pred. R.	Wilcoxon
<i>Bubble Envelope</i>	$M=14.44, IQR=[6-30]$	14.66	0.153
<i>Silk</i>	$M=28.16, IQR=[14-45.8]$	30.06	0.912
<i>Paper Plate</i>	$M=30.00, IQR=[17-53.6]$	21.32	<0.001 *
<i>Brick</i>	$M=36.36, IQR=[20-63]$	50.07	<0.001 *
<i>Glitter Paper</i>	$M=61.39, IQR=[36-72]$	60.00	0.311
<i>Denim</i>	$M=68.45, IQR=[55-82]$	68.34	0.944
<i>Cork</i>	$M=68.68, IQR=[45.8-89]$	65.49	0.431
<i>Metal Mesh</i>	$M=74.61, IQR=[51.1-89]$	79.76	<0.001 *
<i>Plastic Mesh</i>	$M=74.71, IQR=[55-88.9]$	78.57	0.009 *

relationship between our model's predictions and the visual roughness values obtained during the crowd-source study. A statistically significant, very strong positive correlation between model predictions and AMT visual roughness was found $r_s(9) = 0.929, p < 0.001$. A one-sample Wilcoxon signed rank analysis was then conducted on each individual image from our test data set to examine model accuracy in contrast to the entire distribution of values obtained during the crowd-sourcing study. Table 1 displays the output from this analysis. Median values are reported for subjective visual roughness, as well as the interquartile range (IQR), including Q1 and Q3. Predicted values are similar to the distributions of crowd-sourced subjective roughness values for 5 of our 9 test set images, when assessed using Wilcoxon analysis. Our model predicted a significantly higher visual roughness rating for the texture image *Brick* than the median roughness value obtained during our crowd-sourcing study ($diff = 13.71, p < 0.001$). This was also the case for texture images *Metal Mesh* ($diff = 5.15, p < 0.001$), and *Plastic Mesh* ($diff = 3.86, p = 0.009$). The model's prediction value was significantly lower than the median roughness value for *Paper Plate* ($diff = -8.68, p < 0.001$). We speculate that our model did not achieve an even higher accuracy because human visual perception can be inherently inconsistent, as people have their own subjective experience in evaluating roughness. This inconsistency is reflected in some of the large inter-quartile ranges found in our crowd-source data set.

5 VISUO-HAPTIC RENDERING

In the previous section we presented a visual roughness prediction model that takes as input an image texture and returns a predicted visually perceived roughness rating. Having validated the accuracy of this model, our goal was to combine this predicted visual roughness rating with a tactile roughness value for a mid-air haptic stimulus [Ablart et al. 2019]. Ablart et al.'s study assessed draw frequencies from 5 to 100 Hz, and found a linear relationship between mid-air haptic sensation draw frequencies in the range of 25 to 75 Hz, and perceived tactile roughness. Therefore, we only consider this range to avoid degenerate solutions. In this range, the perceived roughness decays linearly with frequency according to $y = -0.0934x + 9.54$ ($R^2 = 0.91$), where y is the roughness on a Likert scale of 1 - 9, and x is the mid-air haptic draw frequency in Hz. We re-scale the 1-9 Likert data ratings from [Ablart et al. 2019]

using a min-max normalization across a range of 0 - 100 and identify them with the predicted visual roughness ratings of our model. The approximate linear relationships of both visual roughness (see Section 4.3), and haptic roughness allowed for straight-forward conversion between each scale without information loss. The resulting visuo-haptic training pipeline is schematically shown in Figure 1.

