# Supplementary material

## Additional analysis of the complex cell directional responses

Figure 1 shows the direction selectivity of 36 of the 144 trained complex cells, elaborating the results given in Figs. 5 and 6. Most complex cells are sensitive to opposite motion directions (dual lobe). Only 2 out of the 36 cells presented have a unilateral response in direction space (4th row, 3rd column and 6th row, 5th column). The sizes and shapes of the lobes vary slightly. The population covers a wide range of different orientations/directions, enabling it to represent a wide variety of visual inputs.

## STDP windows

We have experimented with the following additional STDP windows for the complex cell learning rules:

Step left window:

$$\Delta w_i^{\mathrm{LTD}} = \begin{cases} \eta_{\mathrm{LTD}} : |t_{s-1} - t_i| \le \tau_{\mathrm{LTD}} \\ 0 : |t_{s-1} - t_i| > \tau_{\mathrm{LTD}} \,. \end{cases} \tag{1}$$

Step right window:

$$\Delta w_i^{\mathrm{LTP}} = \begin{cases} \eta_{\mathrm{LTP}} : |t_i - t_s| \le \tau_{\mathrm{LTP}} \\ 0 : |t_i - t_s| > \tau_{\mathrm{LTP}} \,. \end{cases} \tag{2}$$

Step symmetrical window:

$$\Delta w_i^{\mathrm{LTP}} = \begin{cases} \eta_{\mathrm{LTP}} : |t_i - t_s| \le \tau_{\mathrm{LTP}} \\ 0 : |t_i - t_s| > \tau_{\mathrm{LTP}} \,. \end{cases}$$
$$\Delta w_i^{\mathrm{LTD}} = \begin{cases} \eta_{\mathrm{LTD}} : |t_{s-1} - t_i| \le \tau_{\mathrm{LTD}} \\ 0 : |t_{s-1} - t_i| > \tau_{\mathrm{LTD}} \,. \end{cases} \tag{3}$$

Linear window:

$$\Delta w_i^{\mathrm{LTP}} = \begin{cases} \eta_{\mathrm{LTP}}(1 - (t_i - t_s)) : |t_i - t_s| \le \tau_{\mathrm{LTP}} \\ 0 : |t_i - t_s| > \tau_{\mathrm{LTP}} \,. \end{cases}$$
$$\Delta w_i^{\mathrm{LTD}} = \begin{cases} \eta_{\mathrm{LTD}}(1 - (t_{s-1} - t_i)) : |t_{s-1} - t_i| \le \tau_{\mathrm{LTD}} \\ 0 : |t_{s-1} - t_i| > \tau_{\mathrm{LTD}} \,. \end{cases} \tag{4}$$

Exponential window (similar to rule for simple cells):

$$\Delta w_i^{\mathrm{LTP}} = \eta_{\mathrm{LTP}} \, e^{\frac{t_i - t_s}{\tau_{\mathrm{LTP}}}}$$
$$\Delta w_i^{\mathrm{LTD}} = -\eta_{\mathrm{LTD}} \, e^{\frac{t_{s-1} - t_i}{\tau_{\mathrm{LTD}}}} \,. \tag{5}$$

These windows lead to different degrees of orientation selectivity as shown in Fig. 2. The simple step window used in the main text gives the highest amount of orientation selectivity.

## Stereo driving scenes

Continuing the work on stereo driving scenes, we give more detail on the network learned using the MVSEC dataset here. We focus on the sequence "outdoorday1data". This sequence is composed of a stereo event-based recording of a driving scene in a residential street. We learned on 9 regions corresponding to a tiling of the visual scene, similar to Fig. 7b with less but bigger regions (40 by 40 pixels).

