

RELATIVE ADVANTAGE OF TOUCH OVER VISION IN THE EXPLORATION OF TEXTURE

Yoon Ho Bai, Choonseog Park and Yoonsuck Choe

Department of Computer Science, Texas A&M University

Email: yoonbai@gmail.com, cspark13@cs.tamu.edu, choe@tamu.edu

ABSTRACT

Texture segmentation is an effortless process in scene analysis, yet its mechanisms have not been sufficiently understood. A common assumption in most current approaches is that texture segmentation is a vision problem. However, considering that texture is basically a surface property, this assumption can at times be misleading. One interesting possibility is that texture may be more intimately related with touch than with vision. Recent neurophysiological findings showed that receptive fields for touch resemble that of vision, albeit with some subtle differences. To leverage on this, we tested how such distinct properties in tactile receptive fields can affect texture segmentation performance, as compared to that of visual receptive fields. Our main results suggest that touch has an advantage over vision in texture processing. We expect our findings to shed new light on the role of tactile perception of texture and its interaction with vision, and help develop more powerful, biologically inspired texture segmentation algorithms.

1. INTRODUCTION

Visual perception starts from segregation of scenes based on cues related to luminance, color, contours and texture of object surfaces. Moreover, the human visual system uses texture information in order to automatically—or preattentively—segregate parts of the visual scene [1]. Several theories and algorithms exist for texture discrimination based on vision [2, 3]. These models diverge from one another in algorithmic approaches to address texture imagery using spatial elements and their statistics. Even though there are differences among these approaches, they all begin from the assumption that texture segmentation is a visual task.

However, considering that texture is basically a surface property [4], this assumption can at times be misleading. An interesting possibility is that since surface properties are most immediately accessible to touch, tactile perception may be more intimately associated with texture than with vision (it is known that tactile input can affect vision [5]).

Coincidentally, the basic organization of the tactile (somatosensory) system bears some analogy to that of the visual

system [6]. In particular, recent neurophysiological findings showed that receptive fields for touch resemble that of vision, albeit with some subtle differences [7]. To leverage on this, we tested how such distinct properties in tactile receptive fields can affect texture segmentation performance, as compared to that of visual receptive fields.

Our results based on the above ideas suggest that touch has an advantage over vision in texture processing. We expect our findings to shed new light on the role of tactile perception of texture and its interaction with vision, and help develop more powerful, biologically inspired texture segmentation algorithms.

2. METHODS

The most widely used feature generators for texture segmentation is the computational model of the early visual receptive field (of V1 simple cell), the Gabor filter [8]. When generating Gabor features, typically, an input image $I(x, y)$, $(x, y) \in \Omega$ (Ω is the set of pixel locations) is convolved with a 2D Gabor function $G(x, y)$ as follows [9]:

$$G_{\lambda, \theta, \varphi}(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cdot \cos\left(2\pi\frac{x'}{\lambda} + \varphi\right)$$
$$x' = x \cdot \cos \theta + y \cdot \sin \theta, y' = -x \cdot \sin \theta + y \cdot \cos \theta$$

where λ is the wavelength ($1.5 \times$ window size), θ is the orientation preference, φ is the symmetry phase, γ is the aspect ratio, and σ is the standard deviation of the Gaussian envelope. In our experiments we set these values to be $\sigma = 0.56$, $\gamma = 1.0$, and $\varphi = 0.5\pi$. Afterwards, a bank of Gabor filters with eight equidistant preferred orientations, $\theta = k \cdot \frac{\pi}{8}$, ($k = 0, 1, \dots, 7$) was constructed. The tactile counterpart of the V1 simple cell model is the receptive field (RF) for neurons in the somatosensory area 3b [10]. To the best of our knowledge, tactile RFs have not been incorporated in any texture segmentation or computer vision related algorithms.

DiCarlo and Johnson [10] derived the tactile RF model by recording area 3b neural responses to dot patterns using reverse correlation. The main structure of the RFs consists of three Gaussian subfields: central excitatory region accompanied by an inhibitory lobe and a temporally, dynamically lagging inhibitory lobe with respect to the excitation center [10]. Each subfield can be expressed as:

This research was funded in part by NIH/NINDS (#1R01-NS54252).

