Investigating the Effect of Resolution on Texture Classification using Optical Tactile Sensors

Dexter R. Shepherd [1], Phil Husbands[2], Andy Philippides[2] and Chris Johnson[2]

***Abstract*—In this work we gather a large data set of 13 different textures that represent various materials a robot could come into contact with. We use a resilient clear gel within a TacTip [1] sensor to investigate whether lower resolutions can still achieve high accuracy classifying different textures. We found we could go as low as 20% of the original image size and still get above 97% accuracy on the test data.**

I. Introduction

In most environments, terrestrial locomotion takes place on varied substrates. For a locomoting animal or robot to adapt its gait to changes in substrate, it needs to quickly and reliably identify substrate surface properties [2, 3]. Tactile sensory modalities play a large role in the acquisition of this essential environmental feedback. In robotics, optical tactile sensory systems have been shown to be highly effective in this task [1], but they are computationally expensive, and their tactile interfaces can be fragile. In this paper, we show that it is possible to achieve highly accurate surface classification while making significant gains in computational efficiency by reducing the resolution of the video feed, and furthermore that these gains can be achieved with a more durable interface than has been used previously.

II. Methods

The Sensor: We constructed a TacTip (see Fig. 1) with 133 optical markers and with two material types for the TacTip flesh: RTV27905 silicone (the previous standard) and SORTA-CLEAR silicone (SCS) which has a tensile strength of 473 psi. The SCS is less widely used for optical tactile sensing, despite it being cheaper. RTV27905 silicone has lower stiffness than SCS, and therefore is potentially more sensitive. However, it is also more prone to breakage under pressure. As we intend this part of the sensor to be in contact with the ground and supporting a robot's weight during locomotion, we prefer the higher toughness of SCS so wanted to test if the reduced sensitivity compared to RTV27905 had an affect on classification accuracy.

Rig Testing Setup: Our test rig is a TacTip mounted on a 3DOF cartesian robot (figure 1). The TacTip is pressed downwards onto a textured surface. The sensor tip is then dragged across the surface, in varied directions so as to avoid biasing the data-set. The rig was controlled by a RP2040 chip running MicroPython.

Texture Dataset: The texture dataset was gathered by dragging a TacTip across a range of materials at different pressures using the rig in Fig. 1 and recording the images of the deformed pins from the initial point of touch, for approximately 10 seconds. Pressures was determined by how far a sensor is lowered onto the touching point of plate itself. Further pressure was applied by lowering 3mm, then 6mm. The more pressure, the more physical resistance against the sensor there was. We used a range of pressures which equates to a range of forces on the sensor because we want the sensor to be robust across a range of robot weights and forces. Because there are no standard textural datasets, we picked a range of textures that might reasonably cover a floor, ranging from carpets, soft materials, different fabrics and hard materials.

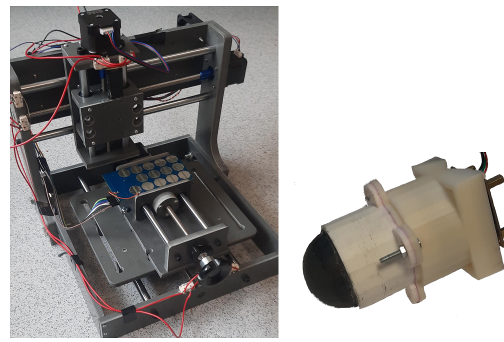
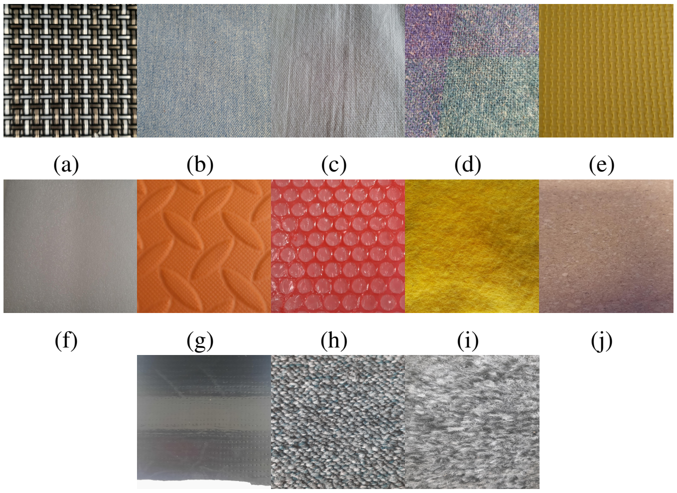


Fig. 1: Test Rig (left panel). The plate (seen as the blue PCB with 15 pads on) has a pressure surface attached held by four spacers the. The material is screwed down on top of the plate. The TacTip (right panel) is attached to the movable arm of the rig.and is angled taright downwards.

The dataset was converted to the dimensions (N,t,w,h) where N is the size of the dataset, t is the number of frames per trial and w and h are the width and height of the image respectively. We converted the images to grey-scale and applied a Sobel filter to make the optical markers stand out [4]. The initial image size is 110×120 pixels, over $t = 20$ frames. The dataset was further augmented by rotating the images 90, 180, and 270 degrees so as to not bias the sensor detection to specific directions of travel.

