

A model of the formation of a self-organized cortical representation of color

A. Ravishankar Rao, Guillermo Cecchi, Charles Peck and James Kozloski
IBM T.J. Watson Research Center
Yorktown Heights, NY 10598
{ravirao,gcecchi,cpeck,kozloski}@us.ibm.com

ABSTRACT

In this paper we address the problem of understanding the cortical processing of color information. Unravelling the cortical representation of color is a difficult task, as the neural pathways for color processing have not been fully mapped, and there are few computational modelling efforts devoted to color. Hence, we first present a conjecture for an ideal target color map based on principles of color opponency, and constraints such as retinotopy and the two dimensional nature of the map.

We develop a computational model for the cortical processing of color information that seeks to produce this target color map in a self-organized manner. The input model consists of a luminance channel and opponent color channels, comprising red-green and blue-yellow signals. We use an optional stage consisting of applying an antagonistic center-surround filter to these channels. The input is projected to a restricted portion of the cortical network in a topographic way.

The units in the cortical map receive the color opponent input, and compete amongst each other to represent the input. This competition is carried out through the determination of a local winner. By simulating a self-organizing map for color according to this scheme, we are largely able to achieve the desired target color map. According to recent neurophysiological findings, there is evidence for the representation of color mixtures in the cortex, which is consistent with our model. Furthermore, an orderly traversal of stimulus hues in the CIE chromaticity map correspond to an orderly spatial traversal in the primate cortical area V2. Our experimental results are also consistent with this biological observation.

1. INTRODUCTION

A fundamental characteristic of the primate brain is the presence of cortical maps. In these maps, responses to similar stimuli features tend to be mapped in a contiguously and orderly fashion in a two-dimensional cortical space, albeit punctuated by discontinuities. These cortical maps are learnt for different input spaces through the process of self-organization. Each location on this map responds best to a preferred stimulus. The best studied spatial map from both the biological and computational viewpoints is the spatial map in visual area V1.¹ In this map, there are neurons which respond best to preferred orientations of lines in the visual input field. The map is spatially organized and exhibits structures such as singularities which arise from the mapping of a high dimensional input space onto a two-dimensional space.

Though color is an important visual cue, its cortical representation has received attention only recently. In this paper we investigate a computational model for the cortical organization of color processing areas. We draw upon biological constraints and from the previously existing literature on self-organizing maps.

We use the term cortical unit as an abstraction that represents a sufficiently small area of cortex with a homogeneous response to the stimulus of interest. This unit can be as small as a single neuron or can be a larger area such as a cortical minicolumn,² which consists of a few hundred neurons with similar response properties.

We begin our task by considering the biological constraints regarding color that are relevant to the scope of this paper.

1. There are three types of cones in the retina, termed S, M and L cones that perform wavelength encoding of spectral light distributions ³[p. 48]. These cones respond best to short, medium and long wavelength distributions respectively. This creates a retinal representation that is analogous to the RGB representation in digital images.

2. The signals generated by the cones are processed to create an opponent color system. A simple mechanism to perform this is to create differences of color signals, such as the difference between red and green signals, and the difference between blue and yellow signals⁴[p. 212]. Another mechanism to create opponent colors is to use antagonistic center-surround spatial filtering⁵[p. 529]. The center-surround spatial filtering refers to a filter kernel whose weights have one sign (all positive or all negative) within a central region, and opposite sign (all negative or all positive) outside the central region, termed the surround. This type of spatial filtering is known to occur at the retinal ganglion cells as well as lateral geniculate nucleus cells (LGN), which convey retinal information to the cortex.
3. The LGN cells project in a diffuse manner to the cortex, while maintaining retinotopy. Thus, nearby locations in the retinal field are mapped to nearby locations in the cortex.
4. Color is processed in clustered and segregated areas of primate V1, known as blobs. Each blob is dedicated to processing one color opponency, red-green or blue-yellow. According to Landisman and Ts'o,⁶ patches of color selectivity extend beyond the blob areas, and include cells representing a mixture of the two color opponent channels. These are termed color-selective bridges between blobs. Thus, there appears to be a representation of mixed colors in area V1.
5. Xiao *et al*⁷ reported that the thin stripe areas of V2 contain maps where the color of a stimulus is represented by the activation at a specific cortical location. They showed an orderly correspondence between the traversal of colors along the hues in a CIE chromaticity diagram and the traversal of cortical locations corresponding to those colors.
6. There is lateral connectivity amongst different cortical units.
7. There are connections between cortical areas such as V1 and V2.

