Surface Texture Classification Based on Transformer Network

Mudassir Ibrahim Awan
Kyung Hee University
miawan@khu.ac.kr

Jeon Seokhee
Kyung Hee University
jeon@khu.ac.kr

Abstract

An acceleration signal is generated when a person interacts with the surface of an object, which carries pertinent information about the surface material. This acceleration signal is unique to each surface and can be used to recognize the surface texture of an object. In this paper we developed a new transformer-based deep learning model for surface texture classification from haptic data. This approach leverages the self-attention process to learn the complex patterns and dynamics of time-series data. To the best of our knowledge this is the first time that the transformer or its variants are used for surface texture classification using tactile information. As a proof of concept, we collected data for 9 different textures and the evaluation experiments showed that the model achieved state-of-the-art classification accuracy.

Keyword

Transformer neural network, Surface Texture Classification

1. Introduction

In past decade texture classification has attracted many researchers due its wide-ranging applications in robotics such as transmitting surface features from an exploratory remote robot. These features can be collected through a rigid tool by stroking it over a surface, which produces vibrations that can uniquely characterize the corresponding surface and reveal its rich haptic attributes [1]. On the contrary object identification and recognition have widely been performed in the area of computer vision. 2D and 3D features form images have been used to classify object and its surface. Unlike tactile based approaches, vision-based approaches require optical sensors which can be restricted by occlusion and lightning constraints [2].

Recently signal analysis and recognition are done by deep learning-based approaches. In [3] CNN is employed to perform signal recognition. Moreover, in [4] 1D-CNN and Bi-LSTM are used in vehicle network anomaly detection.

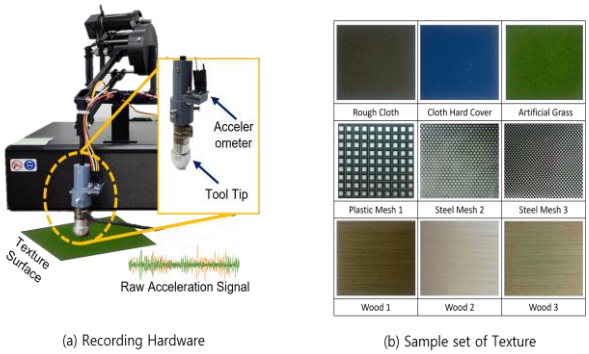

Figure 1. Hardware for sample-set recording

The works cited above served as inspiration for this paper, we develop a novel haptic texture classification approach based on Transformer architecture [5]. Unlike aligned-sequence models such as CNN, RNN, LSTM and their variants, transformer-based model does not process data in ordered sequence manner. Instead, it analyzes the complete sequence of data and use self-attentional mechanisms to understand how the data are interconnected. Therefore, Transformer-based models have the capacity to model time series data with complicated dynamics that are difficult for sequence models to process. In this work we used 9 different textures to show that our transformer-based model can be used for texture classification. The major contribution of this work are as follows:

- A transformer-based model for texture classification using tactile data.
- Results show that the model achieved state of the art classification accuracy.

The remaining paper is structured as follows. Section 2 contains detailed explanation of collected dataset and proposed transformer model. Section 3 discusses evaluation results and lastly in Section 4 we conclude our work.

2. Classification Approach

In this section, the proposed attention based deep learning network scheme will be elaborated by starting with the explanation of data recording (Section 2.1), followed by the proposed surface texture classification model (Section 2.2).

2.1 Hardware Setup and Dataset

Our Hardware setup consists of interaction tool and one sensor attached to it (see Fig. 1(a)). Accelerometer (ADXL335; Analog Device, having resonant frequency of 5.5 kHz and a noise level of 0.126 m/s2) enables us to measure the acceleration of the interaction tool tip. To record these acceleration signals we connected accelerometer to the portable data acquisition device (USB-6220; National Instrument).

In our study, 9 samples in three different groups are prepared (see Fig. 1(b)). The first group consists of fabric material, the second group consist of steel and plastic meshes while the third group is wooden samples. For each texture we collected acceleration data for 10 seconds with 3KHz sampling frequency in 2 different trials. One trial is used for training and another for evaluation. Recorded acceleration signal is also band pass filtered within the range of 25Hz and 1000Hz in order to remove gravity component, noise and the effect of purposeful human motion. Subsequently to reduce dimensionality of the recorded three axis acceleration signal we mapped these signals onto a single axis using DFT321 algorithm [6]. For the input to our deep learning model, we have also normalized the recorded signal so that each texture profile lies within the range of -1 to 1. After preprocessing we have segmented these acceleration signal into 500 samples window in time domain. These segments are used as the input to our Transformer Network.