5.1 Discussion

We note that the proposed pipeline is coupled to the ratings provided by [Ablart et al. 2019], and therefore could be limited by their corresponding test setup, haptic device, and circle stimulus. Appreciating this limitation we remark that our matched rendering pipeline approach can be updated and enhanced with different mid-air haptic devices and non-circle stimuli as required, however we consider such future work as being useful but incremental in nature. Similarly, the proposed pipeline could be extended to other haptic devices (e.g. hand-held devices, wearable haptics or vibration plates). In such use cases, one could adapt our prediction model to the targeted haptic technology. Further, we propose that the model architecture described in the previous section may be repeated for other perceptual texture dimensions and matched to the equivalent dimension for a mid-air haptic stimulus. Testing the effectiveness of such algorithms is pertinent towards the future implementation of haptics that complement powerful graphics and convey multiple texture dimensions beyond that of roughness alone. Finally, we appreciate that the visuo-haptic roughness mapping we implemented, and subsequently test, may not be optimal. For example, we have applied a min-max normalization to both the subjective visual and the tactile roughness ratings, however it is possible that different non-linear re-scalings should be used prior to the matching identification of micro-scale roughness ratings, especially since our data is not normally distributed.

6 VISUO-HAPTIC RENDERING VALIDATION

6.1 Method

Participants were asked to adjust the perceived roughness of a mid-air haptic stimulus so that it would match the texture graphic displayed on a screen. We recruited 21 participants (10 female, 11 male, mean age: 33 ± 7.8). Each participant sat at a desk, facing a 23 inch full-HD computer screen displaying an image texture from our sub-set HaTT test data. An Ultrahaptics UHEV1 haptics device (same model used in [Ablart et al. 2019]) was placed on the desk and programmed to output a mid-air haptic circle with a circumference of 20 cm. The participant's preferred hand was tracked using a Leap Motion controller, so that mid-air haptic stimuli were always centered in the centre of their middle finger, targeting their finger pads. To ensure each mid-air haptic texture was explored in a dynamic way, sensation intensity was increased only when the user actively explored the corresponding visual texture in relation to the hand displayed on screen [Strohmeier and Hornbæk 2017]. Sensations were felt only once a user's hand velocity was above a threshold value (0.05m/s). Participants wore over-ear headphones that generated pink noise to avoid any influence from auditory cues. Participants could adjust the properties of the displayed mid-air haptic stimulus via a slider displayed below the image onscreen.

This slider controlled the draw frequency of the mid-air haptic pattern, yet the participants did not know this. Instead, the cursor was labelled as “rough/smooth”. Participants were instructed to change the haptic roughness using this cursor until, in their opinion, it best matched that of the displayed graphic texture image. The sequence of the graphics displayed to each participant were randomised and repeated 3 times, and all 21 participants were presented with the nine visual stimuli test set used in our previous study (see Figure 2). The study was approved by our internal ethics committee.

6.2 Results

We examined how accurately our model could predict the user-tuned draw frequencies that matched the displayed graphic. A linear regression was applied and fitted to the user-observed values versus the model predicted values of the frequency. An R^2 value of 0.76 was calculated and a mean absolute error of 5.6 was observed.

Note that the collected data from our validation study were not normally distributed when assessed using their corresponding box plots and Shapiro-Wilk tests for normality ($p < 0.05$ for all). Firstly, a Spearman’s rank-order correlation test was run so that the relationship between our model’s predictions on haptic roughness could be evaluated against data captured during our visuo-haptic matching study. Similar to comparisons made between crowd-sourced visual roughness ratings and model predictions, a statistically significant, very strongly positive correlation was found between our model’s prediction and participant’s median haptic roughness assessments, $r_s(9) = 0.933, p < 0.001$.

