

Tactile or Visual?: Stimulus Characteristics Determine Receptive Field Type in a Self-organizing Map Model of Cortical Development

Choonseog Park, Yoon H. Bai and Yoonsuck Choe

Abstract—Tactile receptive fields (RFs) are similar to visual receptive fields, while there is a subtle difference. Our previous work showed that tactile RFs have advantage in texture boundary detection tasks compared to visual RFs. Our working hypothesis was that tactile RFs are better in texture tasks since texture is basically a surface property, more intimately linked with touch than with vision. From an information processing point of view, touch and vision are very similar (i.e., two-dimensional sensory surface). Then, the question is what drives the two types of RFs to become different? In this paper, we investigated the possibility that tactile RF and visual RF emerge based on an identical cortical learning process, where the only difference is in the input type, natural-scene-like vs. texture-like. We trained a self-organizing map model of the cortex (the LISSOM model) on two different kinds of input, (1) natural scene and (2) texture, and compared the resulting RFs. The main result is that RFs trained on natural scenes have RFs resembling visual RFs, while those trained on texture resemble tactile RFs. These results suggest that the type of input most commonly stimulating the sensory modality (natural scene for vision and texture for touch), and not the intrinsic organization of the sensors or the developmental process in the cortex, determine the RF property. We expect these results to shed new light on the differences and similarities between touch and vision.

I. INTRODUCTION

Sensory neurons in the primary sensory cortices in the brain preferentially respond to specific patterns of input. For example, neurons in the primary visual cortex (area 17, V1) have oriented Gabor-like receptive fields (RFs) [1] (Fig. 1). Interestingly, neurons in the somatosensory area 3b exhibit similar RF properties, with a subtle difference [2], [3], [4]. In area 3b, neurons respond to tactile input from the finger tip, and just like in the visual cortex, they only respond to a specific pattern of input. However, there is a difference between the tactile receptive fields and visual receptive fields. Instead of an excitation/inhibition pair as in Gabor patterns, there is an extra third component that is inhibitory, where the position of that component dynamically changes depending on the direction of scan of the tactile patch (Fig. 2). Given that the two sensory modalities (vision and touch) have the same basic spatial organization (i.e., a 2D sensory surface), and that the cortex is a fairly uniform medium, it is curious as to why the two RF types show this kind of difference.

One obvious reason could be that the types of input stimulating the two modalities differ in their statistical

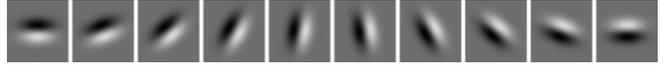


Fig. 1. Visual Receptive Fields (VRFs). Visual cortical receptive fields have a Gabor-like pattern, with different orientation, phase, and spatial frequency. Shown here are oriented Gabor patterns with the same phase and spatial frequency (dark represents inhibitory region and bright excitatory region).

characteristics. Vision is exposed more to natural scenes containing various objects and backgrounds that do not repeat over space (on a large scale), while touch is exposed more to surface texture with a regular repetition of pattern in all directions. In other words, visual RFs could have adapted to deal with natural scenes, while tactile RFs adapted to handle textures.

Texture is basically a surface property, so it may be more intimately related to touch, thus tactile RFs would be better for texture processing than visual RFs. Our previous experiment on texture boundary detection indicates that this could be the case. In [5], we showed that preprocessing with tactile RFs gives better texture boundary detection performance compared to visual RFs (see the Background section for details).

In this paper, we investigate the possibility that tactile RFs and visual RFs emerge based on an identical cortical learning process, where the only difference is in the input type: natural scene vs. texture. We trained a self-organizing map model of the cortex (the LISSOM model [6]) on two different kinds of input, (1) natural scene and (2) texture, and compared the resulting RFs. The main result is that RFs trained on natural scenes have RFs resembling visual RFs, while those trained on texture resemble tactile RFs. These results suggest that the type of input most commonly stimulating the sensory modality (natural scene for vision and texture for touch), not the intrinsic organization or developmental process, determine the RF property in the primary sensory cortices.

The rest of this paper is organized as follows. First, we will provide some background on visual and tactile processing in the brain (Sec. II). Next, we will present the LISSOM self-organizing map algorithm (Sec. III), explain in detail our experimental design, and present our results (Sec. IV). Finally, we will talk about interesting perspectives and issues regarding our results (Sec. V), followed by the conclusion (Sec. VI).

II. BACKGROUND

Visual cortical receptive fields resemble Gabor patterns [1], [7], [8], [9], with specific tuning for orientation, phase,

Choonseog Park is with Texas A&M University, College Station, TX, USA (phone: 979-204-1165, e-mail: cspark13@cs.tamu.edu).

Yoon H. Bai was with Texas A&M University, College Station, TX, USA. He is now with Samsung, Seoul, Korea (email: yoonbai@gmail.com).

Yoonsuck Choe is with Texas A&M University, College Station, TX, USA (e-mail: choe@tamu.edu).

