UNIVERSITÀ DEGLI STUDI DI GENOVA
DIBRIS

A cortical architecture of V1-MT pathway: technical report

Fulvio Missoni
April 26, 2021

This architecture is based on the model for the emergence of the motion-in-depth (MID) selectivity proposed by Sabatini & Solari (2004) [2]. In their work, on the basis of the phase-based techniques, they implement an algorithm to construct a neural-like detector for MID. In particular they proposed that the binocular integration play a key role in the emergence of this perception.

1 General architecture

If we consider a stereo visual system in vergent geometry, each cameras see the visual target from different point of view. The two cameras' views are related, formally: $I[x(t), y(t), t] = I^R[x(t)+\delta_H, y(t)+\delta_V, t]$ where $\delta_H[x(t), t]$ and $\delta_V[y(t), t]$ are respectively, the horizontal and vertical components of the disparity vector $\boldsymbol{\delta}[x(t), y(t), t]$. The first stage of processing consists on a population of not separable spatio-temporal filters (or receptive fields) that models the distributed representation of the stereo image signal on several channels of description: orientation, phase-shift and velocity. In general, computing the convolution of the pair images with a not separable spatio-temporal filter we have:

$$Q^{L/R}(x, y, t; \theta, \psi, v_c) = I^{L/R} * G(x, y; \theta, \psi) f(t; v_c) cos[2\pi(k_0 x_\theta - f_0 t)] \tag{1.1}$$

$$= C^{L/R}(x, y, t; \theta, \psi, v_c) + jS^{L/R}(x, y, t; \theta, \psi, v_c) \tag{1.2}$$

where G and f are respectively the spatial and temporal component of the receptive field and C and S are the real and the imaginary part of the convolution. It is worth to note that the temporal (f_0) and the spatial (k_0) frequencies are linked by the relation $v_c = f_0 k_0$ where v_c is the preferred component velocity. It can be shown that is possible to express a not separable spatio-temporal filter as the combination of separable spatio-temporal filters contributions (**see dedicated section**). These monocular terms can be expressed as follows:

$$S_{01}^{L/R} = C_t^{L/R} + S^{L/R}; \qquad S_{02}^{L/R} = C^{L/R} - S_t^{L/R}; \tag{1.3}$$

$$S_{03}^{L/R} = C_t^{L/R} - S^{L/R}; \qquad S_{04}^{L/R} = C^{L/R} + S_t^{L/R}; \tag{1.4}$$

The spatial component is modelled by analytic function represented by 2D Gabor function:

$$G(x, y; \theta, \psi) = A \cdot e^{\left(-\frac{x_\theta^2 + y_\theta^2}{2\sigma^2}\right)} e^{j2\pi(k_0 x_\theta + \psi)} \tag{1.5}$$

where:

$$x_\theta = x\,cos(\theta - 90°) + y\,sin(\theta - 90°) \tag{1.6}$$

$$y_\theta = -x\,sin(\theta - 90°) + y\,cos(\theta - 90°) \tag{1.7}$$

where k_0 is the radial spatial peak frequency, σ determines the spatial dispersion, ψ is the phase parameter (related to filter symmetry), A is a modulation constant and θ is the orientation angle of the filter respect to the x axis (**see section dedicated for more details on the filter design**). The temporal component, on the other hand, can be modelled through three different models:

1. Exponential decay model

$$f(t; v_c, \tau_c) = B \cdot e^{-t/\tau_c} cos(2\pi f_0 t) \tag{1.8}$$

   where f_0 is the temporal frequency of the filter chosen based on the preferred component velocity value v_c, B is a normalization constant and τ_c is the time constant and determines the filter extension in temporal domain.

2. Gabor model

$$f(t; v_c, \sigma) = B \cdot e^{-t^2/(2\sigma^2)} cos(2\pi f_0 t) \tag{1.9}$$

   where the filter extension is set by σ value instead of τ_c

3. Adelson & Bergen model

$$f(t; k) = B \cdot (kt)^n e^{-kt} \left(\frac{1}{n!} - \frac{(kt)^2}{(n+2)!}\right) \tag{1.10}$$

   where the parameter n assumes the values of 3 and 5, by yielding even and odd temporal profiles, respectively (Adelson & Bergen, 1985 [1]). The parameter k affects both the peak frequency and the bandwidth of the filter, and its values (**see Section dedicated**) can be chosen to obtain a preferred component velocities.