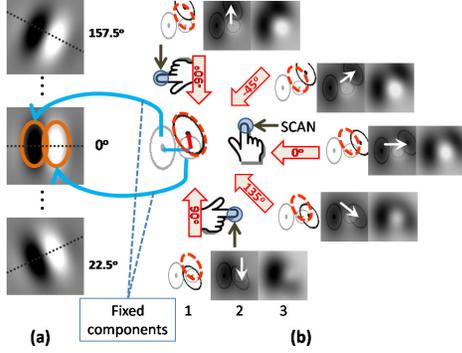


Fig. 1. RF models: (a) Visual RFs showing 157.5° , 0° , and 22.5° orientation preference. (b) Tactile RF's responses to five finger-tip scanning directions (baseline orientation preference is 0° in all cases).

$$G(x, y) = a \cdot \exp\left(-\frac{1}{2}L^T S^{-1}L\right),$$

$$L = \begin{bmatrix} x - \mu_x - \mathbf{v}_x \tau \\ y - \mu_y - \mathbf{v}_y \tau \end{bmatrix}, S = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

where (μ_x, μ_y) represents the center of the subfield, $(\mathbf{v}_x, \mathbf{v}_y)$ represents the stimulus velocity vector, and τ represents the delay of the peak excitation or inhibition with respect to skin stimulation. The center of excitation was fixed to stay at the middle of all tactile models while the complementary inhibition and lagging inhibition centers varied with respect to the excitatory center. The parameters a , σ_x , σ_y , and ρ specify the amplitude, spread, orientation, and elongation (aspect ratio) of the excitatory ($a > 0$) or inhibitory ($a < 0$) component represented by the Gaussian function.

Finally, the three Gaussian subfields are linearly summed to represent the tactile model. In Figure 1, the outline in the middle depicts the initial RF before scanning. The arrows represent scanning directions of the fingertip. From each scan, the resulting RF is illustrated through three diagrams: (1) The excitatory and fixed inhibitory lobe are outlined in gray ellipses and the lagging component is illustrated as dotted (before scanning) and black (after scanning) ellipses; (2) the lagging inhibitory lobe is displaced in the opposite direction of the scan; (3) and the linear summation of arrows listed as fixed orientation components.

This extra degree of freedom of the tactile receptive field (TRF) model (the lagging component) affects the level of occlusion of the excitatory lobe that ultimately determines orientation preference. As with the visual receptive fields (VRFs), a bank of eight TRF models with equidistant orientations was made.

Given the computational models of tactile and visual modalities, our experiments measured the performance of texture segregation through means of detecting texture-defined boundaries from natural and synthetic texture image inputs. We used 18 textures from the widely cited Brodatz texture collection [11] and 18 textures with boundary simply synthesized by appending two textures. For the experiments, we

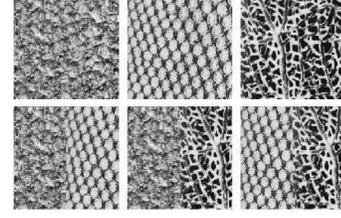


Fig. 2. A sample texture set. The top row shows non-boundary and the bottom row boundary textures.

made six different texture sets so that each set contains three non-boundary textures and three boundary textures. Figure 2 shows one texture set containing three non-boundary textures and three boundary textures. In order to extract the RF response for the given textures, each texture was preprocessed by a Laplacian of Gaussian (LoG) filter, a popular choice for edge detection, followed by a transformation of the edge into detectable discontinuities [12]. The LoG filter is defined as below

$$G_\sigma(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \left(-\frac{x^2+y^2}{2\sigma^2}\right),$$

$$LoG = \Delta G_\sigma(x, y) = \frac{\delta^2}{\delta x^2} G_\sigma(x, y) + \frac{\delta^2}{\delta y^2} G_\sigma(x, y)$$

where σ is the standard deviation (width) of the Gaussian envelope and is set to 0.5 in our experiment.

In order to reproduce equivalent stimuli from tactile sensation from a finger, we examined a certain number of consecutive window patches (frames) sliding across a predefined scanning direction inside the input image. The pixel intensity in the image played the role of surface height in texture surfaces. In this experiment, we used 12 frames with a window size of 15 (or 17) pixels. TRF and VRF filter banks constructed with 8 oriented RFs having identical size with the window patches (frames) were applied by vectorized dot product to the individual window patches, producing a vector consisting of 12 response values.

Specifically, we examined multiple scans from every input image with different scanning directions to accommodate all possible ways of encountering the texture boundaries. Figure 3 shows a group of typical response profiles extracted from random natural texture images scanned at various directions. Texture boundaries were located in the middle of the input image between two distinct textures. As shown, TRF responses show higher amplitudes and correlation along the texture boundary, compared to those of VRFs.

3. EXPERIMENTS AND RESULTS

We compared the classification rate based on the two types of RF responses (TRFs and VRFs), and also analyzed the boundary/non-boundary separability in the two representations.

So far, we obtained TRF and VRF response data from input images with texture-defined boundaries. To expose effects



Fig. 3. 3D visualization of RF responses. Visualization of 100 (left) TRF and (right) VRF responses superimposed on the input image containing a texture-defined boundary in the middle.