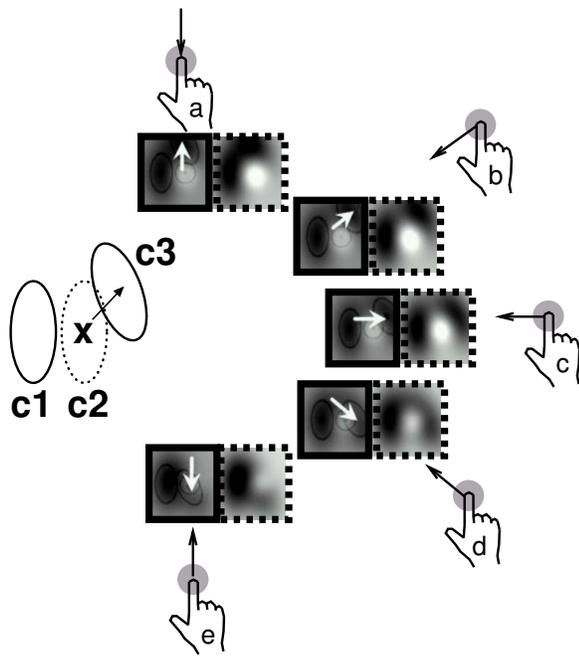


Fig. 2. Tactile Receptive Fields (TRFs). Tactile receptive fields are similar to visual receptive fields (marked C1 and C2, representing inhibitory and excitatory blobs) but there is an added dynamic inhibitory component (marked C3). An interesting feature of this extra inhibitory component is that its position relative to the fixed components C1 and C2 change, centered at “X”, depending on the direction of scan of the tactile surface (e.g., the tip of the index finger). The dynamic component’s shift in position is in the opposite direction of the scan direction. The five groups of figures to the right show how scan direction alters the tactile receptive field property. In each group, the arrow on the finger tip shows the scan direction; the box with a solid outline shows how the dynamic inhibitory component is shifted (white arrow) in the opposite direction of the scan; and the box with the dotted outline shows the resulting tactile receptive field shape. Adapted from [5] (also see [2]).

and spatial frequency. Fig. 1 shows an example of Gabor RF models of varying orientation. In addition to these static features, visual cortical neurons are also sensitive to dynamic features such as the direction of motion of the stimuli [10], [11], [12], [13]. Fig. 3 shows an example of a direction-selective RF in the cat primary visual cortex [10]. Instantaneous RFs in the two-dimensional visual space at times 20, 60, 100, and 120 ms are shown on top, and a continuous integration of the RFs along the vertical is drawn in the bottom plane. The neuron’s spatial preferences change systematically over time, giving it a spatiotemporal preference for a dark vertical line moving horizontally to the right.

As briefly discussed in the introduction, tactile RFs have similar properties as the visual counterpart (Fig. 2), since they are based on the on-off Gabor pattern. However, there is a dynamic component that curiously depends on the direction of scan of the skin patch (e.g., a finger tip). Fig. 4 shows examples of the tactile RFs estimated from the cortical area 3b of an alert monkey and the model predictions [2]. In these plots, dark represents an inhibitory region and bright an excitatory region. Each row in the figure shows the RF estimated from the raw data (left) measured through

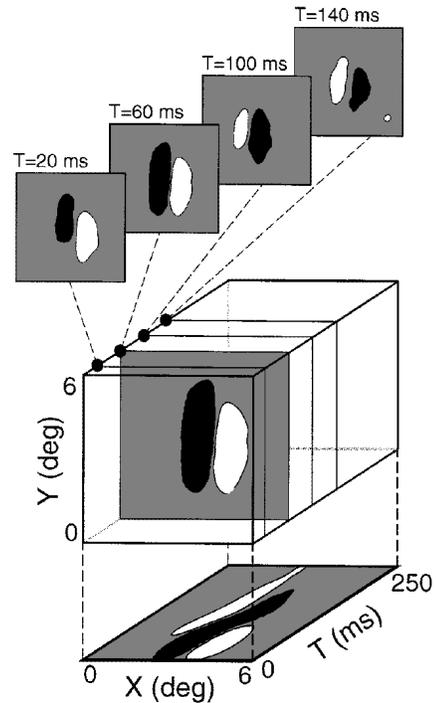


Fig. 3. Spatiotemporal Receptive Fields in Cat. Primary visual cortical neurons show direction (of motion) selectivity in addition to orientation selectivity. The figure shows the spatiotemporal pattern that optimally stimulates a visual cortical neuron in a cat. Here, we can see the dark region of a vertical Gabor pattern moving to the right. Adapted from [10] (as rendered in [6]).

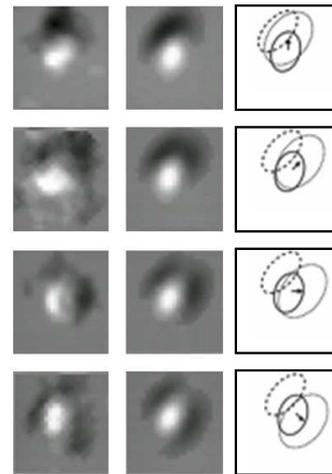


Fig. 4. Tactile Receptive Fields in Monkey Area 3b. RFs resulting from four different scanning directions on the finger tip of a monkey is shown. Each row shows, from the left to the right, (1) the actual measured RF (bright=excitation, dark=inhibition), (2) the three-component model by [2], and (3) the outline of the three-component model. Given the same excitation–inhibition pair, a third (inhibitory) component shows up, and the center of that component shifts its position in the opposite direction of the scan. For example, the top row corresponds to a downward scan, thus the third component shifts up. See Fig. 2 for a detailed explanation. Adapted from [2] (grayscale was inverted to show excitation in white and inhibition in black).

microelectrode recording in area 3b of the alert monkey, the RF predicted by a three-component model (middle), and the positions of the Gaussian components in the model (right). Three ellipses in right panels represent fixed excitatory (thick oval), fixed inhibitory (dashed oval), and lagged inhibitory (thin oval) lobe moving in the opposite direction of the scanning direction (arrow). Depending on the direction of scan, the dark blob (lagged inhibitory region) moves around, altering the final shape of the RF. The resulting RFs, even though they are based on a Gabor pattern, show patterns distinct from visual RFs, such as a donut or a curve.

In brief, visual and tactile RFs share common features but they also have subtle differences. The similarity may be driven by the fact that the visual and tactile sensory surfaces have the same structure (i.e., a 2D sheet), and that they are stimulated by the same underlying spatial environment.

