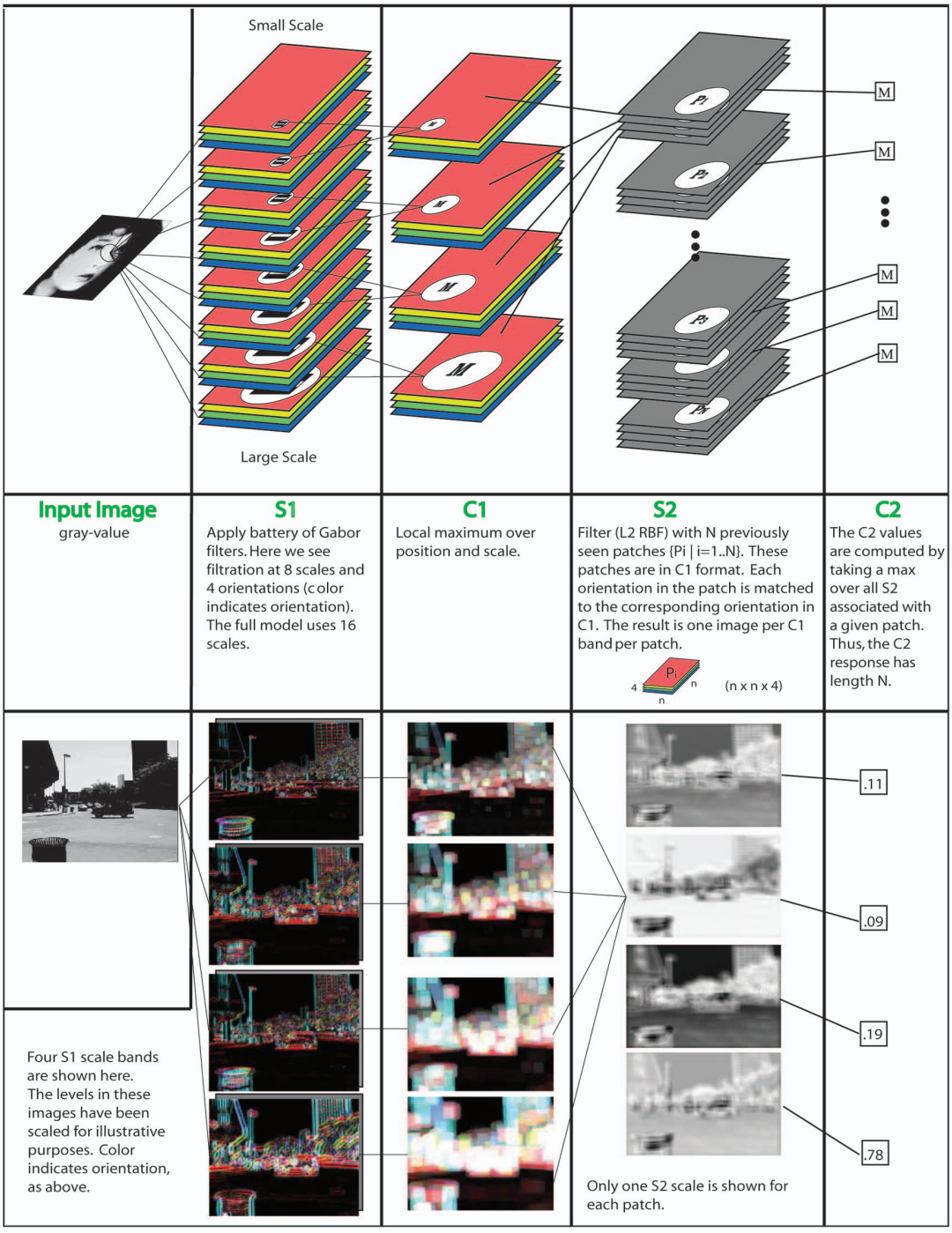

Fig. 1. System overview: A gray-value input image is first analyzed by an array of S_1 units at four different orientations and 16 scales. At the next C_1 layer, the image is subsampled through a local MAX (M) pooling operation over a neighborhood of S_1 units in both space and scale, but with the same preferred orientation. In the next stage, S_2 units are essentially RBF units, each having a different preferred stimulus. Note that S_2 units are tiled across all positions and scales. A MAX pooling operation is performed over S_2 units with the same selectivity to yield the C_2 unit responses.

MAX operation. That is, the response r of a complex unit corresponds to the response of the strongest of its m afferents $(x_1, \ldots, x_m)$ from the previous S_1 layer such that:

$$r = \max_{j=1\ldots m} x_j. \tag{3}$$

Consider, for instance, the first band: $\mathcal{S} = 1$. For each orientation, it contains two S_1 maps: The one obtained using a filter of size 7×7 and the one obtained using a filter of size 9×9 (see Table 1). The maps have the same dimensionality but they are the outputs of different filters. The C_1 unit responses are computed by subsampling these maps using a cell grid of size $N_{\mathcal{S}} \times N_{\mathcal{S}} = 8 \times 8$. From each grid cell, one single measurement is obtained by taking the maximum of all 64 elements. As a last stage, we take a max over the two scales from within the same spatial neighborhood, by recording