TECHNICAL NOTE

APPLICATIONS OF THE EVE SOFTWARE FOR VISUAL MODELING

MICHAEL S. LANDY, LEV Z. MANOVICH and GEORGE D. STETTEN

Department of Psychology and Center for Neural Science, New York University, 6 Washington Place,
New York, NY 10003, U.S.A.

(Received 3 March 1989; in revised form 2 June 1989)

Abstract—EVE, the *Early Vision Emulation* software, is a system for the simulation of early visual processing. EVE has the ability to carry out the operations of a wide variety of models of spatial vision, motion detection and processing, and spatial sampling. We introduce the EVE software and illustrate some of its applications for models of pattern detection, pattern discrimination, and motion detection.

Modeling software Visual models Computer simulation Early vision

Recently, there has been an increase in the use of computer simulation for the analysis of models of early visual processing. Models have been proposed for such tasks as detection and pattern discrimination (e.g. Watson, 1983; Watt & Morgan, 1983; Wilson & Gelb, 1984), motion detection (Adelson & Bergen, 1985; Barlow & Levick, 1965; Marr & Ullman, 1981; van Santen & Sperling, 1985; Watson & Ahumada, 1985), and so on. Each such model involves a large array of processors which is applied to an input stimulus. The processor outputs are combined in some nonlinear fashion in order to derive a measure which can be related to human performance. The complex nature of these models and of test stimuli makes it increasingly difficult to deal with these models analytically. Therefore, a number of researchers have begun to use computer simulation in order to test their models.

Consider, for example, the model proposed by Watson (1983) for spatial pattern detection and discrimination. This model consists of a large number of processors which linearly filter the input image. The processors vary in spatial position, spatial scale, phase and orientation. The processors are scaled in size with retinal eccentricity. Individual processor outputs are given a sensitivity so that the entire array (of some tens of thousands of processors) has a pooled response consistent with a measured contrast sensitivity function. The outputs of the processors are combined (using a Minkowski

vector length) to yield an estimate of d' in a detection task. For discrimination tasks, one computes the length of the vector difference between the model outputs for each of the two patterns to be discriminated. For simple input stimuli, such as sine wave gratings and Gabor patches (Gaussian windowed sine waves), the outputs of individual processors may be directly computed (assuming that the model processors' receptive fields are actually computed using a continuous input image, and not a receptor-sampled version of it). But, in order to apply the Watson model to more complex stimuli, computer simulation is imperative.

We have devised a software system for testing visual psychophysical models. The software is called EVE (the *Early Vision Emulation* software). Computational modeling is difficult and tedious work. EVE is intended to make the process far easier for the researcher. EVE consists of a series of programs written in the C language for use on any machine which uses the UNIX operating system. It requires no programming on the part of the user for a wide variety of applications, but some users may wish to extend its capabilities by adding new modules. For more detailed discussion of the design of EVE, the steps required to compute visual models, and detailed documentation, we refer the interested reader to Landy, Manovich and Stetten (1989) and Landy (1988).

EVE may be used to test any visual model consisting of a number of layers of processors

Fig. 1. A typical early vision model consisting of (A) a sequence of input images with luminance values $A_1(t), \dots, A_8(t)$ at time t ; (B) a layer of processors which calculate the output of a linear spatial filter (e.g. with a Gabor profile), square the result, and then apply a linear temporal filter; (C) a layer of processors which compute differences between pairs of layer 1 processors; and (D) the computation of a summary statistic of the final layer. Typically the computation of the summary statistic represents a "linking hypothesis" which relates model performance to human performance data. [This figure first appeared in *Behavior Research Methods, Instruments, & Computers*, Vol. 21, No. 5, 1989. Reprinted by permission. © Psychonomic Society, Inc.]

which are wired together in a feed-forward manner. Figure 1 illustrates an example of the sort of model which is easily implemented using EVE. The network consists of an input, two layers of processors, and a computation (or *linking hypothesis*) which relates the output of the second layer to performance in a detection task. The input is a temporal sequence of visual images. The first layer has many processors, each of which computes the output of a linear spatial receptive field, squares the result, and then applies a temporal filter. The second layer consists of processors which compute the difference between particular pairs of processors in the first layer. The final stage computes a summary statistic of the activation of the second layer processors.