However, two questions remain to be answered: (1) what could account for the difference in the two RF types?, and (2) what could be the functional benefit of such a specialization? It turns out that we can gain a lot of insight by answering the second question first. In [5], we tested the hypothesis that tactile RFs could be better than visual RFs in texture segmentation, even when the task itself is a vision task. The hypothesis was based on the insight that texture (even visual texture) may be more intimately linked to touch, since texture is basically a surface property. (This insight was also shared in our earlier work on comparing texture segmentation performance on tasks defined in 2D vs. 3D [14], [15] (cf. [16]).)

We used either the oriented Gabor RF models (visual) or DiCarlo and Johnson’s three-component RF model (tactile) as a preprocessing stage in texture segmentation. This allowed us to test how such distinct properties in tactile receptive fields can affect texture segmentation performance. Our results showed that tactile RFs have a consistently higher performance over visual RFs. Fig. 5 shows the results from our previous experiment [5], where six texture sets (a mix of natural and synthetic textures) were used in texture segmentation tasks. Fig. 5 shows the comparison of classification rate for the six texture sets. Upsetting the common assumption in most current approaches that texture segmentation is a vision problem, tactile RFs showed superior performance in texture boundary detection compared to their visual counterpart (t-test: $n=1920$, $p < 0.03$ in all, except for texture set 1, $p=0.27$) [5].

These early results suggest that the unique characteristics in the tactile RFs are functionally capable of dealing with textures in general, including visual texture, i.e., not just tactile surface textures. This leads to an interesting possibility that tactile RFs diverge from visual RFs just because they are exposed to more texture-like inputs rather than natural-scene-like inputs.

Another important observation is that the sensory cortical RFs adapt and develop over time to reflect the input statistics. Researchers have repeatedly shown that sensory cortical organization can dramatically change if the input environment

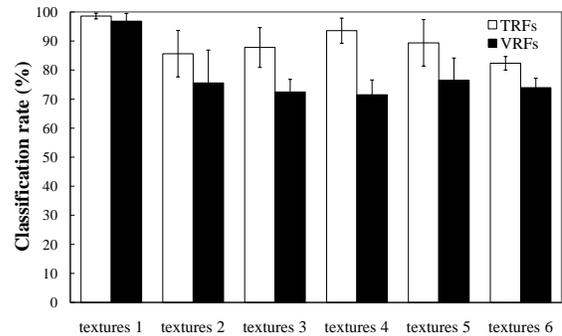


Fig. 5. Comparison of Tactile RF and Visual RF Performance on Texture Boundary Detection Tasks. In all except the first of the six texture sets, the texture boundary detection accuracy with tactile RFs (TRFs, white bars) was significantly better than those with visual RFs (VRFs, black bars). Error bars indicate the standard deviation. Adapted from [5].

is altered [9], [17], [18], and RF properties can change if the input is altered [19], [20] (see [21], [6] for modeling results). Furthermore, the cerebral cortex is quite uniform, where different sensory cortices basically have the same structure and organization [22]. A tour de force in cortical plasticity experiment also showed that such uniformity can support any form of sensory input thrown at the cortex. By plugging in visual signal into the auditory cortex, Sharma et al. showed that the auditory cortex can be transformed to have an organization similar to that of the visual cortex, with orientation-tuned neurons and an overall orientation map layout [23].

With all the pieces of the puzzle in place, we can now formulate an approach: Train a self-organizing model of visual cortical development with different types of input, natural scenes or textures, and observe the RFs that emerge from the process. The expectation is that the model trained with natural scenes will develop visual RFs and those exposed to textures will develop tactile RFs.

III. THE LISSOM MODEL AND SELF-ORGANIZATION

In order to investigate the possibility that tactile RFs and visual RFs emerge based on an identical learning process, we trained LISSOM (Laterally Interconnected Synergetically Self-Organizing Map), a self-organizing map model of the visual cortex [6].

LISSOM was originally developed to model the visual cortex, but it is actually a more general model of how the cortex organizes to represent correlations in the sensory input. Thus, LISSOM should work equally well in modeling the development of non-visual sensory modalities, as demonstrated by [24], where somatosensory cortical development (of the barrel cortex in rodents) was successfully modeled using LISSOM.

Since tactile RFs have a dynamic component, we adopted a variant of LISSOM that can handle dynamically changing input, i.e., LISSOM model of combined orientation and direction map formation. The resulting RFs in the model would have a spatiotemporal pattern.

Fig. 6 shows the LISSOM architecture for orientation and direction selectivity. The description below closely follows [6]. We mainly used the Topographica neural map simulator package for the experiments (<http://topographica.org>), developed by Bednar et al. [6]. The model is similar to a general LISSOM model consisting of two-dimensional sheet of neural units, roughly corresponding to the retina at the input level, ON- and OFF-LGN (Lateral Geniculate Nucleus) channels at the intermediate level, and V1 neurons at the cortical level. LGN units have four sheets with different time delays for each ON and OFF channel so that V1 neurons can use these time-varying inputs to develop spatiotemporal receptive fields.

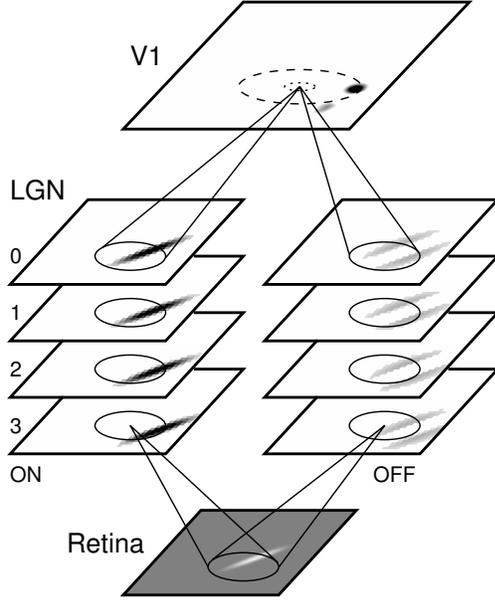


Fig. 6. LISSOM Model of Orientation and Direction Selectivity. Moving input patterns are drawn on the retina in discrete time steps, like frames of a movie. At each time step, the input pattern (or the gaze) is moved slightly on the retina and LGN cells with time step index 3, 2, 1, and 0 each compute their activity with varying delay from the retina. Once all LGN cells have been activated, the initial V1 response is computed based on the responses on the eight LGN sheets. The activity then spreads laterally within V1 through excitatory (small dotted circle in V1) and inhibitory (large dashed circle in V1) connections. Adapted from [6].