It was critical to measure whether the estimation of mid-air haptic feedback varied in contrast to the visual only assessments of roughness captured during our crowd-source exercise. This information would provide an insight towards the feasibility of using perceptual visual roughness estimations as means to design mid-air ultrasonic haptic feedback. Mann-Whitney U tests were run on comparisons between crowd-sourced visual roughness data and visuo-haptic roughness data. Values were scaled differently between data sets, therefore in order to draw any comparisons, visuo-haptic matching roughness values were transformed to fit the 0 – 100 range of the visual roughness scale. Table 2 shows median, and inter-quartile values for both haptic roughness data and visual subjective roughness data from our crowd-sourced study. Assessments of Mann-Whitney results demonstrate that 3 of the 9 images were similar during visual roughness assessment and visuo-haptic matching. Of the remaining 6 images, 4 of these (*Cork*: $p = 0.008$, *Denim*: $p < 0.001$, *Paper Plate*: $p < 0.001$, and *Silk*: $p = 0.009$), visual assessments produced significantly higher perceived roughness values than during the visuo-haptic matching task. In contrast, (*Metal Mesh*: $p < 0.001$, *Plastic Mesh*: $p = 0.01$), produced significantly lower values of perceived roughness than during the visuo-haptic matching task. Finally, a comparison test was performed between data obtained during the visuo-haptic matching task and crowd-sourced visual roughness data using one sample Wilcoxon tests on each of the 9 test set images. Table 3 displays the median, inter-quartile ranges, and Wilcoxon statistical values for each image. Comparisons showed that for 4 images, (*Brick*: $p < 0.001$, *Cork*: $p = 0.038$, *Denim*: $p < 0.001$, and *Silk*: $p < 0.001$), predicted haptic roughness from our model was significantly lower than haptic

roughness during the visuo-haptic matching task. Moreover, predicted haptic roughness for 1 image, (*Metal Mesh*: $p = 0.005$), was significantly higher than haptic roughness during the visuo-haptic matching task. All other comparisons, (*Bubble Envelope*, *Glitter Paper*, *Paper Plate*, *Plastic Mesh*) were similar, thus demonstrating the potential applicability of our visuo-haptic rendering algorithm.

7 DISCUSSION

7.1 Results

As reported in Section 3 and in Table 1, a major part of our rendering algorithm is the visual roughness prediction model that we have tested and reported an accuracy of 92% and a mean absolute error (MAE) of just 4.25. These promising values show that our model can successfully predict the visual perception of roughness in image textures, but leaves some room for further refinement. For instance, a shorter AMT study conducted with more participants could have mitigating possible effects from fatigue in participant responses and therefore refined crowd-sourced data would have been obtained. Another effect could come from the close resemblance between the texture contained in the HaTT database. While constraints in the images facilitate the training of our predictive model, it makes the rating task harder for the participants as the pictures are rather abstract. This can yield to low variance within participants but high variance across participants. A larger database with more instances per texture categories could help mitigate this effect. Note here, that texture is an abstract notion in the first place, therefore variance in the ratings across participants will not disappear completely.

In the validation study involving the visuo-haptic matching task, we have observed a lower prediction accuracy ($R^2 = 0.76$) but with a small mean absolute error ($MAE = 5.65$) and with the rank order still respected ($r_s = 0.933$). We note that the reduced accuracy of the whole pipeline was to be expected. Firstly, our predictive model was trained on visual ratings only, and not visuo-haptic ratings, secondly, because the uncertainties in the visual texture predictor model propagated down to the end-to-end system, and thirdly, due to the variability in the roughness ratings of the mid-air haptic stimulus data taken from [Ablart et al. 2019]. Combining our predictive model with the roughness-to-draw-frequency relationship can only be as accurate as the two components taken separately. The Mann-Whitney test presented in Table 2 further explains this drop in accuracy and suggests a non-linear mapping between the two subjective roughness ratings. For instance, haptic roughness ratings ranged from 11 to 90, while the visual ones from 14 to 74.41. We note that the visuo-haptic pipeline can be further improved by ensuring that the roughness ratings between visual and haptic are re-scaled such that they demonstrate similar statistics and support range, thus ensuring a more congruent visuo-haptic rendering.

The one-sample Wilcoxon signed rank test provides some further details of our model’s accuracy (see Table 3) when contrasting the user selected haptic draw frequencies to our model’s predictions. Observing a mean absolute error (MAE) in draw frequency of 5.65 Hz we seek to compare this to the participants vibrotactile perceptual resolution. Namely, according to a review on psychophysical studies in haptic, the human just noticeable difference (JND) for vibrotactile frequency is can range from 3 - 30% [Jones and Tan 2013]. This suggests that JND for vibrotactile feedback is greater

Table 2: Non-parametric Mann-Whitney group differences between crowd-sourced subjective roughness data (0 – 100) and validation study haptic draw frequency data (scale transformed). Median, and IQR values are reported.