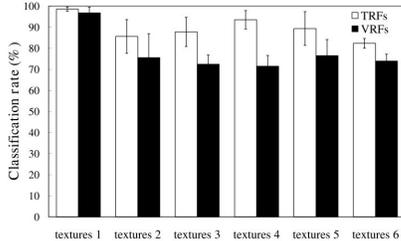


Fig. 4. Comparison of average classification rate for six different texture sets with TRFs and VRFs. In almost all cases, the classification rate with TRFs is better than the one with VRFs (errorbars indicate standard deviation).

of boundary-present responses, we conducted a double-blind test by adding a virtually identical, but controlled experiment without any texture-defined boundaries. On the condition of the same combination of scanning direction, input textures (without any boundaries) and identical parameterizations, we have collected the same amount of controlled data.

We trained a standard back-propagation network (120 input unit, 10 hidden units, 2 output units ([1 0] for boundary and [0 1] otherwise) throughout 200 epochs at learning rate $\eta = 0.5$) to discriminate texture boundary responses from non-boundary responses. On top of training with standard back-propagation, the final decision of detecting a texture boundary was based on voting [13]. We selected five neighboring output vectors from the network and applied the majority rule to finally determine whether the five outputs indicated a texture boundary or non-boundary texture input.

We used 18 random textures from the Brodatz collection and constructed two sets of input images: target-present (texture boundary) images versus target-absent (no texture boundary) images as depicted in Figure 2. Figure 4 shows the resulting classification rate of voted texture boundaries for 6 exemplary sets. The TRF performances were significantly superior to those of the VRF (t-test: $n = 1920, p < 0.03$) except for texture set 1 ($p = 0.27$).

Why are TRFs better than VRFs for texture segmentation? One possible reason is that the nonlinear structure of the TRF is more ecologically suited to the feature of the surface texture than linear structure of the VRF because most textures are

composed of more nonlinear features than linear features.

As we can see in figure 1, a three-component model has curvy lobes between excitatory and inhibitory components because the lagged inhibitory component affects the two fixed components whereas a Gabor filter has linear division between the excitatory and the inhibitory components. Hence, the bank of three-component models may easily extract more nonlinear features in the surface texture than a Gabor filter bank.

To validate this idea, we tested boundary detection with curvy textures and linear textures. Figure 5 shows two types of texture and the comparison of the classification rate. Curvy textures without boundary were synthesized with many segments of circles at different curvatures (0.333, 0.2, and 0.143) and linear textures without boundary were synthesized with lines at different orientations (horizontal, vertical, and slash). As we can see, TRFs show a higher competitive edge on curvy textures.

In order to test the representational power of TRFs as compared to VRFs, we used Fisher’s Linear Discriminant Analysis (LDA). Figure 6 shows the probability density of the LDA distribution extracted from a set of input images. In each case, data from the non-boundary and the boundary case are shown as two separate classes. The plots show that the TRF response feature distribution is more separable than those from the VRF.

Up to this point, we have applied classification result comparisons to determine the relative usefulness of local spatial features represented in TRF and VRF responses. However, this characterizes the joint performance of a feature operator and a subsequent classifier. In other words, subsequent processing of the raw responses such as voting or LDA are liable to alter raw attributes of the initial input space.

Therefore, from the discriminant analysis, we can objectively compare the features only by measuring cluster separability according to the Fisher criterion [9]. Moreover, this was expected and we subsequently validated it with ANOVA, as we found that TRF responses had p-values under 0.05 while VRF responses had higher p-values around 0.10. The results indicate that TRF based-responses have higher separability, better representing boundary-present features.

4. DISCUSSION

The primary aim of this study was to explore and compare performances of texture segregation based on Tactile Receptive Fields (TRFs) and Visual Receptive Fields (VRFs). Our current finding suggests that touch-based texture perception contains more discriminative information than vision-based local spatial features. Statistical measures and classification performances were used to evaluate this characteristic as well as providing insight on analyzing the TRF representation. Due to the extra degree of freedom and component in the RF structure, TRF’s functional implications can accommodate

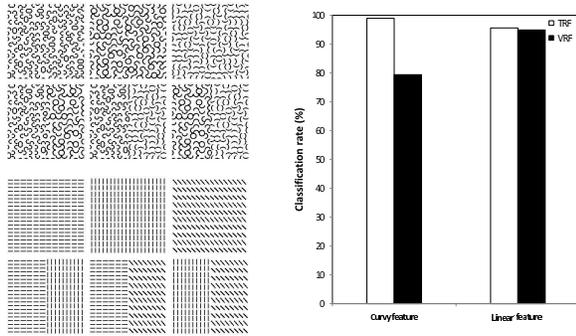


Fig. 5. Comparison of classification rate for curvy and linear textures. Left: Textures with curvy (top two rows) or linear features (bottom two rows) are shown. Right: Performance of TRF vs. VRF on curvy/linear textures are shown.

more complex spatial properties, e.g., curvature, than VRFs. In particular, we have also investigated a preliminary test to analyze the implications of different components that make up an RF and how they contribute to the detection of texture boundaries. By fixing a certain parameter and isolating it from other fixed parameters, we can explore the effectiveness of the isolated parameter.

