Extracting edges in space and time during visual fixations

Lynn Schmittwilken and Marianne Maertens
Science of Intelligence, Technische Universität Berlin

Assuming that vision is stable during fixations, existing edge models typically employ orientation-sensitive spatial mechanisms to mimic human edge processing. However, recent studies suggest that small eye jitters that occur during fixational pauses are functionally relevant for human vision [1]. To test this notion, we therefore developed a spatial edge model with standard components of early vision models (spatial filtering, non-linear normalization, integration), extended it by a temporal domain and fed it with a time-varying input as if sampled by fixational eye movements (FEMs) [2]. The model successfully accounts for human performance in multiple edge tasks and notably does so without relying on orientation-sensitive mechanisms.

The model structure is shown in Figure 1. We simulate the effect of FEMs by applying ocular drift to the retinal input (Fig. 1B) resulting in a time series of slightly shifted input images. Drift is simulated as Brownian motion over $T = 0.2s$ with a diffusion coefficient of $D = 20\frac{arcmin^2}{s}$ and a temporal frequency of $f = 100Hz$. The dynamic input is then filtered in space and time (Fig. 1C). In space, we applied five spatial DoG filters $G_i(f_x, f_y)$ with peak spatial frequencies (SFs) between 0.62 and 9.56 cpd in octave intervals defined as

$$G_i(f_x, f_y) = e^{-2\pi^2 s_i^2 (f_x^2 + f_y^2)} - e^{-8\pi^2 s_i^2 (f_x^2 + f_y^2)}, \quad (1)$$

where f_x and f_y denote the SFs in cpd and $s_{1-5} = [0.016, 0.032, 0.064, 0.128, 0.256]$ deg controls the spatial scale of the DoG filters.

In time, we used a bandpass filter $H(\omega)$ which peaks at 9.52 Hz and no sensitivity to static inputs defined as

$$H(\omega) = m_1\, e^{-(\frac{|\omega|}{m_2})^2} \;/\; (1 + (\frac{m_3}{|\omega|})^{m_4}), \quad (2)$$

where ω denotes the temporal frequencies in Hz, $m_1 = 1$, $m_2 = 22.9$, $m_3 = 8.1$, $m_4 = 0.8$, and $H(\omega = 0) = 0$.

After filtering the dynamic input in space and time, we first integrate the filtered outputs across time by computing the squared mean separately at each spatial scale i (Fig. 1D). Then, we normalize the integrated signals by their mean activation M_i (Fig. 1E). Finally, we sum the normalized signals over all scales i, creating the final 2d model output (Fig. 1F). We quantify model performance by correlating the model output with a ground truth edge template (Fig. 1G).

We tested the model on contour detection in natural scenes (realistic task) and on edge sensitivity in narrow-band noise (controlled task). The model captured human performance reasonably well [2]. Our results show that when considering the spatial and temporal properties of the early visual system, FEMs facilitate human edge processing without relying on orientation-sensitive processes[1].

[1] M. Rucci and J. D. Victor, "The unsteady eye: an information-processing stage, not a bug," *Trends in Neurosciences*, vol. 38, no. 4, pp. 195–206, 2015.

[2] L. Schmittwilken and M. Maertens, "Fixational eye movements enable robust edge detection," *Journal of Vision*, vol. 22, no. 8, pp. 1–12, 2022.

[1] Link to all model code: https://github.com/computational-psychology/schmittwilken2022_active-edge-model

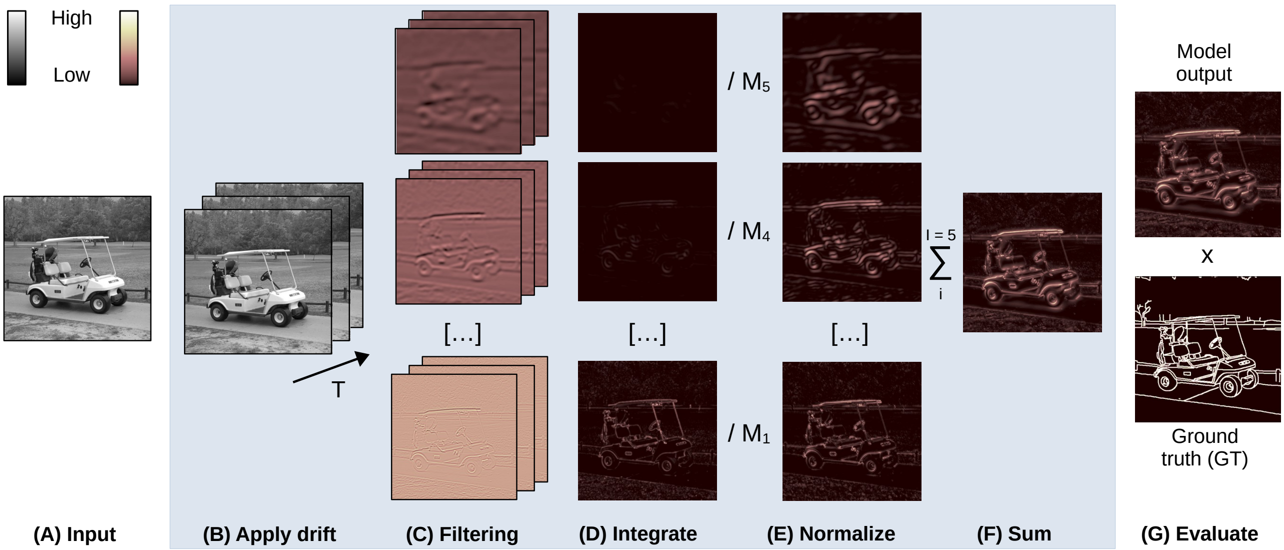

Figure 1: Model structure from [2]