Tex Name	Haptic Roughness (Transformed: 0-100)	Visual Roughness	Mann-Whitney
<i>Bubble Envelope</i>	$M=11, IQR=[2-22(20)]$	$M=14.44, IQR=[6-30(24)]$	0.061
<i>Silk</i>	$M=22, IQR=[10-32(22)]$	$M=28.16, IQR=[14-45.8(31.8)]$	0.009 *
<i>Paper Plate</i>	$M=22, IQR=[11.5-34(22.5)]$	$M=30.00, IQR=[17-53.6(36.6)]$	<0.001 *
<i>Brick</i>	$M=28, IQR=[11-59(48)]$	$M=36.36, IQR=[20-63(43)]$	0.051
<i>Glitter Paper</i>	$M=66, IQR=[42-76(34)]$	$M=61.39, IQR=[36-72(36)]$	0.167
<i>Denim</i>	$M=56, IQR=[32-68(36)]$	$M=68.45, IQR=[55-82(27)]$	<0.001 *
<i>Cork</i>	$M=62, IQR=[36-76(40)]$	$M=68.68, IQR=[45.8-89(43.2)]$	0.008 *
<i>Metal Mesh</i>	$M=90, IQR=[78-96(18)]$	$M=74.61, IQR=[51.1-89(37.9)]$	<0.001 *
<i>Plastic Mesh</i>	$M=82, IQR=[72-92(20)]$	$M=74.71, IQR=[55-88.9(33.9)]$	<0.010 *

Table 3: One sample Wilcoxon signed rank test comparisons between median frequency set by users (User F.), and predicted frequency (Pred. F.) from model. Median and IQR values reported. Frequency values in Hz.

Tex Name	User F.	Pred. F.	Wilcoxon
<i>Bubble Envelope</i>	$M=69.5, IQR=[64.0-74.0]$	67.66	0.611
<i>Silk</i>	$M=64.0, IQR=[59.0-70.0]$	59.97	0.001 *
<i>Paper Plate</i>	$M=64.0, IQR=[58-69.25]$	64.38	0.763
<i>Brick</i>	$M=61.0, IQR=[45.5-69.5]$	49.96	<0.001 *
<i>Glitter Paper</i>	$M=42.0, IQR=[37.0-54.0]$	45.00	0.744
<i>Denim</i>	$M=47.0, IQR=[41.0-59.0]$	40.82	<0.001 *
<i>Cork</i>	$M=44.0, IQR=[37.0-57.0]$	42.25	0.038 *
<i>Metal Mesh</i>	$M=30.0, IQR=[27.0-36.0]$	35.12	0.005 *
<i>Plastic Mesh</i>	$M=34.0, IQR=[29.0-39.0]$	35.71	0.110

than the MAE of our model, and that it is unlikely that participants would be able to perceive such errors. This also suggests that improvements in our draw frequency predictions would have little effect in subjective performance of our algorithm.

7.2 Further Insights and Limitations

Roughness ratings were not normally distributed from rough to smooth. Indeed, both in the visual ratings and the validation task, participants more often gave ratings that tended towards the extremes of the roughness range, but less in-between. This might be due to the existing dichotomy in our language. In the English language, there are no adjectives describing different levels of roughness. A material is either rough or smooth. Following on from this observation, one could hypothesise that roughness rating accuracy during our tactile assessment is not that important.

According to modality appropriateness it is likely that the texture judgement is dominated by visual cues as opposed to tactile cues [Klatzky and Lederman 2010]. Therefore it is acceptable to limit tactile roughness to a reduced set of levels from rough to smooth. Providing comparative assessments between two materials with similar perceptual quantities of roughness, such an approach is viable. This has great implications for modern day applications, while research in mid-air haptic texture rendering closes the gap with visual texture rendering. We also acknowledge the fact that our model training was conducted using 80 images. Future work should increase the size of this training set by augmenting the

HaTT image database with additional image textures to further increase the robustness in our described model.