The approach we have taken in this paper is to first use the above biological constraints to propose an ideal cortical color map in Section 3.1, and to subsequently develop a computational model, as described in Sections 3.2- 3.5. The goal of our modeling effort is to devise a self-organization scheme that creates the ideal cortical color map.

2. BACKGROUND

In this section we briefly survey existing work on the computational cortical models of color processing. Saarinen and Kohonen⁸ presented one of the early models for the self-organized formation of color maps. They used the Kohonen learning algorithm⁹ to train a network that received color inputs, which were derived through spectral modeling. They used a cortex of size 15x15, such that the same input was delivered to all the cortical units. The neighborhood size initially was the same as the cortex size. However, this is not biologically plausible, as the cortex receives topographically mapped inputs, and the intra-cortical connectivity is restricted to local neighborhoods.

Barrow *et al*¹⁰ have developed a model for cortical color blob formation which integrates the processing of oriented line information as well as color information. They modeled the input pathway from retina to cortex through six channels, consisting of two luminance center-surround and four red-green center surround channels. Their simulation results show that the cortex is largely selective to orientation, except at small clusters of cells which show only color selectivity. However, the spatial organization of a large number of color-selective areas was not studied in their model.

Doi *et al*¹¹ present a model for color processing which uses the placement of retinal cones in a mosaic pattern. This creates a more detailed representation of the retinal inputs as compared with other approaches, which assume that L-, M- and S- cone photoreceptors are co-located at a retinal unit. They focus on the processing performed by the retinal units, where the goal is to reduce the redundancy among the outputs of the retinal units. Their work does not analyze the cortical processing of these color signals.

Dow¹² has presented a model for the organization of orientation and color columns on a 2D cortical sheet. He considers the role of color opponency in the creation of this map, and arranges colors around achromatic singularities.

There is a large body of work on cortical organization to monochromatic inputs, which shows that the cortex self-organizes into an orientation map consisting of orientation selective units.¹ However, there is comparatively little work on self-organizing maps devoted to color processing. This provided us the motivation to address this problem in this paper.

3. PROPOSED MODEL

We first present a conjecture for an ideal or expected target color map in Section 3.1, partly because there is insufficient biological data available regarding such a color map. Furthermore, having a suitable target can guide the computational experimentation process, and gives us a reference we can compare our results against.

Next, we investigate the representation of color inputs using two different models: a single color opponent model in Section 3.2 and a double color opponent model in Section 3.3. These color input models then presented to a simulated cortex with specific connectivity patterns as described in Section 3.4. The cortex then self-organizes based on the model described in Section 3.5.

3.1. Target Map

We hypothesize an ideal target map designed to meet the following requirements. These requirements are derived from observations about current theories on opponent color processing, the organization of two dimensional cortical maps, as presented in Section 1, and on our own intuition.

1. All color and luminance values should be represented in the map. In other word, all color ratios, such as cyan, purple, orange, and yellow-green should be represented. Thus, given a hue, it should map onto a specific location in the cortex. This represents a hypothesis we make that generalizes the findings of Xiao *et al.*⁷
2. The achromatic regions, corresponding to black and white should alternate. These represent singularities, where there is no color information, or no color specificity.
3. Opponent colors should not be adjacent to each other, and should be represented on opposite sides of singularities. Thus, the opponent colors red and green fall on opposite sides of a singularity, and so do the colors blue and yellow.
4. The red-green and blue-yellow axes are orthogonal to each other.
5. The representation should tile across a 2D plane. Thus it should be possible to find an elementary area which when tiled can cover the two-dimensional plane.
6. The color at a location can be mixed from the colors of its nearest neighbors.
7. A mechanism is required for the intensity encoding of the color stimulus. A proposed mechanism is that the number of active units around a cortical unit representing a specific hue is proportional to the input intensity of that hue.