An input consists of four sequential frames of an image, moving across the retina at a certain location and direction. At each time step t , the frame t is presented on the retina, and the activities of two LGN ON/OFF cells with time t are calculated. The fixed weights for the LGN ON units are computed as:

$$L_{xy,ab} = \frac{\exp\left(-\frac{(x-x_c)^2+(y-y_c)^2}{\sigma_c^2}\right)}{\sum_{uv} \exp\left(-\frac{(u-x_c)^2+(v-y_c)^2}{\sigma_c^2}\right)} - \frac{\exp\left(-\frac{(x-x_c)^2+(y-y_c)^2}{\sigma_s^2}\right)}{\sum_{uv} \exp\left(-\frac{(u-x_c)^2+(v-y_c)^2}{\sigma_s^2}\right)}, \quad (1)$$

where $L_{xy,ab}$ is the weight from the retinal receptor (x, y) in the receptive field to an LGN ON neuron (a, b) with center

(x_c, y_c) , and σ_c defines the width of the central Gaussian and σ_s the width of the surround Gaussian. We set the size of the central Gaussian to 0.07385 and the size of the surround Gaussian to 0.29540. The OFF neuron weights are the negative of the ON weights.

After receiving input from the retina, the LGN units compute their responses as a squashed weighted sum of the total received activation:

$$\xi_{ab} = \sigma\left(\gamma_L \sum_{xy} \chi_{xy} L_{xy,ab}\right), \quad (2)$$

where ξ_{ab} is the response of the LGN ON/OFF unit (a, b) , χ_{xy} is the activation of retinal unit (x, y) within the receptive field of (a, b) , $L_{xy,ab}$ is the afferent weight from (x, y) to (a, b) , and γ_L is a constant scaling factor for LGN's afferent weight. To produce activity for low-contrast inputs of images, we set γ_L to 4.7 which is double that of the Gaussian width. Here, σ is a piecewise-linear approximation of the sigmoid activation function:

$$\sigma(s) = \begin{cases} 0, & s \leq \theta_l \\ (s - \theta_l)/(\theta_u - \theta_l), & \theta_l < s < \theta_u \\ 1, & s \geq \theta_u \end{cases}, \quad (3)$$

where s is the activation level of the neuron, θ_l is the lower bound and θ_u is the upper bound. While we used the default value for the initial upper bound ($\theta_u = 0.038$) in the Topographica package, the initial sigmoid lower bound was set to a lower value ($\theta_l = 0.076$) than the default value to allow responses to low-contrast stimuli. Both of them are gradually adjusted during self-organization to enhance the performance.

Fig. 7 shows typical inputs, and Fig. 8 shows how the inputs are sampled, to be fed into the LGN layers. Moving input patterns following a scanning direction are drawn on the retina in discrete time steps, like frames of a movie. At each time step, LGN cells compute their activities based on the moving input pattern on the retina.

After all four frames are drawn on the retina, one after another, the LGN sheets are activated in sequence, with the prescribed time delay. Then, each V1 neuron computes its initial response projected from the activation on all eight LGN ON/OFF sheets. The initial response of V1 neurons is computed as a weighted sum of activation received from the LGN and then passed through a sigmoid activation function:

$$s_{ij} = \gamma_A \left(\sum_{ab \in ON} \xi_{ab} A_{ab,ij} + \sum_{ab \in OFF} \xi_{ab} A_{ab,ij} \right) \quad (4)$$

$$\eta_{ij}(0) = \sigma(s_{ij}), \quad (5)$$

where $\eta_{ij}(0)$ is the initial response of V1 neuron (i, j) , s_{ij} is the afferent activation of V1 neuron (i, j) , ξ_{ab} is the activation of LGN ON/OFF neuron (a, b) in the receptive field of V1 neuron (i, j) , $A_{ab,ij}$ is the afferent weight, and γ_A is a constant scaling factor for the afferent weight. We set the γ_A to 1 which is a default value for LISSOM simulations in Topographica.

After the initial computation, V1 calculate lateral excitatory and inhibitory contributions to settle the activity:

$$\eta_{ij}^{\text{new}}(t) = \sigma(s_{ij} + \gamma_E \sum_{kl} \eta_{kl}^{\text{pre}}(t-1) E_{kl,ij} - \gamma_I \sum_{kl} \eta_{kl}^{\text{pre}}(t-1) I_{kl,ij}), \quad (6)$$

where $\eta_{kl}^{\text{pre}}(t-1)$ is the activity of the neighbor the V1 neuron (k, l) in the previous time step, $E_{kl,ij}$ is the excitatory lateral connection weight connecting neuron (i, j) and (k, l) , $I_{kl,ij}$ is the inhibitory lateral connection weight, and γ_E and γ_I are scaling factors that determine the strength of excitatory and inhibitory lateral interactions.

Because image patterns have significant long-range correlations, and inhibitory weights spread over a larger area, the lateral interaction strength γ_E and γ_I were set to 0.9 and -0.9 to keep the balance between excitatory and inhibitory lateral weights approximately constant.