Test results show that two leading factors—lagging center orientation and orientation preferences—of the RF structure have stronger links to the superior performance of TRF over VRF based boundary detection.

5. CONCLUSION

The main novelty and contribution of this paper is in the use of tactile receptive field responses for texture segmentation. Furthermore, we showed that touch-based representation is superior to its vision-based counterpart when used in texture boundary detection.

Tactile representations were also found to be more discriminable (LDA and ANOVA). We expect our results to help better understand the nature of texture perception and build more powerful texture processing algorithms.

6. REFERENCES

- [1] A. Thielscher and H. Neuman, “A computational model to link psychophysics and cortical cell activation patterns in human texture processing,” *J of Comp Neurosci*, vol. 3, no. 22, pp. 255–282, 2006.
- [2] B. Julesz, “Texton gradients: the texton theory revisited,” *Biol Cybern*, vol. 54, no. 4-5, pp. 245–251, 1986.
- [3] J. Malik and P. Perona, “Preattentive texture discrimination with early vision mechanisms,” *J of Optical Soc of Am A*, vol. 7, no. 5, pp. 923–931, 1990.

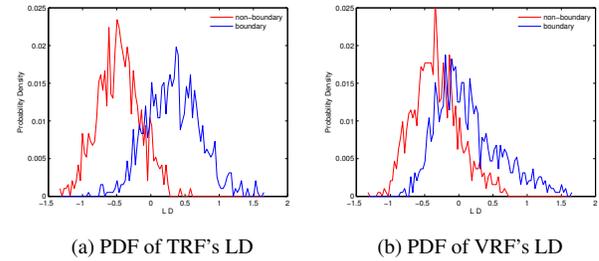


Fig. 6. Linear discriminant (LD) distributions of TRF and VRF responses. Visualization of 100 PDFs of the (a) tactile and (b) visual RF responses' LD are shown.

- [4] S. Oh and Y. Choe, “Segmentation of textures defined on flat vs. layered surfaces using neural networks: Comparison of 2d vs. 3d representations,” *Neurocomputing*, vol. 70, no. 13-15, pp. 2245–2255, 2007.
- [5] C. Spence, F. Pavani, and J. Driver, “Crossmodal links between vision and touch in covert endogenous spatial attention,” *J of Exp Psy: Human Perception and Performance*, vol. 26, no. 4, pp. 1298–1319, 2000.
- [6] E. Deibert, M. Kraut, S. Kremen, and J. Hart, “Neural pathways in tactile object recognition,” *Neurology*, vol. 52, no. 7, pp. 1413–1417, 1999.
- [7] S. Bensimaia, P. Denchev, J. Dammann, J. Craig, and S. Hsiao, “The representation of stimulus orientation in the early stages of somatosensory processing,” *J of Neurosci*, vol. 28, no. 3, pp. 776–786, 2008.
- [8] J. Jones and L. Palmer, “An evaluation of the two dimensional gabor filter model of simple receptive fields in cat striate cortex,” *J of Neurophysi*, vol. 58, no. 6, pp. 1233–1258, 1987.
- [9] S. Grigorescu, N. Petkov, and P. Kruizinga, “Comparison of texture features based on gabor filters,” *IEEE Trans on Image Proc*, vol. 11, no. 10, pp. 1160–1167, 2002.
- [10] J. DiCarlo and K. Johnson, “Spatial and temporal structure of receptive fields in primate somatosensory area 3b: Effects of stimulus scanning direction and orientation,” *J of Neurosci*, vol. 20, no. 1, pp. 495–510, 2000.
- [11] P. Brodatz, *Textures: A Photographic Album for Artists and Designer.*, Dover Publication, New York, NY, 1966.
- [12] T. Randen and J. Husoy, “Multichannel filtering for image texture segmentation,” *Optical Engineering*, vol. 33, no. 88, pp. 2617–2625, 1994.
- [13] L. Hansen and P. Salamon, “Pattern analysis and machine intelligence,” *Neural Network Ensembles*, vol. 12, no. 10, pp. 993–1001, 1990.