These requirements have been used to produce Figure 1, which shows the hypothesized ideal cortical color map. The achromatic regions, corresponding to black and white in Figure1(a) represent singularities. Figure1(b) shows an interpolated version of the map in Figure1(a) to produce a greater variety of color mixtures. Figure1(a) may be interpreted as a coarse rendition of the ideal color map, and Figure1(b) as a fine-grained rendition. One can see that the requirements specified above have been met in the rendition of the ideal color map, such as the representation of all color ratios, and the tiling nature of the representation.

Having presented an ideal cortical color map, we move to the problem of arriving at this map through a self-organizing computational model. The first aspect we cover is the representation of the color information at the retinal input level. We describe two different models for representing color input information.

3.2. Color representation using single color opponency

The first model is an opponent color model, as described by Boynton⁴[p. 212] and used by Saarinen and Kohonen.⁸ As in Boynton, we use the notation of RGB for the outputs of the R, G and B cones. The RGB values can also be interpreted as the outputs of L, M and S cones. These values are transformed into a luminance channel and two opponent color channels according to the following set of equations.

$$\begin{aligned} L &= R + G \\ C_1 &= (R - G) \\ C_2 &= (R + G) - B \end{aligned} \tag{1}$$

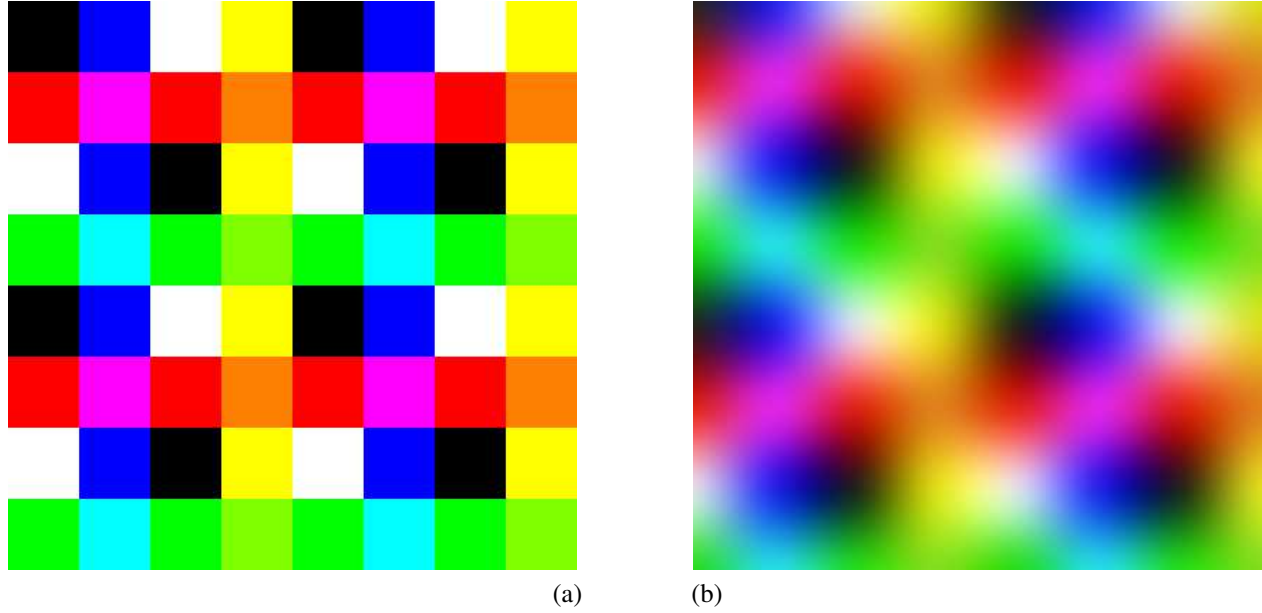


Figure 1. (a)The ideal color map at low resolution. (b)The ideal colormap at a higher resolution.

where C_1 and C_2 represent the two chromatic channels. We note that C_1 corresponds to a red-green opponent channel and C_2 corresponds to a yellow-blue opponent channel. We assume that the responses, L , C_1 and C_2 can be created from an input color image by cells in the retina. We co-locate the three color planes in a single unit, which is a reasonable simplification since the cones project to the same cortical area.