Model: We employed a convolutional neural network (CNN) for pattern recognition. Our arbitrarily chosen architecture was made up of two convolutional layers, a max pooling layer and two linear layers before the output layer. The dataset was concatenated into a line of T images before applying min-max scaling over the dataset. Each convolutional layer used a kernel of 3×3. The max pooling kernel was 2×2 leading into a linear layer where the size was calculated using equation $L = 64 \times h \times T \times 0.25 \times w \times 0.25$. The output layer was the number of classes, and calculating the resultant class used the



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(k) (l) (m)

Fig. 2: Our texture dataset was developed to try and align with previous datasets. Our texture dataset was developed using a range of everyday floor coverings to try and align with previous datasets which often use common lab materials. The textures were (a-m): Interlaced mat. Denim Jeans. Cotton. Wool. Foam with small groves (referred Efoam). Foam with smooth surface (referred Ffoam). Foam with large groves (referred Gfoam). Bubble wrap. Felt. Cork. Concentrated rubber. Short carpet. Long carpet.

maximum value for the class prediction. The full augmented (added rotated images) dataset was split into 80% training and 20% testing. We used the stochastic gradient descent algorithm with 100 epochs per model for training and a 0.005 learning rate. A hyperparameter search was undertaken to evaluate the importance of temporal size, and start position of the sensor in order to classify texture. We could use as little as the first four frames of the trial to achieve an accuracy of around 99% on the testing dataset.

III. Results and Discussion

To assess the importance of the resolution of the optical sensor, We experimented by downsampling the images from the database with a range of scaling values of: 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 65, 60, 70, 80, 90 and 100 %. All resizing used the area interpolation method. At 5% the image is pixelated and to a human incomprehensible. 10% is still pixelated but some detail of movement can be seen. The CNN model was trained on each data set, and the accuracy gathered, for each resolution. This experiment was repeated five times to gather an average and maximum performing network. The trends shown in figure 3 make it clear that low resolutions lead to a poorer performance and that anything below 20% is impeded in accuracy. However, resolutions above 20% can perform well. The differences in results between the two types of silicone are negligible, meaning that we can increase durability with no real cost in terms of accuracy. The minimal image size should cover a frame of 24×88 pixels (88 being

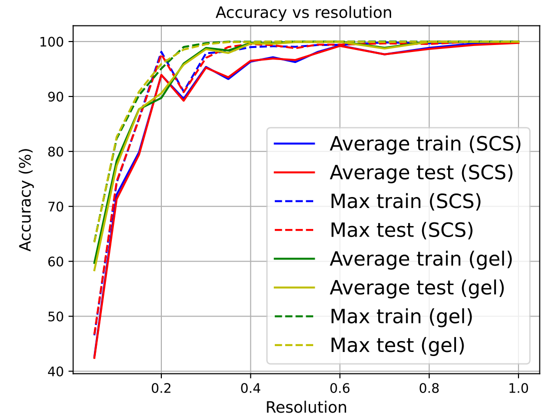


Fig. 3: Texture classification accuracy for different image resolutions using four frames. Averages shown in solid line, best performance shown as dotted. Results are shown for both types of silicone: SCS and gel; as indicated in the legend.

the concatenated four temporal images of 24×22). These dimensions represent 20% of our original cropped image.

IV. Conclusion

We have demonstrated that it is possible to distinguish between 13 surface textures with an accuracy of over 97% using a system composed of a TacTip, minimal preprocessing, and a CNN. Remarkably, we found that we could reduce the resolution of the TacTip video feed by up to 80% without compromising performance, thereby dramatically reducing the computational costs of this system. We also found that we could achieve this while simultaneously increasing the durability of the tactile interface, leading to increased physical reliability.

Acknowledgments

This work was funded by the Leverhulme Trust

References

[1] N. F. Lepora, ”Soft biomimetic optical tactile sensing with the TacTip: A review,” extitIEEE Sensors Journal, vol. 21, no. 19, pp. 21131–21143, 2021.

[2] V. Dürr and A. F. Krause, ”Tactile sensing in insects,” in extitEncyclopedia of Computational Neuroscience, D. Jaeger and R. Jung, Eds. New York: Springer, 2014.

[3] A. O. Krushynska, F. Bosia, M. Miniaci, and N. M. Pugno, ”Spider web-structured labyrinthine acoustic metamaterials for low-frequency sound control” New Journal of Physics, vol. 19, no. 10, p. 105001, 2017.

[4] M. Comi, A. Tonioni, M. Yang, J. Tremblay, V. Blukis, Y. Lin, N. F. Lepora, and L. Aitchison, ”Snap-it, tap-it, splat-it: Tactile-informed 3D Gaussian splatting for reconstructing challenging surfaces,” extitarXiv preprint arXiv:2403.20275, 2024.