Finally, we note that our approach has only focused on texture roughness but has not addressed any of the other possible dimensions of texture. We expect that adding additional dimensions of texture [Okamoto et al. 2013] could increase realism and enhance the visual experience with evermore congruent haptics. Moreover, our method described herein can be applied to alternative haptic devices (non-contact and contact). In such use cases, one would be expected to adapt the subjective haptic roughness model to the targeted haptic technology.

8 CONCLUSION

This work explored the utility of incorporating the perception of visual roughness in images to aid the design of mid-air haptic textures that recreate an appropriate and expected estimation of tactile roughness. We approached this challenge by developing a machine learning based prediction model that could estimate visually perceived roughness with a high degree of accuracy (~92%). From this, in combination with results from perceptual studies on tactile roughness and mid-air haptic feedback, we established a relationship between the visual perception of roughness and mid-air haptic feedback input parameters. By combining visual roughness estimations and tactile roughness produced by mid-air haptic feedback, we produced a method by which both image texture and haptic feedback can be presented in a harmonious manner. To validate this, we conducted a user study showing our method was accurate to ~76%. To our knowledge, this is the first attempt to unify visual texture perception with mid-air haptic feedback. Future research will explore how our method can be applied to other texture dimensions so that visual perception can further aid the design of richer mid-air haptic feedback. Our work highlights how visual texture perception should be carefully considered when rendering textured objects and surfaces in digital environments to enhance the user experience.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 801413; project H-Reality.