Thus, the input layer consists of a 2D sheet of units, arranged in a rectangular grid, such that each unit produces a 3-tuple, (L, C_1, C_2) that encodes the spectral wavelength properties of a received light distribution. We perform an additional step of equalizing the range of the L , C_1 and C_2 values to be in the range $[-0.5, 0.5]$ through a linear transformation.

3.3. Alternate color representation using double color opponency

In this representation, the color opponent red-green and yellow-blue channels described above undergo a spatially opponent center-surround filtering, as shown in Figure 2. Consider a unit whose receptive field (the set of units that it receives inputs from) consists of two concentric circles. The term spatially opponent means that the central portion of the receptive field, indicated by the shaded circle in Figure 2, is processed with a sign that is opposite to the sign of the surround, indicated by the larger circle. Thus, one has combinations of the type positive-center vs. negative surround, which is termed “on-center”, or negative-center vs. positive surround, which is termed “off-center”. An “on-center” unit responds to a positive stimulus in the center region, and is inhibited by a positive stimulus in the surround. This filtering occurs at a pre-cortical stage, such as in the retinal ganglion cells or the LGN (lateral geniculate nucleus).

3.4. Connectivity patterns

In order to avoid the impression that the cortical area we are simulating is necessarily area V1, we do not model the retina to LGN to cortex connectivity. Rather, we consider inputs from an input layer to the cortical sheet which is responsible for categorizing color. The inputs could well arise from a lower-level cortical area. The only requirement is that the inputs convey the color opponent information as described in Sections 3.2 or 3.3.

Each input unit projects to multiple cortical units as shown in Figure 3(a). As shown in the figure, the projection pattern is circularly symmetric. Each input unit projects topographically to cortical units within a circle of radius r_{IC} , where the subscript indicates input layer to cortex. Other projection patterns may be used as well, though they do not affect the qualitative nature of the results.

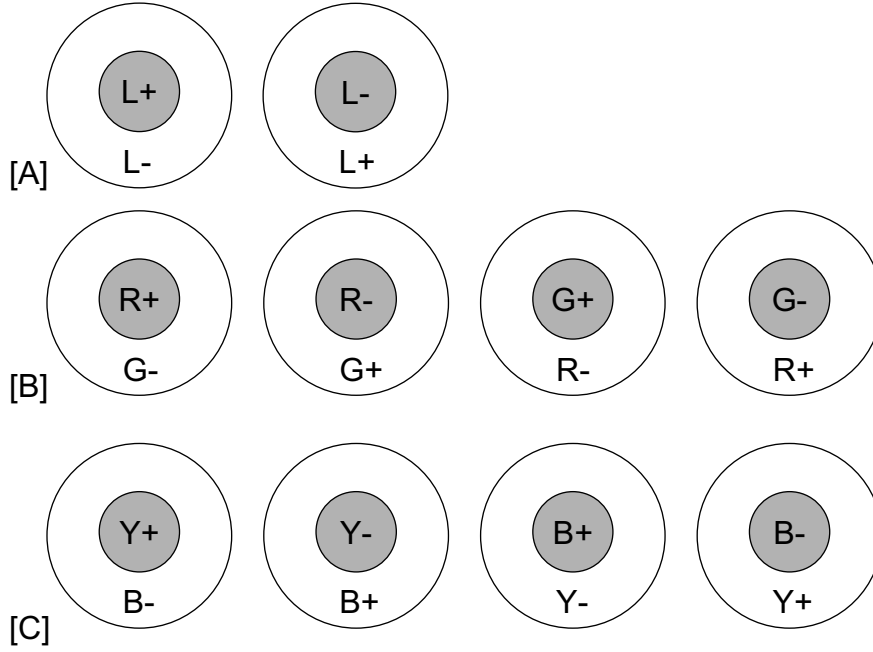


Figure 2. (A) Center-surround on and off organization of luminance channel (B) center-surround on/off organization of red-green opponent channel (C) center-surround on/off organization of yellow-blue opponent channel