The algorithm proposed in (Sabatini & Solari, 2004 [2]) computes the temporal derivative of disparity as a combination of the real and imaginary part of the convolution (C, S) and its temporal derivatives (C_t, S_t). For the calculation of these last two components it is necessary to make a clarification because the numeric differentiation is an operation extremely sensible to noise. It is possible to demonstrate than asking that the temporal filter f and its temporal derivative f' be in quadrature is a sufficient condition (**see dedicated section on temporal filter design**) Thus, we obtain:

$$C_t^{L/R}(x,y,t;\theta,\psi,v_c) = f' * C_t^{L/R}(x,y,t;\theta,\psi) \tag{1.11}$$

$$S_t^{L/R}(x,y,t;\theta,\psi,v_c) = f' * S_t^{L/R}(x,y,t;\theta,\psi) \tag{1.12}$$

Binocular integration: simple cells

To obtain the oriented (in the spatio-temporal domain) monocular receptive fields the approach proposed in (Adelson & Bergen, 1985 [1]) is followed. The binocular convergence of these signals permits to obtain the simple cells response selective to the component motion direction (orthogonal to its spatial orientation), four with left ocular dominance:

$$S_1 = (1-\alpha)S_{01}^L - \alpha S_{02}^R; \tag{1.13}$$

$$S_2 = (1-\alpha)S_{02}^L + \alpha S_{01}^R; \tag{1.14}$$

$$S_3 = (1-\alpha)S_{03}^L - \alpha S_{04}^R; \tag{1.15}$$

$$S_4 = (1-\alpha)S_{04}^L + \alpha S_{03}^R; \tag{1.16}$$

and four with right ocular dominance:

$$S_5 = \alpha S_{01}^L - (1-\alpha)S_{02}^R; \tag{1.17}$$

$$S_6 = \alpha S_{02}^L + (1-\alpha)S_{01}^R; \tag{1.18}$$

$$S_7 = \alpha S_{03}^L - (1-\alpha)S_{04}^R; \tag{1.19}$$

$$S_8 = \alpha S_{04}^L + (1-\alpha)S_{03}^R; \tag{1.20}$$

Each cell is characterized by its spatial orientation θ in the x-y domain (associated to the preferred spatial orientation of the stimulus) and its orientation in the x_θ-t domain (that identifies its preferred component velocity v_c).

Energy model: complex cells (I layer)

The subpopulation responses of the complex cells (first-layer) is obtained by the energy model (Adelson & Bergen, 1985 [1]), therefore the sum of the squared outputs of a pair of simple cells in quadrature:

$$C_{11} = S_1^2 + S_2^2; \tag{1.21}$$

$$C_{12} = S_3^2 + S_4^2; \tag{1.22}$$

$$C_{13} = S_5^2 + S_6^2; \tag{1.23}$$

$$C_{14} = S_7^2 + S_8^2; \tag{1.24}$$

The invariance of the response with respect to the contrast of the visual signal is obtained through a normalization stage (**for more detail see normalization section**):

$$C_{1,i} = \frac{C_{1,i}}{\frac{a_1}{4}\sum_{k=1}^{4} C_{1,k} + a_2}; \quad i = 1,2,3,4 \tag{1.25}$$

where, a_1 controls the normalization with respect to the population of complex cells, and a_2 is a constant value.

Opponent energy unit: complex cells (II layer)

Finally, combining two complex cells with same ocular dominance the responses of C_{21} and C_{22} cells are obtained:

$$C_{21} = w_2 C_{12} - w_1 C_{11} \tag{1.26}$$

$$C_{22} = w_3 C_{13} - w_4 C_{14} \tag{1.27}$$

where, $w_i \in [0,1]$; $i = 1,2,3,4$. By construction, between the two receptive fields there is a constant displacement value equal to $\pi/2$. The parameters of the responses of the monocular filters control the tuning selectivity of these cells at a certain oriented disparity vector $\delta^{\theta}_{pref} = \Delta\psi_\theta / k_0$ and a certain motion velocity along the direction orthogonal to its receptive field orientation (θ). These cells result markedly selective to the binocular disparity and, in this way, are capable to provide an affordable estimate of disparity in $[-\Delta, \Delta]$ range, where $\Delta = \Delta\psi^{max}_{\theta} / k_0$ can be defined as the maximum disparity value detectable by the population.