As in Fig. 1, models implemented using EVE consist of a sequence of arrays of processors wired in a feed-forward fashion. The inputs to each array of processors can include a temporal sequence of images (arrays of luminance values) or the outputs of previously computed arrays of processors. The computation performed by a particular array of processors consists of a receptive field computation for each processor (either linear or nonlinear), followed by a sequence of spatially pointwise operations

which treat each processor output independently. Spatially pointwise operations include instantaneous nonlinearities, temporal filtering, addition of noise, etc. The sample positions in each array are unconstrained; they can form a square array, a hexagonal array, or a set of arbitrarily specified locations. EVE can be used to compute the results of such a model as applied to a variety of stimuli, and then compute statistics of the activity of processors in each stage of the model. These statistics may then be related to human performance with similar stimuli, and compared to performance of alternative models.

EXAMPLE APPLICATIONS OF EVE

Here, we discuss several examples of using the EVE software to explore visual models. The list of examples is intended to give an indication of the gamut of uses to which EVE might be applied. It is by no means exhaustive. EVE can also be applied to problems of binocular brightness summation by using two input layers, one for each eye. One could use the motion detection mechanisms discussed below as an input to a further motion processing stage, such as the computation of speed.

The Watson spatial model

As our first example, consider the model for spatial pattern detection and discrimination described by Watson (1983). This model applies a large set of linear processors to its input (Watson refers to them as *sensors*). The processors come in eight families, indexed by the preferred spatial frequency of the central processors in that family: 0.25, 0.5, 1, 2, 4, 8, 16 and 32 cpd. Within a family, processors are placed in a series of rings, always sampled to conform with the Nyquist criterion for the preferred spatial frequency. The processors scale in size with eccentricity. Within a given family, and at a given spatial sampling location, there are ten processors: sine and cosine phase Gabors with one of five possible orientations (0, 36, 72, 108 or 144 deg). Each processor is given a sensitivity which is a function of its size and spatial frequency family. The sensitivities are the only free parameters in the model, and are set so as to make the model output for simple grating patches conform to a measured contrast sensitivity function.

The output of the processors is considered as a large processor output vector. Any given

spatial input pattern results in a particular output vector. The discriminability of two patterns is modeled as a monotonic function of the distance between their resulting output vectors. The detectability of a pattern is modeled as a function of the length of the resulting output vector, which is the same as the discriminability of the pattern from the zero uniform field. The length is computed using a Minkowski metric with an exponent of 3.5. The use of an exponent greater than 2.0 has been shown to be a reasonable model of performance under stimulus uncertainty (Pelli, 1981, 1985).

It is easy to compute the results of this model using EVE. The model fits precisely the model format suggested by Fig. 1. The model consists of two layers. The first layer contains a large number of processors indexed by spatial position, preferred spatial frequency, orientation, phase, and sensitivity. Each receptive field output is computed for a given stimulus, resulting in a large vector of processor outputs. The second layer simply computes a gross measure of activity in the first layer (the vector length), which is related to the detectability of the input pattern.

We have run this simulation for a number of frequencies and orientations of input pattern. A typical input pattern and subfamily output are illustrated in Fig. 2a and b, which were created and displayed using EVE. The individual processors are visible as dark or bright squares corresponding to a receptive field result which is positive or negative. The strongest output came from the middle processor as expected, since it was tuned to 4 cpd, had a profile which was in sine phase, and was located at the zero crossing of the input grating stimulus. The other processors were located on a series of rings centered on the central one, and their responses grew progressively weaker with eccentricity because of the eccentricity scaling of the model (the outermost sensors had a preferred spatial frequency nearly an octave lower than the central processor). Processors with very weak outputs were represented using a medium grey tone; all of the outermost processors were weakly activated.

A variety of model results are given in Table 1. As expected, a particular subfamily yielded the greatest response to a stimulus with the orientation of that subfamily (Table 1a). A given family responded well to the spatial frequency to which it was tuned (Table 1b) but usually responded more strongly to somewhat lower spatial frequencies. Because of the eccen-

Table 1. Summary results from the Watson (1983) model of spatial pattern detection and discrimination. All tables give the response of a subset of the processors to a particular input grating. The vector length of that subset of processors is shown using a Minkowski metric with an exponent of 3.5, normalized so that the largest response in each table is 100. (a) Response to a 4 cpd vertical grating by 4 cpd processor sub-families with various orientation preferences (pooled across processors in sub-families with that orientation and either phase). (b) Response to various input spatial frequencies by the 4 cpd vertical processor sub-family (pooled across processors in the 4 cpd sub-families with 0 deg orientation and either phase). (c) Response by each of six processor families to each of eight input spatial frequencies (pooled across orientation and phase)