Cortical units are laid out in a two dimensional array. Each cortical unit receives connections from a local neighborhood of radius r_{CC} where the subscript denotes cortico-cortical. This is shown in Figure 3(b), where a cortical unit in the center, denoted in black, has a local neighborhood shown by the shaded units. This local neighborhood serves two purposes. First, it allows the central unit to determine if it is a local winner in a competition by comparing its value with those in this neighborhood. Second, in the event that the central unit is a local winner, the weight vectors of all the elements in the local neighborhood are updated along with the winner. We use a two-dimensional circularly symmetric Gaussian weighting function centered on the local winner so that the effect of a winner on its local neighborhood diminishes with distance.

3.5. Algorithm for weight updates

The basic operation of the network is as follows. Let X denote the input vector from the input layer to the cortex. Let D denote the number of dimensions in the input vector X . The range of each input dimension is $[-1, 1]$. We will use a subscript j to index the input unit, and a superscript k to indicate the index of the input dimension, where $k \in \{1, \dots, D\}$.

Let w denote a synaptic weight, which represents the strength of the connection between two units. Let the subscript ij denote a connection from the i^{th} input unit to the j^{th} cortical unit. A superscript k similarly indexes the weight corresponding to the k^{th} input dimension. The output y_j of the j^{th} cortical unit is given by

$$y_j = \sum_k \sum_i w_{ij}^k X_i^k \quad (2)$$

Here the cortical unit combines the responses from the D different dimensions from the input layer.

The next step is for each cortical unit to determine whether it is a winner within its local neighborhood. Let N_j denote the local neighborhood of the j^{th} cortical unit (which excludes the j^{th} unit). Let m index the cortical units within N_j .

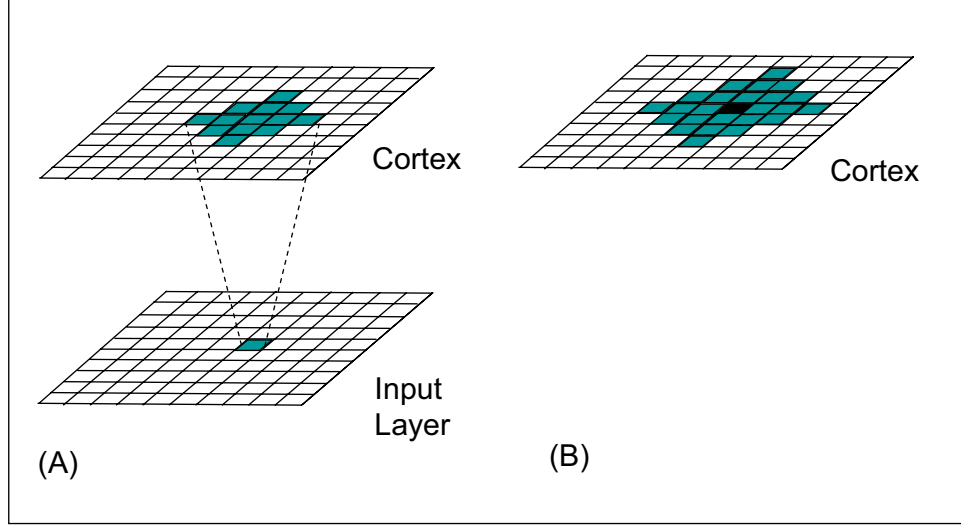


Figure 3. (A) Mapping from the input layer to the cortex. (B) Cortico-cortical connections and local neighborhoods formed by these connections

Thus, unit j is a local winner if

$$\forall m \in N_j, y_j > y_m \quad (3)$$

This is a local computation for a given cortical unit. Furthermore, the size of the cortex is larger than the size of the local neighborhood. This forms a major departure from previous work such as Saarinen and Kohonen,⁸ where every cortical unit was connected to every other cortical unit, and only one winner was selected for the entire network. The connectivity scheme we have chosen is more realistic, and allows multiple winners to exist in the simulated cortex.