REFERENCES

- Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. TensorFlow: A system for Large-Scale Machine Learning. *Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation* (2016), 44. <https://dl.acm.org/citation.cfm?id=3026899> <https://dl.acm.org/loginbiblio.poligran.edu.co/citation.cfm?id=3026899>
- Damien Ablart, William Frier, Hannah Limerick, Orestis Georgiou, and Marianna Obrist. 2019. Using Ultrasonic Mid-air Haptic Patterns in Multi-Modal User Experiences. In *2019 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE)*. IEEE, 1–6.
- Soledad Ballesteros, José M. Reales, Laura Ponce De Leon, and Beatriz Garcia. 2005. The perception of ecological textures by touch: Does the perceptual space change under bimodal visual and haptic exploration? *Proceedings - 1st Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems; World Haptics Conference, WHC 2005 April* (2005), 635–638. <https://doi.org/10.1109/WHC.2005.134>
- Henry Binns. 1936. Visual and tactual ‘judgement’ as illustrated in a practical experiment. *British Journal of Psychology* 27 (1936), 404–410.
- Phil Brodatz. 1966. *Textures: a photographic album for artists and designers*, by Phil Brodatz. Dover publications.
- Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2011. Amazon’s Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science* 6, 1 (2011), 3–5.
- Tom Carter, Sue Ann Seah, Benjamin Long, Bruce Drinkwater, and Sriram Subramanian. 2013. UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 505–514.
- Jesús Chamorro-Martínez, Pedro Manuel Martínez-Jiménez, José Manuel Soto-Hidalgo, and Belén Prados-Suárez. 2016. An adaptive fuzzy approach for modeling visual texture properties. *Fuzzy Sets and Systems* 286 (2016), 86–113. <https://doi.org/10.1016/j.fss.2015.09.008>
- Jamie Chilles, William Frier, Abdenaceur Abdouni, Marcello Giordano, and Orestis Georgiou. 2019. Laser Doppler Vibrometry and FEM Simulations of Ultrasonic Mid-Air Haptics. In *2019 IEEE World Haptics Conference (WHC)*. IEEE, 259–264.
- Daniel Cremers, Mikael Rousson, and Rachid Deriche. 2007. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *International journal of computer vision* 72, 2 (2007), 195–215.
- Heather Culbertson, Juan José López Delgado, and Katherine J Kuchenbecker. 2014. One hundred data-driven haptic texture models and open-source methods for rendering on 3D objects. In *2014 IEEE Haptics Symposium (HAPTICS)*. IEEE, 319–325.
- Carlos W.D. de Almeida, Renata M.C.R. de Souza, and Ana Lúcia B Candeias. 2010. Texture classification based on co-occurrence matrix and self-organizing map. In *2010 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 2487–2491.
- Derek DiFilippo and Dinesh K. Pai. 2005. The AHI: An audio and haptic interface for contact interactions. In *ACM SIGGRAPH 2005 Courses*. ACM, 164.
- Emilia Djonov and Theo Van Leeuwen. 2011. The semiotics of texture: From tactile to visual. *Visual Communication* 10, 4 (2011), 541–564. <https://doi.org/10.1177/1470357211415786>
- Timothy Dozat. 2016. Incorporating Nesterov Momentum into Adam. *ICLR Workshop* 1 (2016), 2013–2016.
- Euan Freeman, Ross Anderson, Julie Williamson, Graham Wilson, and Stephen Brewster. 2017. Textured Surfaces for Ultrasound Haptic Displays. *Proceedings of the 19th ACM International Conference on Multimodal Interaction - ICMI '17 Demos* (2017), 7–8. <https://doi.org/10.1145/3136755.3143020>
- Sam Frish, Mykola Maksymenko, William Frier, Loïc Corenthy, and Orestis Georgiou. 2019. Mid-Air Haptic Bio-Holograms in Mixed Reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 348–352.
- Thomas Fuller. 1732. *Gnomologia: Adagies and Proverbs; Wise Sentences and Witty Sayings, Ancient and Modern, Foreign and British*. Gale ECCO, Print Editions (27 May 2010). <https://books.google.co.uk/books?id=3y8JAAAAQAAJ>
- Daniel Hajas, Dario Pittera, Antony Nasce, Orestis Georgiou, and Marianna Obrist. 2020. Mid-air haptic rendering of 2D geometric shapes with a dynamic tactile pointer. *IEEE Transactions on Haptics* 13, 1 (2020).
- Fraser Halley. 2012. *Perceptually relevant browsing environments for large texture databases*. Ph.D. Dissertation. Heriot-Watt University.
- Robert M. Haralick and Linda G. Shapiro. 1991. Computer and Robot Vision. *IEEE Robotics & Automation Magazine* 1 (1991), 28–48.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *Proceedings of the IEEE international conference on computer vision* (2015), 1026–1034.
- Takayuki Hoshi. 2017. Gradual Phase Shift to Suppress Noise from Airborne Ultrasound Tactile Display. *Transactions of the Virtual Reality Society of Japan (in Japanese)* 22, 3 (2017), 293–300.
- Takayuki Hoshi, Masafumi Takahashi, Takayuki Iwamoto, and Hiroyuki Shinoda. 2010. Noncontact tactile display based on radiation pressure of airborne ultrasound. *IEEE Transactions on Haptics* 3, 3 (2010), 155–165. <https://doi.org/10.1109/TOH.2010.4>
- Thomas Howard, Gerard Gallagher, Anatole Lécuyer, Claudio Pacchierotti, and Maud Marchal. 2019. Investigating the recognition of local shapes using mid-air ultrasound haptics. In *2019 IEEE World Haptics Conference (WHC)*. IEEE, 503–508.