After the activity settles, the afferent and lateral connection weights of V1 neurons are modified according to the Hebbian learning rule:

$$W_{pq,ij}^{\text{new}} = \frac{w_{pq,ij}^{\text{cur}} + \alpha X_{pq} \eta_{ij}}{\sum_{uv} (w_{uv,ij} + \alpha X_{uv} \eta_{ij})}, \quad (7)$$

where $w_{pq,ij}^{\text{cur}}$ is the current connection weight from neuron (p, q) to (i, j) , $w_{pq,ij}^{\text{new}}$ is the new connection weight, α is the learning rate for each type of connection, X_{pq} is the presynaptic activity after settling, and η_{ij} is the activity of neuron (i, j) after settling.

To enhance the resulting self-organization of the lateral inhibitory weights into long range regions, the lateral inhibitory learning rate α was updated over time; 0.090365 at first, 0.090365×2 at 1,000 iterations, 0.090365×3 at 2,000 iterations, and 0.090365×5 at 5,000 iterations.

IV. EXPERIMENTS AND RESULTS

The main experiment we did was to test the possibility that tactile RFs and visual RFs emerge base on an identical learning process, where the only difference is in the input environment, natural scene vs. texture. We used the self-organizing map model of the cortex (the LISSOM model) on two different kinds of input, (1) natural scene and (2) texture, and analyzed the spatiotemporal properties of the resulting RFs. Fig. 7 shows the two types of input patterns.



Fig. 7. Sample Input Patterns. The top row shows natural scenes and the bottom row textures used in our experiments. Note that the texture set have texture elements are varying scales.

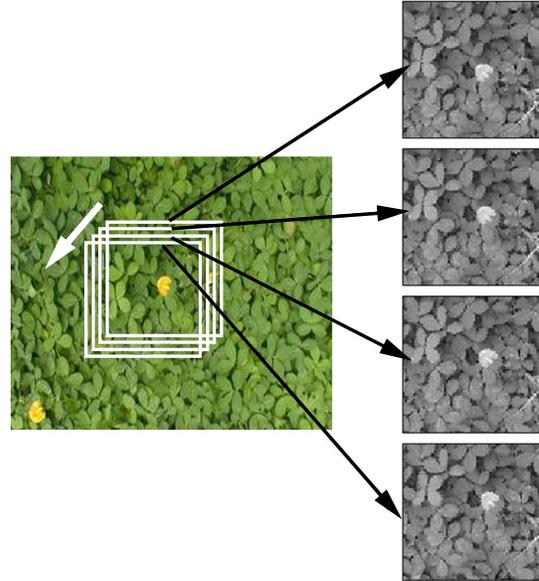


Fig. 8. Generation of Dynamic Input by Scanning the Gaze on an Image. Given a large image, a small region the size of the retina in the LISSOM model is attended to. Motion of the gaze window results in a sequence of inputs being generated on the LISSOM retina, which in turn activates the LGN ON/OFF sheets, one by one depending on the sheet's built-in delay.

We used a variant of the LISSOM that can learn both the orientation and the direction of the visual stimuli. Given an image, we randomly picked an initial location and moved the gaze window in a random direction along a straight line at a fixed interval as shown in Fig. 8. Each gaze location gave a 48×48 image patch that was the same size as the retina in the LISSOM simulation. All four images in each image type were used to generate the input sequence for each simulation. All the LGN sheets were 24×24 in size, and the single V1 sheet was 48×48 in size.

All simulations in this work were based on the same set of default parameters in the Topographica package, with small modifications described in the previous section. Specifically, we set the speed of the input pattern (i.e., the number of retinal units the pattern moves between time steps) to 4 pixels/time step. This value was determined experimentally, as neurons showed the best selectivity for direction of motion at that speed. After training for several thousand iterations (usually between 10,000 to 20,000 iterations), the network developed patterned afferent and lateral connections for retina to LGN, LGN to V1 and V1 to V1. We focused on the projection from LGN to V1 because they represent the spatiotemporal character of the RF.

Fig. 9 shows the self-organized RFs of six representative neurons trained with natural scenes. Nearly all neurons developed spatiotemporal RFs strongly selective for both direction and orientation. That is, each neuron is highly responsive to a line with a particular orientation moving in a direction perpendicular to that orientation. The receptive fields consist of white (excitatory) and black (inhibitory) lobes according to the preferred orientations and direction of the neuron,

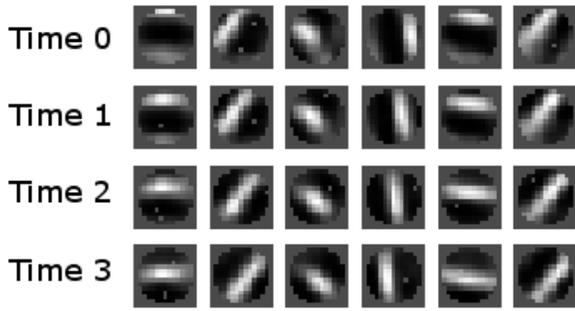


Fig. 9. RFs Resulting from Self-Organization on the Natural Scene Input Set. Six spatiotemporal RFs from the natural scene experiment are shown. Each column corresponds to an individual neuron's RF, and each row represents the different time-lag. Within each column, we can see that the pattern moves in a direction perpendicular to the orientation preference.

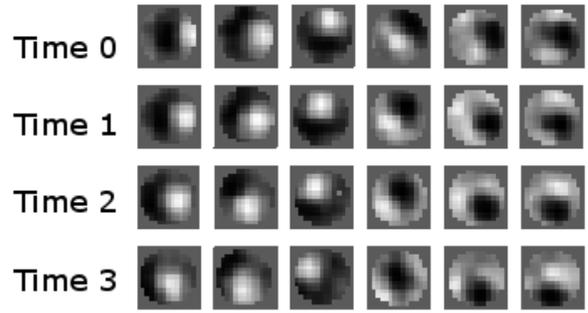


Fig. 11. RFs Resulting from Self-Organization on the Texture Input Set. Six spatiotemporal RFs from the texture experiment are shown. Each column corresponds to an individual neuron's RF, and each row represents the different time-lag. The RF shapes resemble the ring-like shape of tactile RFs found in the experimental literature [2].