Once the local winners are determined, their weights are updated to move them closer to the input vector. If cortical unit j is the winner, the update rule is

$$w_{ij}^k \leftarrow w_{ij}^k + \mu(X_i^k - w_{ij}^k) \quad (4)$$

for each input dimension k , and where i indexes those input units that are connected to the cortical unit j , and μ is the learning rate. μ is typically set to a small value, so that the weights are incrementally updated over a large set of input presentations.

In addition, the weights of the cortical units within the neighborhood N_j , denoted by the index m , are also updated to move closer to the same input, but with a weighting function $f(d(j, m))$, where $d(j, m)$ is the distance from the unit m to the local winner j . This is given by

$$w_{im}^k \leftarrow w_{im}^k + f(d(j, m))\mu [X_i^k - w_{im}^k] \quad (5)$$

Finally, the weights are normalized as follows. Let $\|N_j\|$ be the norm of all the incident weights at node j .

$$\|N_j\| = \sum_k \sum_i w_{ij}^k{}^2 \quad (6)$$

Then the i^{th} feedforward weight at the j^{th} cortical unit is calculated as

$$w_{ji}^k \leftarrow w_{ij}^k / \|N_j\| \quad (7)$$

Note that the normalization is carried out over all the input dimensions. One can normalize over individual input dimensions as well.

4. EXPERIMENTAL METHODS

We created two models to test, based on the two types of inputs, consisting of single color opponency or double color opponency.

4.1. Single opponent color model

We used an input layer consisting of 15x15 units. The input was created by choosing random RGB values in the range [0,1], to create input arrays of uniform color. We then apply the transformations of equations 1 to generate the appropriate inputs. The inputs were normalized to be in the range in $[-1, 1]$, and consisted of $D = 3$ dimensions.

A radius of $r_{IC} = 3$ was used to generate a topographic mapping from the input layer into the cortex. We modeled the cortex with an array consisting of 30x30 units. The intra-cortical connectivity was created with the parameter $r_{CC} = 5$.

For the weight updates, the function f was chosen to be a Gaussian that tapers to approximately zero at the boundary of the local neighborhood, ie at r_{CC} . The learning rules in section 3.5 were applied to learn the afferent weights.

The learning rate μ was set to 0.01. Learning was performed over 100,000 iterations.

4.2. Double color opponent model

We used an input layer consisting of 15x15 units. The input was created by choosing random RGB values in the range [0,1], to create input arrays of uniform color. We then apply the transformations of section 3.3 to generate the appropriate color channel signals. In this case the input is 10 dimensional, with two values from the luminance channels, four from the red-green opponent channels and four from the blue-yellow opponent channels, as shown in Figure 2. Thus, the cortical afferent input has 10 dimensions (two for luminance and eight for chrominance). The center and surround were implemented with Gaussian filters, such that the radius of the surround was approximately twice the radius of the center, and the area under each filter was normalized. The center was implemented with a 3x3 Gaussian, and the surround with a 7x7 Gaussian. The cortical inputs were normalized to be in the range [-1,1].

The subsequent stages of input to cortex mapping, intra-cortical connectivity and weight update calculations are identical to the single opponent color model, and employ the same set of parameters described in Section 4.1.

5. EXPERIMENTAL RESULTS

The final configuration of the network after training is visualized as follows. For each cortical unit, we determine the RGB input which causes the maximum response within this unit. The cortical unit is then colored with the same RGB value as its preferred input.

The result after applying the single color opponent model of Section 4.1 is shown in Figure 4. By comparing Figure 4 with Figure 1, we see that many of the desired features of spatial organization are present in the results. The color opponent red-green and blue-yellow regions are organized such that they are in opposition across a singularity. All color mixtures, such as cyan, orange, and purple are represented in the map. However, the black and white regions do not alternate. This may be due to artifacts of the self-organization process, including the parameters used for the learning and competitive processes.