(a)						
Processor orientation	Relative response					
0	100.0					
36	5.8					
72	1.0					
108	1.0					
144	5.8					
(b)						
Input spatial frequency	Relative response					
0.25	1.8					
0.5	1.5					
1	12.5					
2	100.0					
4	93.6					
8	0.7					
16	0.2					
32	0.1					
(c)						
	Processor family (preferred SF at fovea)					
Input SF	1	2	4	8	16	32
0.25	0.31	2.13	3.17	2.87	0.39	0.00
0.5	5.20	7.36	2.76	2.70	0.38	0.00
1	21.09	64.37	12.57	2.45	0.36	0.00
2	0.22	83.12	100.00	13.37	0.33	0.00
4	0.05	0.77	93.46	94.74	1.86	0.00
8	0.02	0.24	0.92	76.10	12.87	0.01
16	0.01	0.07	0.29	0.77	9.91	0.05
32	0.00	0.03	0.11	0.28	0.11	0.04

tricity scaling, the nominally "4 cpd" family actually contained only one processor tuned to 4 cpd (the central one) and a great number of sensors tuned to frequencies as much as an octave lower (for a 4 deg wide processor array). When the families were weighted by different sensitivities as suggested by Watson, we found that the greatest contribution to the total model response was not always by the subfamily which was best tuned to the stimulus, but rather by mistuned subfamilies with higher overall sensitivity. This was especially true at the lowest spatial frequency (0.25 cpd, Table 1c), where the greatest contribution to the response was from the 4 and 8 cpd processor families. (The 0.25

and 0.5 cpd processor families were not computed because their receptive fields were all larger than the stimulus.) Cosine phase processors had nonnegligible responses at low frequencies (even at d.c.), and therefore, one can see a substantial response from the more numerous, more sensitive, higher frequency (4–8 cpd) processors to the 0.25 cpd stimulus, dominating the response of the model as a whole to this pattern.

Finally, as one would expect of such a large model, there were difficulties simulating such a vast array of processors. Typically, one would prefer to run the entire array of processors on the same (digitized) input stimulus. But, if the stimulus is sufficiently well sampled to avoid major input sampling artifacts in the 32 cpd family, then one must have a huge input image, resulting in unfeasible amounts of computation and storage. If the input image is sampled reasonably densely for the low frequency families, it may still be poorly sampled for the mid-frequency families, resulting in a large artifactual response from these numerous, sensitive processors. For the simulations reported here we used a 4×4 deg stimulus with 100 input samples per deg. This resulted in sufficient sampling for the 32 cpd family (the Nyquist rate is 64 samples per deg), but in a long amount of time to compute the simulations. Alternatively, one can determine (by pilot simulations) which families are important to a particular discrimination, and tailor the input sampling to those processor families.

The kind of computation we have just described is not specific to the Watson model. For example, the line element model for pattern discrimination described by Wilson (1986) is quite similar. It also consists of a set of linear processors characterized by spatial position and receptive field profile (in this case a difference of three Gaussians). Each receptive field computation is followed by a nonlinearity, resulting in an output vector for a given pattern. The distance between two output vectors is again the model prediction for pattern discriminability. The first processor layer comprises a resampling (the receptive field computation) followed by a spatially pointwise operation (the nonlinearity applied to each receptive field output). This model can be computed easily using EVE.

Retinal sampling

Because of EVE's generality with respect to spatial sampling, it is well-suited to exploring the consequences of retinal (and subsequent)

spatial sampling. Figure 2c shows the result of hexagonally sampling a portion of the sine wave grating shown in Fig. 2a using Gaussian sampling functions, as computed and displayed using EVE. We have used this technique to create simulated foveal retinal images. The processing was based on two simplifying assumptions: (1) the sampling array was a perfect hexagonal lattice with a receptor spacing of 30 sec; and (2) the sampling function (which models the combined effects of the optical point-spread function and receptor aperture) was a Gaussian with a width at half height of 1.3 min. This technique was applied to a pair of 1×1 deg sine wave grating patches (4 and 8 cpd). The resulting hexagonally sampled images were used as input for the Watson spatial model outlined above. Since the EVE routines used in the Watson simulation are general with respect to the input image sampling properties, this was a very simple task. Table 1 lists results for a 4×4 deg input image which had dense square sampling. Using hexagonal sampling and a 1×1 deg stimulus, the orientation tuning was broader (8.3% of maximum for the 36 deg family, compared to 5.8% in Table 1a). The spatial frequency tuning of each family was similar to Table 1b. For square sampling, the 8 cpd family was slightly more sensitive to a 4 cpd stimulus than the 4 cpd family. For the hexagonally sampled image, the 4 cpd family was three times more sensitive to a 4 cpd grating patch than the 8 cpd family. This was a simple consequence of using a smaller grating patch (1 deg diameter rather than 4 deg for the square sampling), resulting in a far weaker effect of the eccentricity scaling. This is an example of a new result which is easy to calculate using the EVE software.

