PERCEPTUAL CODING FOR 3D RECONSTRUCTION

Christopher W. Tyler and Spero C. Nicholas
Smith-Kettlewell Eye Research Institute, San Francisco C 94115, USA

Paper ID: 66

ABSTRACT

A primary issue in 3D reconstruction is the real-time efficacy of different coding methods for the multiple decisions among competing 3D solutions. A common model framework making such coding decisions is the boundary limited drift-diffusion model, which has been developed in parallel in various branches of science from quantum physics to economics. A common property of all such models is the linear increase in variance of the diffusion processes over time, implying an inability to focus on the current information in the environment, and the inevitability of a forced random decision in the absence of any reliable evidence. We have developed an alternative, more plausible model framework for Bayesian information accumulation that solves both problems and provides an accurate account of many features of context effects in human 3D reconstruction performance.

Index Terms— 3D reconstruction, decision-making, Bayesian, drift diffusion models, neural networks

1. BAYESIAN 3D RECONSTRUCTION

A key problem faced by computer vision systems operating in the real world is understanding the 3D structure of the image array that forms their input. To be effective, such systems will most likely need to incorporate a set of 3D models representing the Bayesian information about the structure of the world they inhabit and use the incoming information to decide which of these models offers the best representation of the world providing the visual input information. For example, if the vision system in question has the task of navigating a vehicle in a freeway environment, it has to model the disposition and movements of all the vehicles in the neighborhood and decide on the appropriate course of action. This task requires deciding which of the visual blobs in its purview represent moving vehicles and computing their velocities and likely trajectories. At least for human operators, these tasks require decisions as to where each object is located in 3D space at any moment in time, and is therefore a time-limited decision process.

2. DRIFT DIFFUSION MODELS

A common model framework for decision-making in such coding decisions is the boundary limited drift-diffusion model, which been developed in various branches of science under the names Random Walk, Brownian Diffusion, Boltzmann Transport Equation, Markov Process, Fokker-Planck Equation, Lévy Flight, Kolmogorov Equation, Wiener Process, and Ohrnstein-Uhlenbeck Process. These diffusion processes are defined by the common property that each state of the system differs from the state previous time step by a small random perturbation. However, it is the nature of all such models is that the variance of diffusion processes increases linearly over time. They differ in this respect from zero-memory random processes, which have no state dependencies and a constant variance over time.

The increasing variance of diffusion processes implies that they are only suitable for modeling noise processes with a defined (recent) starting point, since the noise variance would become indefinitely large as the starting point became progressive longer ago in time. This is not a problem in discrete computer decision models since they can assume a reset to zero just before the initial data input, but it becomes problematic for continuous monitoring to detect the occurrence of arbitrarily occurring events or realistic modeling of human decision-making processes. In this sense, the drift diffusion models have no intrinsic mechanism for focusing on the current information in the environment. In general, if a drift-diffusion mechanism has unlimited time to come to a decision, it will eventually be forced to a random decision even in the absence of any reliable evidence.

To address these deficiencies in the drift-diffusion models, we have developed an alternative, more plausible model framework for Bayesian information accumulation that solves the problem of how to focus on the current information and avoids the random decision problem. This framework also provides an accurate account of many features of human 3D reconstruction performance, providing an informative link between computational and neuronal mechanisms of 3D reconstruction.

3. BAYESIAN BIASING OF THE DIFFUSION MODEL

In its simplest form, the drift-diffusion model is an unbiased random-walk process. In application to decision models, however, it is biased by incoming information about the problem to be solved, and corresponds to the accumulation of positive or negative information relative to a particular solution. When sufficient information has accumulated to reach a preset positive or negative decision boundary, the decision is reached. As a model of neural information processing, the random process is the voltage inside a neuron governing the decision behavior to indicate a choice, and the decision boundary is the firing threshold for the neuron to emit an axonal spike.

In the context of 3D object recognition, the decision becomes more complex. The task is no longer a Yes/No decision, but a multidimensional choice as to which of many previously encoded objects is represented by the blob of information corresponding to each object in the visual field. Thus, the decision process may be considered a multidimensional random walk as the information accumulates relative to each of many object interpretations. In one- and two-dimensional random walks, the asymptotic distribution of locations visited is a Gaussian (or normal) distribution and expected value is the origin, expressed in the form that the probability of ultimately returning to the original starting point is 1. The information biasing adds linearly to this distribution such that the expected value corresponds to the cumulative information added to the random walk process.