The result of applying the double opponent color model in section 4.2 is shown in Figure 5. The organization in this model also agrees broadly with the ideal color map. There are singularities present in Figure 5 where red and green areas lie on opposite sides, and blue and yellow lie on opposite sides of a singularity. However, areas corresponding to black and white are not delineated as in Figure 4. This is due to the fact that the luminance channels do not contain any information for uniform color images. We used an alternate input training set consisting of color blobs of slowly varying luminance, and obtained results that are comparable to Figure 5.

Color mixtures are represented as well, though the number of hues is not as large as in Figure 4. The reason for this is the low-frequency nature of the inputs, which causes the double-opponent channels to carry highly correlated signals. Hence, the double-opponent system needs to be tested with inputs of different spatial frequency, rather than just the low

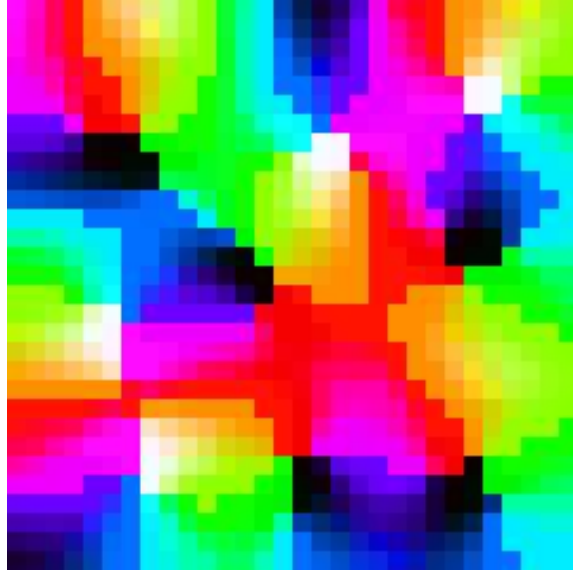


Figure 4. Result of applying the single color opponent model in Section 4.1.

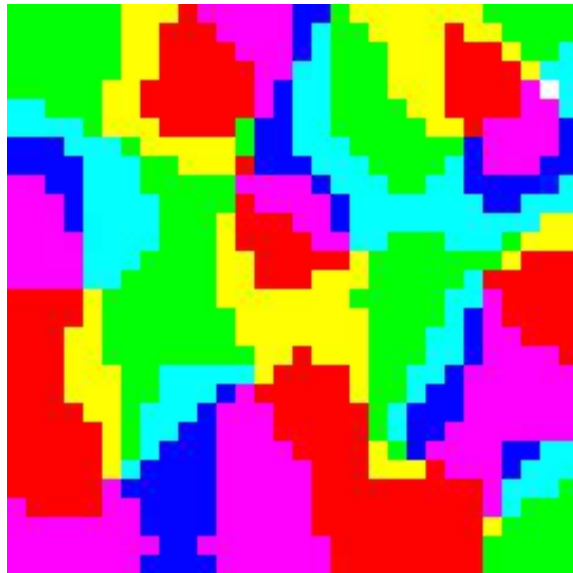


Figure 5. Result of applying a 10 channel double-opponent color model

frequency inputs studied in this paper. However, this may introduce other effects such as the sensitivity of units to the lines of orientation in addition to color.

We implemented a learning method using the local winner for the sake of computational convenience. There are other learning methods that have been used, such as the one used by Bednar,¹³ which incorporates a local network with excitatory and inhibitory connections. However, the methods used by Bednar and surveyed by Erwin¹ have been applied to monochrome input images of natural scenes, and primarily produce cortical maps depicting orientation selectivity.

5.1. Comparison with neuroscientific data

The computational results presented in this paper support the view of Landisman and Ts'o,⁶ as we are able to obtain a cortical representation for the mixture of colors. In particular, Landisman and Ts'o⁶[Fig. 14] show a schematic diagram indicating the presence of cells sensitive to orange, that lie between red and yellow blob centers. This is similar to the results presented in this paper.