- Yasushi Ikei and Masashi Shiratori. 2002. TextureExplorer: A tactile and force display for virtual textures. In *Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002*. IEEE, 327–334.
- Lynette A. Jones and Hong Z. Tan. 2013. Application of Psychophysical Techniques to Haptic Research. *IEEE Transactions on Haptics* 6, 3 (jul 2013), 268–284. <https://doi.org/10.1109/TOH.2012.74>
- David Katz. 1925. Der Aufbau der Tastwelt. *Zeitschrift für Psychologie* (1925).
- Cédric Kervegant, Félix Raymond, Delphine Graeff, and Julien Castet. 2017. Touch hologram in mid-air. In *ACM SIGGRAPH 2017 Emerging Technologies*. ACM, 23.
- Roberta L. Klatzky and Susan J. Lederman. 2010. Multisensory texture Perception. In *Multisensory Object Perception in the Primate Brain*. 221–230. <https://doi.org/10.1007/978-1-4419-5615-6>
- Redouan Korchiyine, Sidi Mohamed Farssi, Abderrahmane Sbihi, Rajaa Touahni, and Mustapha Tahiri Alaoui. 2014. A Combined Method Of Fractal And GLCM Features For MRI And CT Scan Images Classification. *CoRR abs/1409.4559* (2014). arXiv:1409.4559 <http://arxiv.org/abs/1409.4559>
- Philip Kortum. 2008. *HCI beyond the GUI: Design for haptic, speech, olfactory, and other nontraditional interfaces*. Elsevier.
- Tejas D. Kulkarni, William F. Whitney, Pushmeet Kohli, and Josh Tenenbaum. 2015. Deep convolutional inverse graphics network. In *Advances in neural information processing systems*. 2539–2547.
- Susan J. Lederman and Susan G. Abbott. 1981. Texture perception: studies of intersensory organisation using a discrepancy paradigm, and visual versus tactual psychophysics. *Journal of Experimental Psychology: Human Perception and Performance* 7, 4 (1981), 902–915.
- Susan J. Lederman and Roberta L. Klatzky. 2009. Haptic perception: A tutorial. *Attention, perception & psychophysics* 71 (2009), 1439–1459.
- Tommy Löfstedt, Patrik Brynolfsson, Thomas Asklund, Tufve Nyholm, and Anders Garpebring. 2019. Gray-level invariant Haralick texture features. *PLoS ONE* 14, 2 (2019), 1–18. <https://doi.org/10.1371/journal.pone.0212110>
- Benjamin Long, Sue Ann Seah, Tom Carter, and Sriram Subramanian. 2014. Rendering volumetric haptic shapes in mid-air using ultrasound. *ACM Transactions on Graphics* 33, 6 (2014), 1–10. <https://doi.org/10.1145/2661229.2661257>
- Karon MacLean and Mario Enriquez. 2003. Perceptual Design of Haptic Icons. *Eurohaptics 2003 Conference* (apr 2003). <https://doi.org/10.1.1.138.6172>
- Yasutoshi Makino, Yoshikazu Furuyama, Seki Inoue, and Hiroyuki Shinoda. 2016. HaptoClone (Haptic-Optical Clone) for Mutual Tele-Environment by Real-time 3D Image Transfer with Midair Force Feedback. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), 1980–1990. <https://doi.org/10.1145/2858036.2858481>
- Jonatan Martinez, Adam Harwood, Hannah Limerick, Rory Clark, and Orestis Georgiou. 2019. Mid-Air Haptic Algorithms for Rendering 3D Shapes. In *2019 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE)*. IEEE, 1–6.
- David J. Meyer, Michaël Wiertelowski, Michael A. Peshkin, and J. Edward Colgate. 2014. Dynamics of ultrasonic and electrostatic friction modulation for rendering texture on haptic surfaces. In *2014 IEEE Haptics Symposium (HAPTICS)*. IEEE, 63–67.
- Vinod Nair and Geoffrey Hinton. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the 27th international Conference on Machine Learning* (2010), 807–814. <https://doi.org/10.1.1.165.6419> arXiv:1111.6189v1
- Shogo Okamoto, Hikaru Nagano, and Yoji Yamada. 2013. Psychophysical dimensions of tactile perception of textures. *IEEE Transactions on Haptics* 6, 1 (2013), 81–93. <https://doi.org/10.1109/TOH.2012.32>
- Geoffrey Oxholm, Prabin Bariya, and Ko Nishino. 2012. The scale of geometric texture. In *European conference on computer vision*. Springer, 58–71.
- Claudio Pacchierotti, Stephen Sinclair, Massimiliano Solazzi, Antonio Frisoli, Vincent Hayward, and Domenico Prattichizzo. 2017. Wearable haptic systems for the fingertip and the hand: taxonomy, review, and perspectives. *IEEE transactions on haptics* 10, 4 (2017), 580–600.
- John R. Phillips and Kenneth O. Johnson. 1981. Tactile spatial resolution. II. Neural representation of Bars, edges, and gratings in monkey primary afferents. *Journal of Neurophysiology* 46, 6 (1981), 1192–1203. <https://doi.org/10.1152/jn.1981.46.6.1192> arXiv:https://doi.org/10.1152/jn.1981.46.6.1192 PMID: 6275041
- Sylvia C. Pont and Jan J. Koenderink. 2005. Bidirectional texture contrast function. *International Journal of Computer Vision* 62, 1–2 (2005), 17–34. <https://doi.org/10.1023/B:VISL.0000046587.42611.2c>
- Ismo Rakkolainen, Antti Sand, and Roope Raisamo. 2019. A Survey of Mid-Air Ultrasonic Tactile Feedback. In *2019 IEEE International Symposium on Multimedia (ISM)*. IEEE, 94–944.
- Gregory Reardon, Yitian Shao, Bharat Dandu, William Frier, Ben Long, Orestis Georgiou, and Yon Visell. 2019. Cutaneous Wave Propagation Shapes Tactile Motion: Evidence from Air-Coupled Ultrasound. In *2019 IEEE World Haptics Conference*