Using information provided in the dataset, we can estimate the network ability to learn disparities, and to which extent. We have detailed information on the setup, including the baseline length $B$ between the two event-based cameras (10 cm), the intrisic focal length $f$ in pixels (given by the calibration file, 223 pixels) and the depth of the scene $D$. According to this equation:

$$D = \frac{Bf}{d} \tag{6}$$

we can relate the disparity $d$ in pixels of the learned simple cells receptive fields to the depth of the scene. Considering a 10 by 10 pixel receptive field, we estimated the disparity ranging from -5 to 5 pixels, depending on the relative shift between the left and right subfield. We use the absolute values of the disparity measures as we are only interested in the depth computation. We can therefore detect depth ranging from $\frac{0.1 \times 223}{5} = 4.46$ meters up to theoretically infinite depth.

In order to have a comparison point, we used the ground truth depth maps present along the event-based sequence of the dataset. We present those values for the 6 first regions in Fig. 3 in blue. We observe a wide range of depth, going from a few meters up to a 100 m and more. We removed
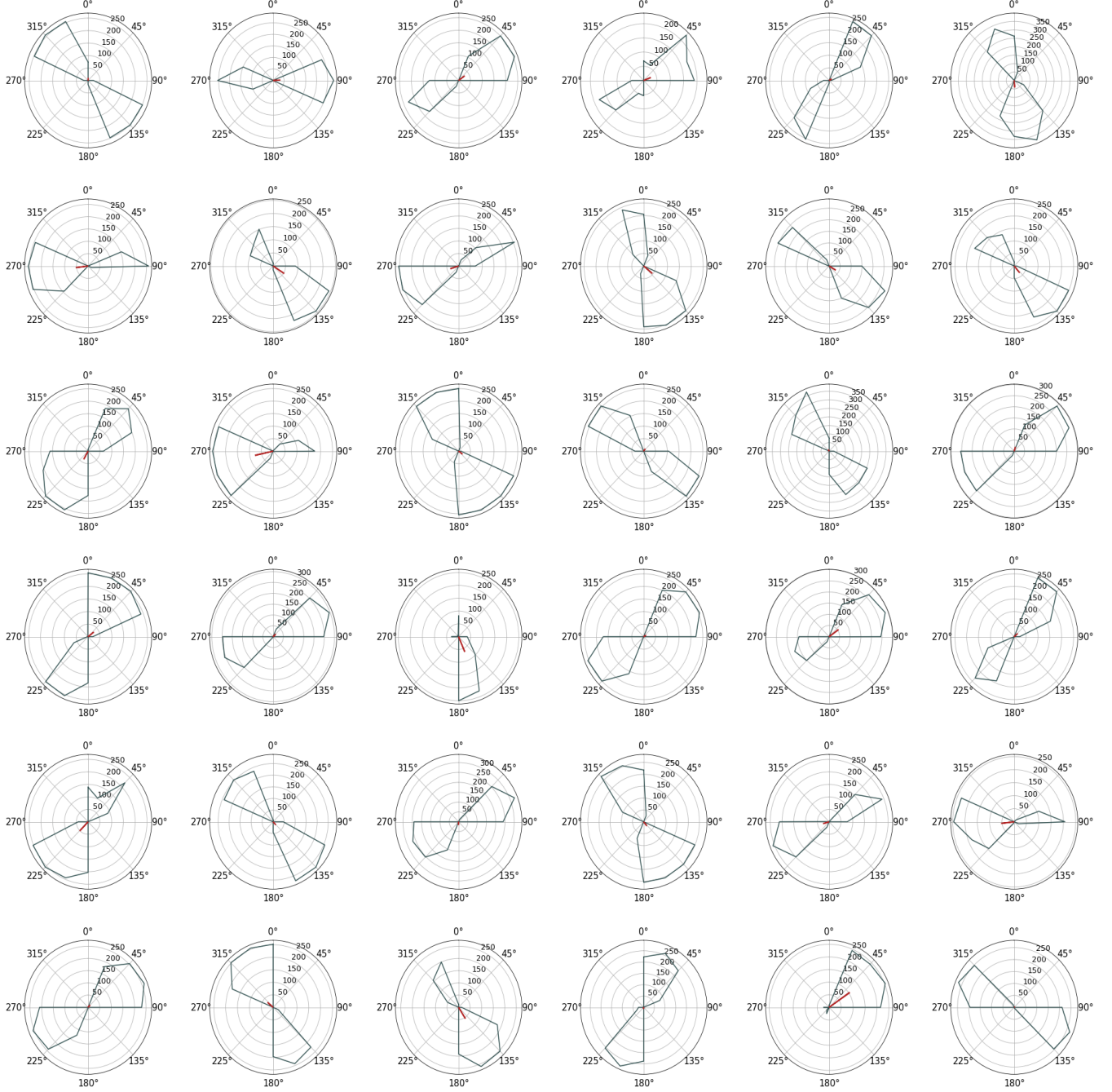
Figure 1: 36 first (out of 144) complex cell response in direction space. Results are obtained by measuring spike activity for gratings of multiples orientations.