Figure 4 from Xiao *et al*⁷ show a sequence of hue locations in cortical area V2. They show that to reach a blue cortical area from a red cortical area, one crosses orange, yellow, yellow-green, and cyan regions in that order. The result shown in Figure 4 of our paper shows an identical spatial ordering in the cortical map derived computationally.

We must caution that our model only considers a single stage of cortical processing. The neuroscientific literature points to a far more complicated picture that involves specific connections between areas V1 and V2 for the processing of orientation and color information.¹⁴

Dow¹² has presented a model for the 2D cortical organization of color. As pointed out in his paper, his model is not optimized to represent color opponency, where opponent colors are on opposite sides of a singularity. However, Dow has tried to address the concomitant representation of orientation and color in the same map, which is a problem we have not addressed in this paper.

6. CONCLUSIONS

Though color is an important visual cue, the mechanisms underlying color processing in the visual cortex are still unclear. This paper addresses this problem by proposing a specific spatially organized cortical map that takes color opponency into account. Further, we developed a computational model that drew inspiration from biology, and whose operation creates cortical color maps that exhibit many of the desirable characteristics of such maps. We are able to achieve the representation of color mixtures that are organized in an opponent fashion around singularities. Our results appear to agree with recent neurophysiological findings.

Our work can be extended by using more precise models for the interaction between color signals earlier in the pathway, including interaction between cones and LGN cells in the retino-geniculate pathway, and interaction within the LGN. We also need to model interactions between the different cortical areas that process color, such as V1, V2 and V4.

Further experimental development is required on the computational front as well, to incorporate realistic excitatory/inhibitory lateral interactions with a minimal set of assumptions that are biologically plausible. We are continuing work in this direction.

REFERENCES

1. E. Erwin, K. Obermayer, and K. J. Schulten, "Models of orientation and ocular dominance columns in the visual cortex: A critical comparison.," *Neural Computation* **7**(3), pp. 425–468, 1995.
2. V. B. Mountcastle, "An organizing principle for cerebral function: The unit module and the distributed system," in *The Mindful Brain*, G. M. Edelman and V. Mountcastle, eds., MIT Press, Cambridge, MA, 1978.
3. B. Wandell, *Foundations of vision*, Sinauer Associates, Sunderland, USA, 1995.
4. R. Boynton, *Human Color Vision*, Holt, Rinehart and Winston, New York, 1979.
5. E. R. Kandel, J. H. Schwarz, and T. M. Jessell, *Principles of Neural Science*, McGraw Hill, New York, 2000.
6. C. E. Landisman and D. Y. Ts'o, "Color processing in macaque striate cortex: electrophysiological properties," *J. Neurophysiology* **87**, pp. 3138–3151, 2002.
7. Y. Xiao, Y. Wang, and D. J. Felleman, "A spatially organized representation of colour in macaque cortical area v2," *Nature* **421**, pp. 535–539, January 2003.

8. J. Saarinen and T. Kohonen, "Self-organized formation of colour maps in a model cortex," *Perception* **14**, pp. 711–719, 1985.
9. T. Kohonen, "Self-organization and associative memory," 1985.
10. H. G. Barrow, A. J. Bray, and J. M. L. Budd, "A self-organizing model of 'color blob' formation," *Neural Computation* **8**, pp. 1427–1448, 1996.
11. E. Doi, T. Inui, T. Lee, T. Wachtler, and T. Sejnowski, "Spatiochromatic receptive field properties derived from information-theoretic analyses of cone mosaic responses to natural scenes," *Neural Computation* , pp. 397–417, 2003.
12. B. M. Dow, "Orientation and color columns in monkey visual cortex," *Cerebral Cortex* **12**, pp. 1005–1015, 2002.
13. J. A. Bednar, *Learning to See: Genetic and Environmental Influences on Visual Development*. PhD thesis, Department of Computer Sciences, The University of Texas at Austin, 2002. Technical Report AI-TR-02-294.
14. A. W. Roe and D. Y. Ts'o, "Specificity of color connectivity between primate v1 and v2," *J. Neurophysiology* **82**(5), pp. 2719–2730, 1999.