- (WHC). IEEE, 628–633.
- Zhimin Ren, Hengchin Yeh, and Ming C Lin. 2010. Synthesizing contact sounds between textured models. In *2010 IEEE Virtual Reality Conference (VR)*. IEEE, 139–146.
- Davide Rocchesso, Francesco Saverio Cannizzaro, Giovanni Capizzi, and Francesco Landolina. 2019. Accessing and selecting menu items by in-air touch. In *Proceedings of the 13th Biannual Conference of the Italian SIGCHI Chapter: Designing the next interaction*. 1–9.
- Ruth Rosenholtz. 2014. Texture Perception. *Oxford Handbook of Perceptual Organization* (2014).
- John C. Russ and F. Brent Neal. 2015. *The Image Processing Handbook* (7th ed.). CRC Press, Inc., USA.
- Isa Rutten, William Frier, Lawrence Van den Bogaert, and David Geerts. 2019. Invisible touch: How identifiable are mid-air haptic shapes?. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.
- Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview. *Neural networks* 61 (2015), 85–117.
- Paul Strohmeier and Kasper Hornbæk. 2017. Generating Haptic Textures with a Vibrotactile Actuator. *Proc. CHI '17* (2017). <https://doi.org/10.1145/3025453.3025812>
- Hideyuki Tamura, Mori Shunji, and Yamawaki Takashi. 1978. Textural Features Corresponding to Visual Perception. *IEE Transaction on systems, man, and cybernetics* SMC-8, 6 (1978), 460–473.
- Fumiaki Tomita and Saburo Tsuji. 1990. Computer Analysis of Visual Textures. *Computer Analysis of Visual Textures* 102, 0 (1990), 6221. <https://doi.org/10.1007/978-1-4613-1553-7>
- Mihran Tuceryan and Anil K. Jain. 1998. Texture Analysis. *The Handbook of Pattern Recognition and Computer Vision* (1998), 207–248. <http://fmi.uni-sofia.bg/courses/graphics/imageprocessing/papers/texture-review.pdf>
- Alison I Weber, Hannes P Saal, Justin D Lieber, Ju-Wen Cheng, Louise R Manfredi, John F Dammann, and Sliman J. Bensmaïa. 2013. Spatial and temporal codes mediate the tactile perception of natural textures. *Proceedings of the National Academy of Sciences of the United States of America* 110, 42 (2013), 17107–12. <https://doi.org/10.1073/pnas.1305509110>
- Michael Wiertlewski, Daniele Leonardi, David J Meyer, Michael A Peshkin, and J Edward Colgate. 2014. A high-fidelity surface-haptic device for texture rendering on bare finger. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 241–248.
- Yun Zhang. 1999. Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. *ISPRS journal of photogrammetry and remote sensing* 54, 1 (1999), 50–60.
- Steven W. Zucker and Demetri Terzopoulos. 1980. Finding structure in Co-occurrence matrices for texture analysis. *Computer Graphics and Image Processing* 12, 3 (1980), 286–308. [https://doi.org/10.1016/0146-664X\(80\)90016-7](https://doi.org/10.1016/0146-664X(80)90016-7)