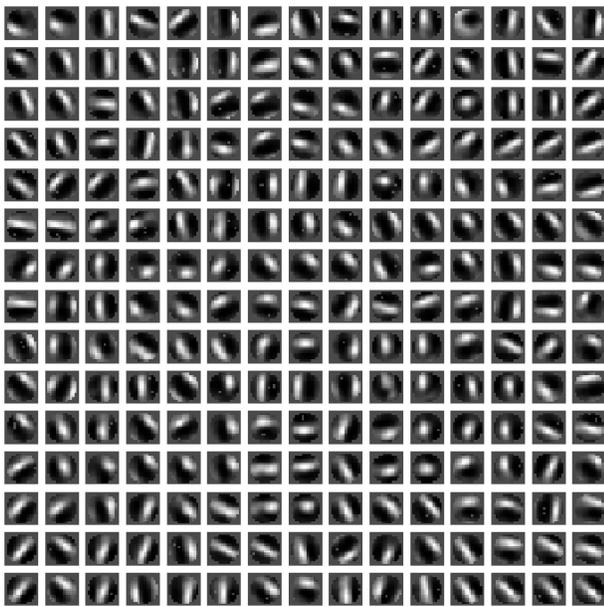


Fig. 10. RFs Resulting from Self-Organization on the Natural Scene Input Set. From the 48×48 cortex, only 15×15 are plotted (roughly every 3rd RF) for a detailed view of the RFs. The RFs mostly resemble visual RFs.

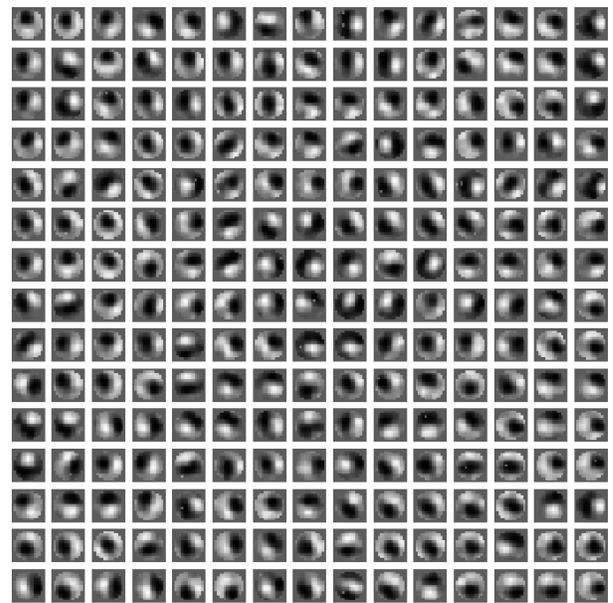


Fig. 12. RFs Resulting from Self-Organization on the Texture Input Set. From the 48×48 cortex, only 15×15 are plotted (roughly every 3rd RF) for a detailed view of the RFs. The RFs mostly resemble tactile RFs.

showing spatiotemporal preference. Such properties of the receptive fields are similar to those of the receptive fields of neurons found experimentally in the visual cortex [10] (cf. Fig. 3).

The overall layout of the RFs developed in this simulation, trained with four natural scenes, are shown in Fig. 10 by plotting roughly every 3rd neuron horizontally and vertically. A number of two-lobed RFs can be seen with strong orientation preferences except some neurons which have nonlinear-shape and respond to all directions. This figure only shows the first frame among the total of four (note that these are spatiotemporal RFs).

The self-organized RFs produced from LISSOM after 20,000 training iterations with the texture input set are visualized in Fig. 11. The neurons developed spatiotemporal RFs strongly resembling tactile RFs found in the experimental

literature (e.g. Fig. 4). Excitatory and inhibitory components of each neuron consists of ring and blob-like features as found in [2]. For example, the first three column in Fig. 11 closely resemble the RFs in Fig. 4. Note that these RF shapes arise not because circular texture elements dominate the texture input set we used. On closer observation, the texture input set we used show texture elements at varying scales, and also, the size of the receptive field (15×15) is usually smaller than the round or oval objects in the texture input set. There are interesting variations (last three columns in Fig. 11) where the polarity is reversed, i.e., instead of an excitatory region in the middle and inhibitory region in the surround, these RFs have an inhibitory region in the middle and the excitatory region in the surround. Fig. 12 shows the overall organization of the cortical map (roughly every 3rd neuron's RF is shown).

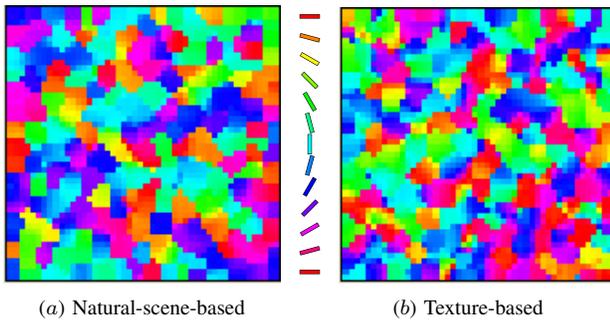


Fig. 13. Orientation Maps. The orientation maps resulting from (a) the natural scene experiment and (b) the texture experiment are shown. The color-key for the orientation is shown in the middle.

The results show that exposure to different kinds of input can drive an identical underlying cortical learning model to develop two different kinds of RFs, tactile or visual. RFs trained on natural scenes resemble visual RFs, while those trained on texture resemble tactile RFs. These results suggest that the type of input most commonly stimulating the sensory modality (natural scene for vision and texture for touch), and not the intrinsic organization or developmental process, determine the RF property.