MID detectors

Combining the opponent energy unit responses for motion as Sabatini-Solari model a sub-population cells selective for the motion-in-depth (or *stereomotion*):

$$C_3(x, y, t; \theta, \psi, v_c) = C_{21}(x, y, t; \theta, \psi, v_c) + C_{22}(x, y, t; \theta, \psi, v_c) \tag{1.28}$$

this selectivity ability is paid for with a total absence of sensitivity to static disparity. The cell's response is maximum for stimulus in motion in depth (approaching or moving away) that generate on both retinas a velocity vector with the same gain v_c and orientation θ (in x-y plane), but opposite direction. The preferred velocity value is determined by the temporal frequency of spatio-temporal filter. Varying the parameters v_c and θ is possible to build a MID detectors population selective for several component velocity vectors, where v_c identifies the gain vector and θ the orientation. Pooling the signals provided by the C3 cells spatially and along the orientation channels we obtain an MT *pattern motion* cell. The spatial pooling is realized through a local weighted average, assuming gaussian weights. The orientation pooling, on the other hand, is obtained by means of positive and negative weights, that model the excitatory and inhibitory effects from afferent connection from V1 to MT neuron. Formally the MT *pattern motion* neuron's response sensitive to speed v_c in d direction can be expressed as:

$$MT(x, y, t; d, \psi, v_c) = \sum_{\theta_i=\theta_1}^{\theta_N} w_d(\theta_i) \sum_{(x',y')} G(x - x', y - y') C_3(x', y', t'; \theta_i, \psi, v_c) \tag{1.29}$$

where $G(x,y)$ is 2D gaussian function centered in (x',y') point and determines the weight values, and $w_d(\theta) = cos(\theta - d)$ represents the linear weights that gives rise to the MT cells *tuning*. The Q values of direction of the v_c vector (d) span from 0 to 2π.

2 Technical notes on filter design

2.1 Separability of Gabor filters

Not separable spatio-temporal filters

As we have seen a spatio-temporal filter is described by the equation 1.2 that rewritten here:

$$H(x,y,t;k_0,f_0,\theta,\psi,v_c) = G(x,y;\theta,\psi)f(t;v_c)cos\left[2\pi(k_0x_\theta - f_0t)\right] \tag{2.1}$$

where G and f are respectively the spatial and temporal component (described by the eqq.(1.5)-(1.10)) of the receptive field, f_0 and k_0 are respectively the temporal and the spatial frequencies (linked by the relation $v_c = f_0k_0$, where v_c is the preferred component velocity), θ is the orientation of the spatial filter in x-y domain and ψ is the phase value of the spatial filter. This function satisfies the separability property and, therefore, can be written as a combination of temporal and spatial components. Also the 2D Gabor function satisfies this property therefore, a more efficient computational approach to obtain the distributed representation for several orientation and phase-shift channels can be used (**see sections dedicated**). From the trigonometric formulas we have:

$$cos(\alpha - \beta) = cos(\alpha)cos(\beta) + sin(\alpha)sin(\beta) \tag{2.2}$$

therefore, from 2.1:

$$H(\cdot) = G(\cdot)f(\cdot)\left[cos(\omega_0x_\theta)cos(\omega_{0t}) + sin(\omega_0x_\theta)sin(\omega_{0t})\right] \tag{2.3}$$

where $\omega_0 = 2\pi k_0$ is the spatial pulsation and $\omega_{0t} = 2\pi f_0$ is the temporal pulsation. One of possible combination is:

$$H(\cdot) = G(\cdot)cos(\omega_0x_\theta)f(\cdot)cos(\omega_{0t}) + G(\cdot)sin(\omega_0x_\theta)f(\cdot)sin(\omega_{0t}) \tag{2.4}$$

Formally, with this formula we obtain a pure not separable spatio-temporal filter. A modelling choice to generalize it, it is to add a parameter $\eta \in [0,1]$ to unbalance the terms. Thus, the general expression of not-separable spatio-temporal filter is obtained:

$$H(\cdot) = G(\cdot)cos(\omega_0x_\theta)f(\cdot)cos(\omega_{0t}) + \eta G(\cdot)sin(\omega_0x_\theta)f(\cdot)sin(\omega_{0t}) \tag{2.5}$$

If $\eta = 0$ a pure even (separable) spatio-temporal RF is obtained. If $\eta = 1$ a pure oriented (not-separable) spatio-temporal RF is obtained.