Motion detection

EVE may also be used with models of motion detection. For example, we have simulated the motion energy model of Adelson and Bergen (1985). The input sequence (which was generated using EVE) was a 2 deg wide patch of 4 cpd sine wave grating which oscillated back and forth sinusoidally (one cycle per 40 time steps), with an amplitude of oscillation of 0.35 deg. At each processor location, a sine-phase and cosine-phase Gabor receptive field was cross-correlated with the input. The output of each of the sine- and cosine-phase Gabor operators was processed by each of two temporal filters. Both temporal filters were bandpass, but one had a

Fig. 2(a). *Caption on p. 335.*

Fig. 2(b). *Caption opposite.*

Fig. 2(c).

Fig. 2. A typical input and output from the Watson (1983) model of spatial pattern detection and discrimination as computed using EVE. (a) A 4 deg patch of 4 cpd vertical sine wave grating. (b) The outputs of all processors in the sub-family of processors which are tuned to 4 cpd at the fovea, are all in sine phase, and are all tuned to vertical gratings. The result of correlating each receptive field with the input stimulus is coded so that positive outputs result in a dark spot in the image at the processor position, and negative outputs result in a light spot. The array consisted of a series of rings in which the more eccentric processors were tuned to lower and lower frequencies, and hence were less responsive to the 4 cpd input. Processors with weak outputs are pictured using a medium grey tone; all of the outermost processors were weakly activated. (c) The result of sampling a 0.25 deg patch of grating in (a) using a hexagonal grid of processors with Gaussian sampling functions.

Fig. 3. Results of the simulation of an array of Adelson and Bergen (1985) motion processors. The stimulus oscillated left (for 10 time steps), then right (for 20 time steps), then left (for 20 time steps), then right again (for 10 time steps) in a sinusoidal manner. The solid line gives the grating position x as a function of time. The dotted line is the instantaneous velocity dx/dt of the grating. The dashed line is the average motion processor output. All curves were normalized to give a maximum value of 1.0. The opponent motion output was generally negative for leftward motion, and positive for rightward motion. The stimulus speed during the middle of an oscillation was sufficiently high (during time step 23, for example) to alias to motion in the opposite direction for these processors, given the poor temporal sampling of the input stimulus.

more delayed impulse response than the other (see Adelson and Bergen, 1985, for details of the filters). Thus, the resulting outputs had been filtered by the composition of a spatial and a temporal filter. In other words, in each case a space-time separable filter had been applied. Linear filters were constructed from these by sums and differences which were no longer space-time separable functions of the input sequence, and which had receptive fields which were oriented in space-time (i.e. were tuned for motion direction). The power in each of these processors (the square) was then computed, followed by a sum of the two phases in each direction. Finally, the opponent motion signal was computed as the difference between the

processors which prefer leftward and rightward motion. A sample of the average outputs from the final layer of processors is given in Fig. 3. We can see both appropriate responses to motion and, when the velocity was too high relative to the temporal sampling rate of the image sequence, fast rightward motion produced artifactual leftward motion responses (and vice versa). The Adelson and Bergen motion detector does not alias fast rightward motion into leftward responses for continuous motion, but does give spurious responses to poorly sampled motion. This is not a fault of the model *per se*, but rather a consequence of spurious spatio-temporal Fourier components in the sampled stimulus itself.

Both van Santen and Sperling (1985) and Adelson and Bergen (1985) have shown that the elaborated Reichardt detector (or ERD, Reichardt, 1957, 1961; van Santen & Sperling, 1985) and the motion energy detector (Adelson & Bergen, 1985) are identical at their final outputs for particular realizations of each detector. Although the proof is mathematically trivial, we have demonstrated it using a simulation of a simple ERD. This version of the ERD used the same space-time separable filter outputs from the Adelson and Bergen simulation. Nonlinear subunits tuned to leftward and rightward motion were then computed by cross-multiplying appropriate linear filter outputs. Then, leftward- and rightward-tuned subunits were set in opponency. This set of detectors, when scaled by a factor of 4, produced precisely the same set of processor outputs as the Adelson and Bergen simulation described above. Of course, the two models differ in the outputs available at intermediate stages, which has important consequences for relating these models to the physiology (Emerson, Bergen & Adelson, 1987). The subunit responses of each model to the input stimulus are available as a by-product of the EVE simulations, which easily demonstrates the models' differences as well as their similarities.