Our experiments also turned up interesting preliminary results. The global organization of the RFs can be visualized just like for biological orientation maps, by labeling each neuron by the preferred angle. Fig. 13 shows the overall organization of the two maps. The orientation map developed with natural scene input shows similar characteristics as those found in the visual cortex, with smooth transition across neighboring orientation domains. The map also has more red and cyan than other colors, which means high selectivity for near horizontal and vertical orientation, which is also observed in experiments. It is interesting to note that the map trained on textures (i.e., those that develop tactile RFs) also has a rough orientation map. Note that in the tactile map case, the concept of orientation is less well defined than the visual map case, since not all RFs have a clear orientation preference. Thus, the selectivity may be low. Such a prediction (orientation map, and low selectivity) can be verified experimentally in area 3b, using optical imaging techniques.

V. DISCUSSION

The main contribution of this work is to have shown what drives a virtually identical learning medium (the cerebral cortex) to specialize and diverge, to represent different sensory modalities, visual vs. tactile. By comparing two structurally similar modalities of touch and vision, along with the assumption that texture (whether visual or tactile) is basically a surface property, we have shown that it is the input statistics (natural-scene-like or texture-like) that determine the learned RF type, not the direct sensory modality. These and our earlier results [14], [15], [5] show an intricate relationship among touch, texture, and surface property in 3D.

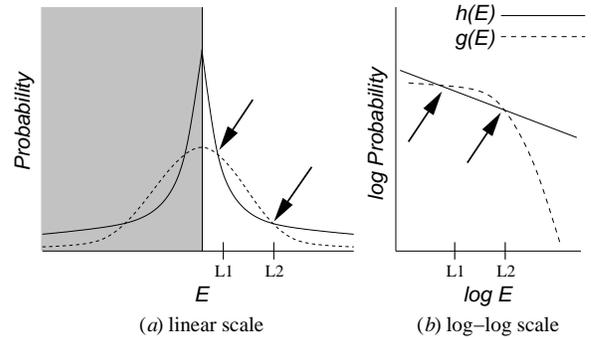


Fig. 14. Power-law vs. Baseline Gaussian Distribution. The response distribution $h(E)$, where E is the activation level, and the matching Gaussian distribution $g(E)$ that has the same standard deviation is illustrated. When the two curves intersect at the point marked L2, the probability of the response distribution becomes higher than the baseline Gaussian distribution. This point has been shown to play an important role in saliency thresholding [25], [26]. Note that the shaded part in (a) is only for illustrative purposes (to show the peakedness of $h(E)$), since the response E is always positive. Adapted from [25].

In the background section, we raised two questions, regarding (1) the functional role and (2) the developmental origin of tactile RFs. There are some more interesting properties of the tactile RF that can link to its functionality. In an earlier work, we have shown that the power-law-like response distribution in a visual cortical neuron model can help subsequent stages in the cortical processing to easily extract salient features in the input, such as edges and contours [25], [26] (cf. [27]). The basic idea was that the heavy-tail part in the power-law distribution, compared to a Gaussian baseline, can accurately predict the response threshold (Fig. 14).

Interestingly, the response distribution of the tactile map also shows a power-law property, even though the spatiotemporal structure of the RF is different from the visual RFs (Figs. 15 and 16). However, we should note that the input statistics was different, natural-scene-like vs. texture-like. Thus, we can speculate that one goal of early sensory processing is to generate RF coding that maps the specific input distribution into a canonical response distribution that can be easily utilized by a similar second stage of processing, regardless of the modality. These observations raise the interesting possibility that later stages in multi-modal sensory processing may share a common, integrated mechanism, thanks to the customized encoding done at the early stages of processing. (In a similar spirit, [28] and [29] show how requirements in subsequent stages of cortical processing can dictate initial sensory encoding strategy.)

The observations above can lead to interesting future research. For example, we can check if response distribution of tactile maps responding to natural scene inputs maintain the power-law property. Our prediction is that the power-law property will not be maintained in such a case. A dual experiment with visual maps can also be done. We can also extend the saliency thresholding approach we developed in [26] to the tactile domain, exploiting the power-law response distribution.

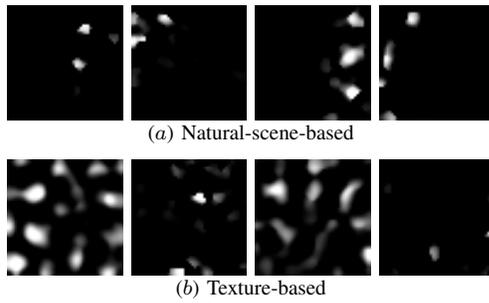


Fig. 15. Cortical Response. Cortical responses of (a) the natural-scene-based cortex and (b) the texture-based cortex are shown (bright represents high and dark represents low activity). Both show a sparse activation profile. See Fig. 16 for the response histogram.

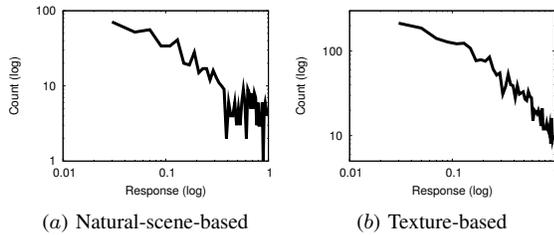


Fig. 16. Response Distribution. The response histogram of (a) the natural-scene-based RFs and (b) the texture-based RFs are shown, in a log-log scale. The histograms were calculated from the response matrix in Fig. 15. Both show a power-law distribution. (Note that the first and the last bin are not plotted, because those values were artificially exaggerated due to the use of the piece-wise linear activation function that had a hard lower and upper bound [Eq. 3].)