Note on temporal component

Two questions still need to be discussed about temporal filter design: 1) how can equations (1.12) and (2.5) be related? 2) how to set the temporal frequency to avoid aliasing? We start from the first question. We express the temporal derivative of the eq. (1.8):

$$f'(t) = -\frac{1}{\tau} e^{(-t/\tau)} cos(\omega_{0t} t) - e^{(-t/\tau)} sin(\omega_{0t} t) \tag{2.6}$$

Assuming that the first term is negligible we can consider that the temporal derivatives is obtained with the quadrature filter of f. This approximation is valid when $\tau << 1$. Another conditions on ω_{0t} and t (with fixed tau) can be obtained considering the periodicity of the two trigonometric functions.

2D Gabor filters

The use of 2D Gabor filters is essential to compute bidimensional features like local orientation, disparity and optic flow. To this aim it is necessary to create a 2D set of Gabor filters with the same peak frequency k_0, to orient them to cover the whole space in the frequency domain. Moreover, if the binocular aspect is considered also several values of phase-shift between left and right side must be considered. At the computational level, to consider such a large number of filters is very expensive. The separability property is a way to reduce this cost. A 2D Gabor filter is expressed by the eqq. (1.5) and (1.7). Combining the two equation and set $\psi = 0$ we obtain the even symmetry Gabor filter:

$$G_e(x, y; k_0, \theta) = g_e \; cos\left[\omega_0 x cos(\theta - 90°) + \omega_0 y sin(\theta - 90°)\right] \tag{2.7}$$

where

$$g_e(x_\theta, y_\theta; \sigma) = A \cdot \frac{1}{2\pi \, \sigma_x \, \sigma_y} exp\left[-\frac{1}{2}\frac{x_\theta^2 + y_\theta^2}{\sigma^2}\right] \tag{2.8}$$

Combining the equations (2.2) and (2.7) we obtain:

$$\begin{aligned} G_e &= gcos(\omega_0 x \; cos\xi) cos(\omega_0 y \; sin\xi) + gsin(\omega_0 x \; cos\xi) sin(\omega_0 y \; sin\xi) && (2.9)\\ &= gcos(\omega_0 x \; sin\theta) cos(\omega_0 y \; cos\theta) - gsin(\omega_0 x \; sin\theta) sin(\omega_0 y \; cos\theta) && (2.10)\end{aligned}$$

Considering the vertical g_x and the horizontal g_y components of the filter g we have:

$$\begin{aligned} g_x &= \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{1}{2\sigma^2}(xcos\xi - ysin\xi)^2\right] && (2.11)\\ &= \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{1}{2\sigma^2}(xsin\theta + ysin\theta)^2\right] && (2.12)\end{aligned}$$

$$\begin{aligned} g_y &= \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{1}{2\sigma^2}(xsin\xi + ycos\xi)^2\right] && (2.13)\\ &= \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{1}{2\sigma^2}(-xcos\theta + ycos\theta)^2\right] && (2.14)\end{aligned}$$

Defining:

$$cs = g_x cos(\omega_0 x sin\theta) \tag{2.15}$$

$$cc = g_y cos(\omega_0 y cos\theta) \tag{2.16}$$

$$ss = g_x sin(\omega_0 x sin\theta) \tag{2.17}$$

$$sc = g_y sin(\omega_0 x cos\theta) \tag{2.18}$$

We can rewrite equation (2.10) as:

$$G_e = cs * cc - ss * sc \tag{2.19}$$

The odd part can be obtained in a similar way:

$$G_o = ss * cc + cs * sc \tag{2.20}$$

2.2 Normalization model

From my notes

2.3 Aperture problem

From my notes (cite articles!!)

Contents

References

[1] Edward H Adelson and James R Bergen. Spatiotemporal energy models for the perception of motion. *Josa a*, 2(2):284–299, 1985.

[2] Silvio P Sabatini and Fabio Solari. Emergence of motion-in-depth selectivity in the visual cortex through linear combination of binocular energy complex cells with different ocular dominance. *Neurocomputing*, 58:865–872, 2004.