the 3 bottom regions which are much more limited in terms of depth since they are directed towards the road and the car (ranging from 2 up to 5 meters). The central region (2<sup>nd</sup> row 2<sup>nd</sup> column in the figure) present the biggest depths, since it is oriented towards the center of the scene.

In grey, we plotted the computed depth histogram us-

ing the learned receptive fields. Our method for computing disparity is precise up to a minimum of 1 pixel. In order to compare to the ground truth provided in the dataset, we used bins of unequal length using equation 6. We limited the last bin to 70 meters since it can theoretically represent any depth from 22.33 meters up to infinity. Comparing the
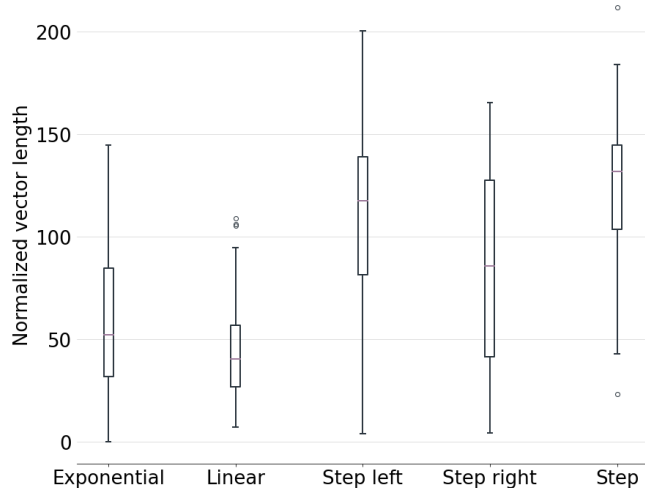
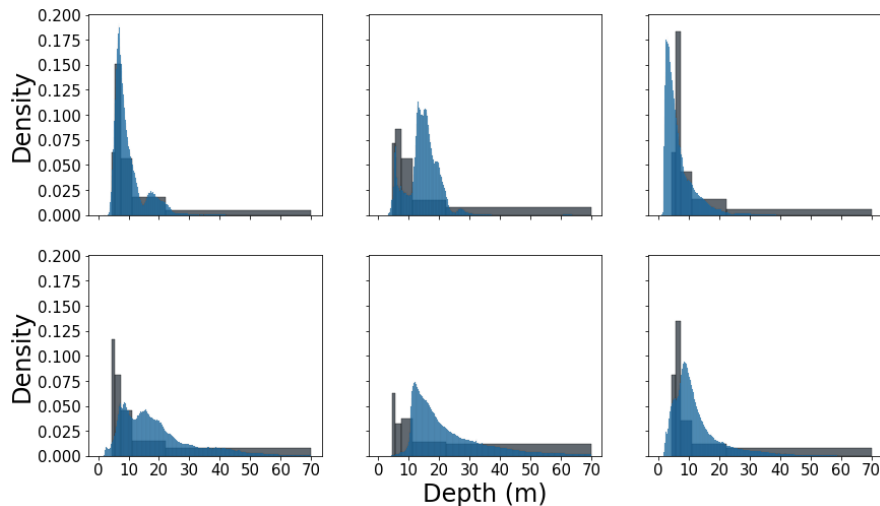Figure 2: Orientation selectivity for 5 different complex cell STDP windows.