VI. CONCLUSION

The main finding of this work is that a common cortical development framework (LISSOM) can develop two different RF types, just based on the type of input presented during training. The results suggest that texture in general, whether it is tactile or visual in origin, contributes to the emergence of the unique properties similar to those observed in tactile RFs in area 3b of the somatosensory cortex. This is an interesting result that helps us better understand the intimate relationship among texture, surface, and touch, and further strengthens our earlier finding that tactile RFs can outperform visual RFs in texture segmentation tasks. We expect our new tactile-oriented approach to texture segmentation to complement the traditional visually oriented approach, and help us better understand the nature of texture.

REFERENCES

- [1] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex," *Journal of Neurophysiology*, vol. 58, no. 6, pp. 1233–1258, 1987.
- [2] J. J. DiCarlo and K. O. Johnson, "Spatial and temporal structure of receptive fields in primate somatosensory area 3b: Effects of stimulus scanning direction and orientation," *Neuroscience*, vol. 20, pp. 495–510, March 2000.
- [3] E. Deibert, M. Kraut, S. Kremen, and J. Hart, "Neural pathways in tactile object recognition," *Neurology*, vol. 52, pp. 1413–1417, 1999.
- [4] S. Bensimaia, P. Eechev, J. D. Maass III, J. Craig, and S. Hsiao, "The representation of stimulus orientation in the early stages of somatosensory processing," *Neuroscience*, vol. 28, pp. 776–786, 2008.
- [5] Y. H. Bai, C. Park, and Y. Choe, "Relative advantage of touch over vision in the exploration of texture," in *Proceedings of the 19th International Conference on Pattern Recognition, Tampa, FL, 2008*, in press.

- [6] R. Miikkulainen, J. A. Bednar, Y. Choe, and J. Sirosh, *Computational Maps in the Visual Cortex*. New York, NY: Springer, 2005.
- [7] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision Research*, vol. 20, pp. 847–856, 1980.
- [8] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *Journal of Physiology*, vol. 195, pp. 215–243, 1968.
- [9] —, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology (London)*, vol. 160, pp. 106–154, 1962.
- [10] G. C. DeAngelis, G. M. Ghose, I. Ohzawa, and R. D. Freeman, "Functional micro-organization of primary visual cortex: Receptive-field analysis of nearby neurons," *Neuroscience*, vol. 19, pp. 4046–4064, 1999.
- [11] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman, "Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development," *Journal of Neurophysiology*, vol. 69, no. 4, pp. 1091–1117, April 1993.
- [12] A. Shmuel and A. Grinvald, "Functional organization for direction of motion and its relationship to orientation maps in cat area 18," *The Journal of Neuroscience*, vol. 16, pp. 6945–6964, 1996.
- [13] M. Weliky, W. H. Bosking, and D. Fitzpatrick, "A systematic map of direction preference in primary visual cortex," *Nature*, vol. 379, pp. 725–728, 1996.
- [14] 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, pp. 2245–2255, 2007.
- [15] —, "Texture segmentation in 2D vs. 3D: Did 3D developmentally precede 2D?" in *Proceedings of the 2004 International Conference on Development and Learning [electronic]*, J. Triesch and T. Jebara, Eds. UCSD Institute for Neural Computation, 2004, pp. 175–182.
- [16] K. Nakayama and S. Shimojo, "Experiencing and perceiving visual surfaces," *Science*, vol. 257, pp. 1357–1363, 1992.
- [17] D. H. Hubel, T. N. Wiesel, and S. LeVay, "Plasticity of ocular dominance columns in monkey striate cortex," *Philosophical Transactions of the Royal Society of London Series B*, vol. 278, pp. 377–409, 1977.
- [18] D. H. Hubel and T. N. Wiesel, "Sequence regularity and geometry of orientation columns in the monkey striate cortex," *Journal of Comparative Neurology*, vol. 158, pp. 267–294, 1974.
- [19] B. Chapman and M. P. Stryker, "Development of orientation selectivity in ferret primary visual cortex and effects of deprivation," *Journal of Neuroscience*, vol. 13, no. 12, pp. 5251–5262, Dec. 1993.
- [20] M. Weliky and L. C. Katz, "Disruption of orientation tuning in visual cortex by artificially correlated neuronal activity," *Nature*, vol. 386, no. 6626, pp. 680–685, Apr 17 1997.
- [21] Y. Choe and R. Miikkulainen, "Contour integration and segmentation in a self-organizing map of spiking neurons," *Biological Cybernetics*, vol. 90, pp. 75–88, 2004.
- [22] V. B. Mountcastle, "An organizing principle for cerebral function: The unit module and the distributed system," in *Neuroscience, Fourth Study Program*, F. O. Schmitt, Ed. Cambridge, MA: MIT Press, 1979, pp. 21–42.
- [23] J. Sharma, A. Angelucci, and M. Sur, "Induction of visual orientation modules in auditory cortex," *Nature*, vol. 404, pp. 841–847, 2000.
- [24] S. Wilson, "Self-organisation can explain the mapping of angular whisker deflections in the barrel cortex," Master's thesis, The University of Edinburgh, Scotland, United Kingdom, 2007.
- [25] H.-C. Lee and Y. Choe, "Detecting salient contours using orientation energy distribution," in *Proceedings of the International Joint Conference on Neural Networks*. IEEE, 2003, pp. 206–211.
- [26] S. Sarma and Y. Choe, "Salience in orientation-filter response measured as suspicious coincidence in natural images," in *Proceedings of the 21st National Conference on Artificial Intelligence*, Y. Gil and R. Mooney, Eds., 2006, 193–198.
- [27] D. J. Field, "What is the goal of sensory coding?" *Neural Computation*, vol. 6, pp. 559–601, 1994.
- [28] E. Salinas, "How behavioral constraints may determine optimal sensory representations," *PLoS Biology*, vol. 4, pp. 2383–2392, 2006.
- [29] Y. Choe, H.-F. Yang, and N. Misra, "Motor system's role in grounding, receptive field development, and shape recognition," in *Proceedings of the Seventh International Conference on Development and Learning*, 2008.