2.2 Haptic Transformer Network

An overview of our Transformer-based Haptic texture classification network can be seen in Fig 2. Our model partially follows the original Transformer Architecture [5] as we only deployed transformer encoder block and did not use decoder block. The transformer encoder architecture is considered sufficient in time-series classification applications [7]

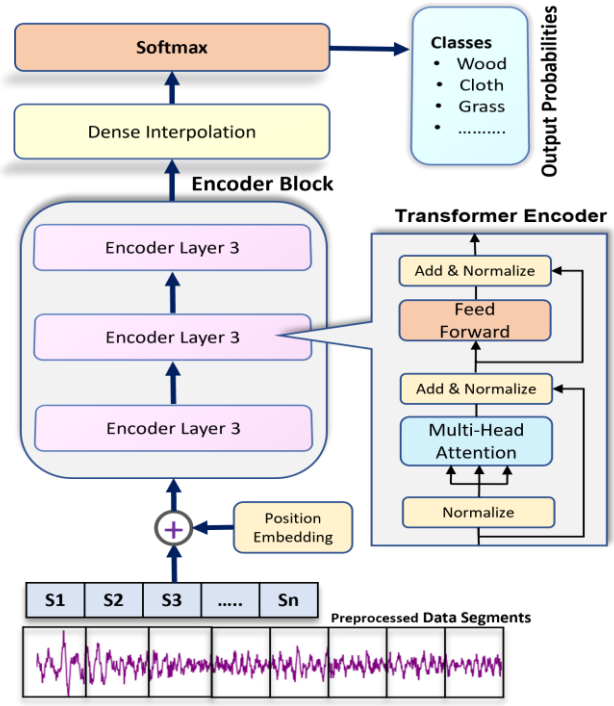

Figure 2. Transformer Network for Haptic Texture Classification

while encoder-decoder network performs well in time-series forecasting tasks [8]. Additionally, the word-embedding layer was eliminated because it was part of the original transformer scheme, which was used to do Natural Language Processing (NLP) tasks, and it was utilized to transform word sentences to numerical vectors.

The input to our network is the segments of preprocessed single-axis acceleration signals collected from each texture while the output is class label vector. These input signals first go through positional encoding. This step is essential to encode the sequential information of the input signal and it can be performed by element-wise addition of the input vector with a positional encoding vector. We used sine and cosine function encoding in our work [8]. Vector containing input acceleration signal along with position embedding is then fed into three encoder layers containing three attention heads. We chose the dimension of each encoder as (500,1). Each encoder layer is identical and contains two sub layers: self-attention sub layer and feed forward sub layer. We used 1D-CNN with kernel size (1x1) and Relu as its activation function followed by a fully connected layer as a feed-forward sub layer. Moreover, each sub-layer is also followed by a normalization layer. To make a final prediction, the encoder block output is

sent as input to the Dense Layer followed by Softmax layer of size (1 x 9). The Softmax layer produces probability distribution vector for all classes and final corresponding class can be reported by the maximum probability label.

Training Parameters: Through a number of training sessions, we selected the optimum hyperparameters for the model. Finally, Adamax was employed as model optimizer with learning rate of 0.001 and batch size of 32. Sparse categorical cross entropy as loss function was used while the number of epochs was set to 500. Additionally, we used dropout as regularization technique for each of the encoder layers to boost the classification performance and prevent the model over-fitting. 0.2 dropout-rate is used for each layer.

3. Results

In this section we present results from our transformer-based model. These findings come from the test acceleration profiles, which are different from the training acceleration data. During evaluation we followed the same preprocessing steps described in Section 2.1. Accuracy, Precision, Recall, and F1 score are the four performance evaluation metrics used in the experiment. Each performance index is calculated by the confusion matrix. Table 1 depicts the results of aforementioned evaluation metrics while Fig 3 exhibits the mean accuracy of each texture. It is noted that our proposed model achieved average accuracy of 98.87% and F1 score of 96.89 % which shows the significance of attention-based (Transformer) algorithm in texture classification applications.

Table 1. Performance evaluation metrics

Performance Metrics	Results
Accuracy	98.87%
Recall	97.1%
Precision	96.7%
F1 Score	96.89 %

4. Conclusion

In this work, we proposed a transformer-based model for texture classification using haptic data. We observed that our model exhibited the compelling performance and

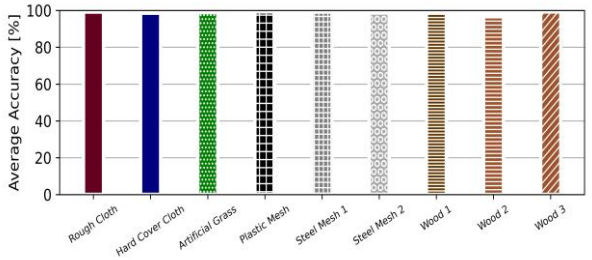

Figure 3. Accuracy score for each texture

achieved 98.87% average accuracy on collected dataset. Although real world applications have plenty of objects with various surface textures. Therefore, in the future we would like to show the effectiveness of our algorithm with more textures.

Reference

1. Culbertson, Heather, et al "Modeling and rendering realistic textures from unconstrained tool-surface interactions." IEEE transactions on haptics 7.3 (2014).
2. Hassan, Waseem, et al. "Towards universal haptic library: Library-based haptic texture assignment using image texture and perceptual space." IEEE transactions on haptics 11.2 (2017): 291-303.
3. Strese, Matti, et al. "Multimodal feature-based surface material classification." IEEE transactions on haptics 10.2 (2016).
4. Sun, Heng, et al. "Anomaly detection for In-Vehicle network using CNN-LSTM with attention mechanism." IEEE Transactions on Vehicular Technology 70.10 (2021).
5. Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
6. Kuchenbecker, Katherine J., et al. "Dimensional reduction of high-frequency accelerations for haptic rendering." (2010).
7. Natarajan, Annamalai, et al. "Convolution-free waveform transformers for multi-lead ECG classification." 2021 Computing in Cardiology (CinC). Vol. 48. IEEE, 2021.
8. Wu, Neo, et al. "Deep transformer models for time series forecasting: The influenza prevalence case." arXiv preprint arXiv:2001.08317(2020)