Figure 3: Depth estimation histogram from the MVSEC dataset sequence for the first 6 visual regions (blue). Histogram of estimated depths of the learned receptive fields for the first 6 visual regions (grey).

two figures, we observe a reasonable correlation between the 2 distributions. The set of simple cell receptive fields form a coherent basis spanning a wide range of disparity tuning.

Figure 4 shows the simple cells receptive fields learned on the driving sequence for the 9 different regions. We observe a coherent tuning to orientation considering the learning scene, i.e., a majority of oblique orientations in the corners, and vertical and horizontal orientations for the sides and center.
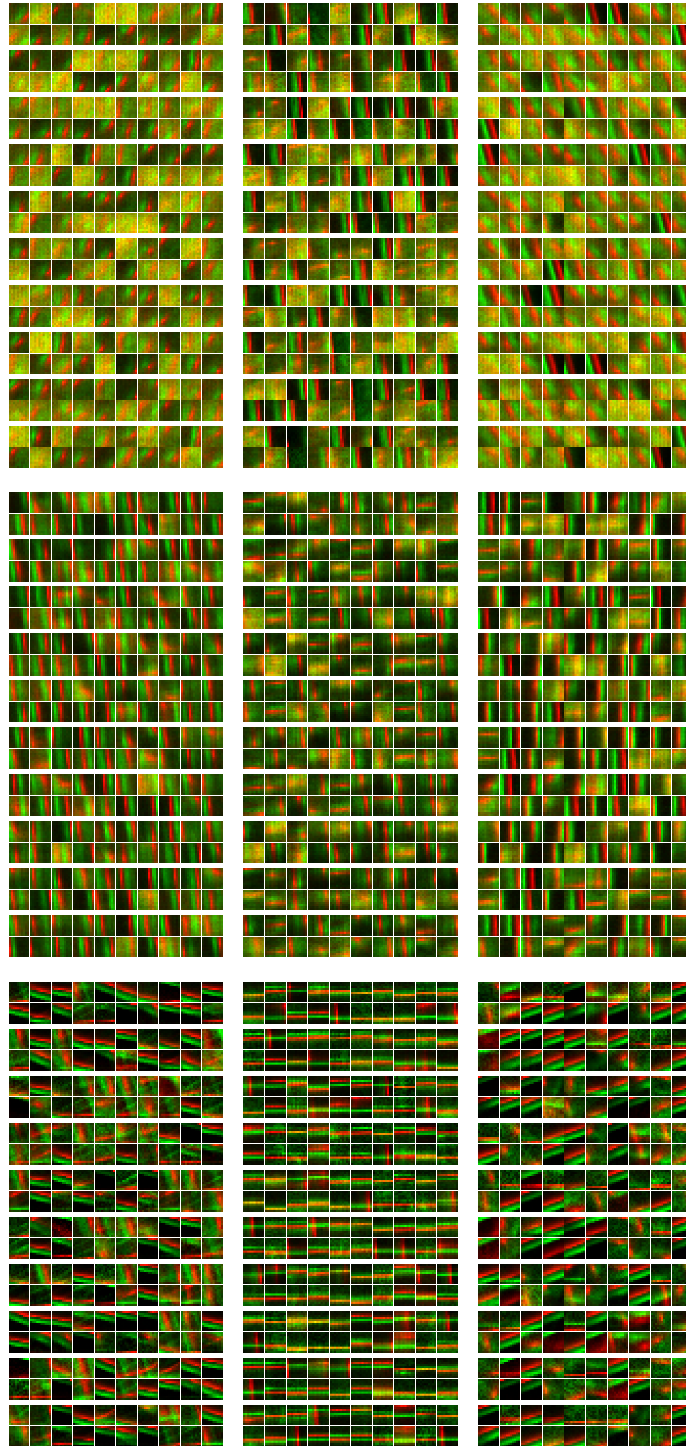
3

Figure 4: Simple cell receptive fields learned on nine different image regions of a driving sequence from the MVSEC dataset. We present the left and right subfield (from the left and right event-based camera) on top of each other.