CONCLUSIONS

We have described some applications of the EVE system for the simulation of early vision. We hope that EVE will be used by modelers in the field of visual sciences*, and that it will make it easier for researchers to share and compare models, and to apply them to new paradigms. For this reason, we are making EVE freely

*EVE may be obtained by mailing either a 1/2-inch tape or 1/4-inch tape cartridge to Michael Landy, New York University, Department of Psychology, 6 Washington Place, Rm 961, New York, NY 10003, U.S.A. Alternatively, it may be obtained via anonymous ftp to Sun file server 'vml.psych.nyu.edu' (internet number 128.122.132.4, the file is 'eve.tar.Z' in directory 'pub'). EVE should run on any UNIX machine with no, or only minor changes. Please notify us that you are using EVE so that we may inform you of new versions and bug fixes. We would appreciate copies of useful extensions made to EVE by others so that we may include them in subsequent versions.

available to researchers. There are two major drawbacks to the use of EVE. First, as mentioned above, brute force simulations of large scale models (such as the Watson simulation described above) can be extremely slow and computationally intensive. EVE makes it easy to carry out such a simulation, but it does not ameliorate the amount of computational resources required. Second, EVE is restricted to entirely feed-forward models. This excludes models which require feedback from later stages to earlier ones (e.g. for gain control) or feedback within a processor layer (e.g., lateral inhibition). We are planning to extend EVE for use as a tool for visual neural modeling as well. As such, we plan to add to EVE the capability of simulating networks that are not entirely feed-forward, including both forms of feedback just mentioned.

As it stands, EVE should be useful as a simulation tool in a number of areas including spatial detection and discrimination, sampling, motion detection and discrimination, and velocity computation. Other possible areas include lightness computation, binocular brightness combination, etc.

Acknowledgement—The design and implementation of the EVE software was supported primarily by a grant from the National Science Foundation, Information Science and Technology Grant IST-8418867, and in part by a grant from the Office of Naval Research, Grant ONR-N00014-K-0077. Special thanks are due, as always, to Robert Picardi for technical assistance, and to J. Anthony Movshon, Sofia Würger, and Charles Chubb for careful editing. Portions of this work have been presented at the *Annual Meeting of the Association for Research on Vision and Ophthalmology*, Sarasota, Florida, May 6 1988, and published in abstract form (Manovich & Landy, 1988).

REFERENCES

- Adelson, A. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2, 284–299.
- Barlow, H. B. & Levick, W. R. (1965) The mechanism of directionally selective units in rabbit's retina. *Journal of Physiology, London*, 178, 477–504.
- Emerson, R. C., Bergen, J. R. & Adelson, E. H. (1987). Movement models and directionally selective neurons in the cat's striate cortex. *Society for Neuroscience Abstracts*, 13, 1623.
- Landy, M. S. (1988). The EVE Early Vision Emulation Software: Reference Manual. Mathematical Studies in Perception and Cognition 88-11, New York University.
- Landy, M. S., Manovich, L. Z. & Stetten, G. D. (1989). All about EVE: The Early Version Emulation software. *Behavior Research Methods, Instrumentation, & Computers*, in press.
- Manovich, L. Z. & Landy, M. S. (1988). EVE: Software for psychophysical modeling. *Investigative Ophthalmology & Visual Science* (Suppl.) 29, 447.
- Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B*, 211, 151–180.
- Pelli, D. G. (1981). Effects of visual noise. Ph.D. Thesis, University of Cambridge, Cambridge, England.
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, 2, 1508–1532.
- Reichardt, W. (1957). Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems. *Zeitschrift Naturforschung*, 12b, 447–457.
- Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In Rosenblith, W. A. (Ed.) *Sensory communication*. New York: Wiley.
- van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A*, 2, 300–321.
- Watson, A. B. (1983). Detection and recognition of simple spatial forms. In Braddick, O. J. & Sleigh A. C. (Eds.) *Physical and biological processing of images* (pp. 100–114). New York: Springer.
- Watson, A. B. & Ahumada, A. J. Jr (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A*, 2, 322–342.
- Watt, R. J. & Morgan, M. J. (1983). The recognition and representation of edge blur: Evidence for spatial primitives in human vision. *Vision Research*, 23, 1465–1477.
- Wilson, H. R. (1986). Responses of spatial mechanisms can explain hyperacuity. *Vision Research*, 26, 453–469.
- Wilson, H. R. & Gelb, D. J. (1984). Modified line-element theory for spatial-frequency and width discrimination. *Journal of the Optical Society of America A*, 1, 124–131.