For dimensions of three and higher, however, the random walk process has the remarkable property that the probability of returning to the origin is less than 1, so the concept of an expected value loses its meaning. Although the probability distribution is still a hyperspherical Gaussian, the current value is essentially adrift in empty space. Thus, if the dimensionality of the diffusion space is 3, the probability of ever returning to the origin drops to an irrational value of about 0.34, and it falls progressively as the dimensionality of the random-walk space increases (Finch, 2003). Given that the coding space of shapes of recognizable objects is hyperdimensional, the decision space as to which object is present in a given location in space must itself be hyperdimensional, creating inherent difficulties for biased diffusion as a model of the 3D object recognition process. In neural terms, decision based on the independent outputs of 1-3 neurons may have a defined expected value, but those that require canvassing

4 or more neurons will have a noise space with no definable expected value under the classic drift-diffusion model.

These considerations highlight the need for a decision model that a) has a stable noise variance over time, to avoid making decisions until an appropriate level of information has accumulated, and b) operates realistically in a hyperdimensional environment in both discrete decision-making and vigilance tasks.

4. A LEAKY DRIFT-DIFFUSION MODEL

To address the deficiencies of the drift-diffusion model, we have developed an asymptotic leaky version of the diffusion process that automatically returns to baseline with some time constant. Leaky drift-diffusion models for the human decision process have been proposed by Usher and colleagues [3,4], but these authors have not extended it to the asymptotic level of a steady-state model of the neural decision process, as we do here. They focus on the dynamics of the leaky diffusion process in its transient domain, where the variance is still in the increasing regime, whereas we consider its drift-diffusion behavior once the variance has stabilized to its asymptotic level.

The value of this asymptotic approach is as a model of neuronal noise and neuronal decision processes, in which the leaky exponential time constant is a model of the membrane time constant governing the intracellular voltage following an input spike. There is general agreement that the intracellular voltage is a integrative random process with sequential dependencies, and it is well-established that the variance of the intracellular voltage is a stable property that does not increase over time. Thus, the natural way to incorporate all these properties is the leaky drift-diffusion model (LDD).

Information accumulation (I) is modeled with the following Itō process:

$$\mathrm{I}(T+t) = \int_0^{T+t} \mu\, \mathrm{u}(x-T)\, dx + \int_0^{T+t} \exp\left(-\frac{T+t-x}{\tau}\right) \dot{W}(x)\, dx$$

where $\dot{W}$ is the derivative with respect to time of a Wiener process, u is the unit step function, τ is the time constant of the leak, T is a pre-stimulus epoch that allows the integrated random variable to reach a quasi-steady state, and μ scales the rate of information accumulation relative to the standard deviation of the noise. At large T, the variance of I asymptotes to $\tau/2$.

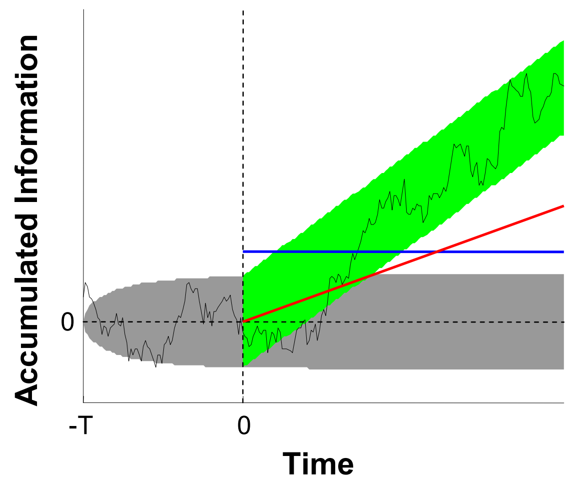

Fig. 1. Simulations of the LDD. Mean ± 1 standard deviation of unbiased noise signal in the absence of a stimulus (gray patch). Mean ± 1 standard deviation of accumulated information once stimulus is presented (green patch). Fixed threshold (blue line). Optimally discriminating threshold (red line). Example random signal (black line).

The decision making criterion, C, was modeled as follows

$$C(T+t) = \sqrt{c_0^2 + (c_1 \mu t)^2}$$

For an individual stimulus of duration $t>0$, the probability of a correct detection is defined as the probability that $\mathrm{I}(T+t) > C(T+t)$ when $\mu>0$

$$p = \Phi\left(\mu t - \sqrt{c_0^2 + (c_1 \mu t)^2}\right)$$

where Φ is the standard normal cumulative distribution function. The probability of a false alarm is given by

$$q = \Phi\left(-\sqrt{c_0^2 + (c_1 \mu t)^2}\right)$$

5. EXPERIMENTAL METHODS

The computational LDD was validated in the context of a basic 3D surface reconstruction task. The reconstruction domain was stereoscopic matching of points in a 3D dynamic noise environment to form the 3D images of flat disks defined by binocular disparity alone. Human subjects were shown targets consisting of a pair of stereoscopic disks preceded and followed by fields of dynamic noise dots that were binocularly uncorrelated, and appeared as a 3D snowstorm. Viewed with one eye, there was no detectable structure or events of any kind.

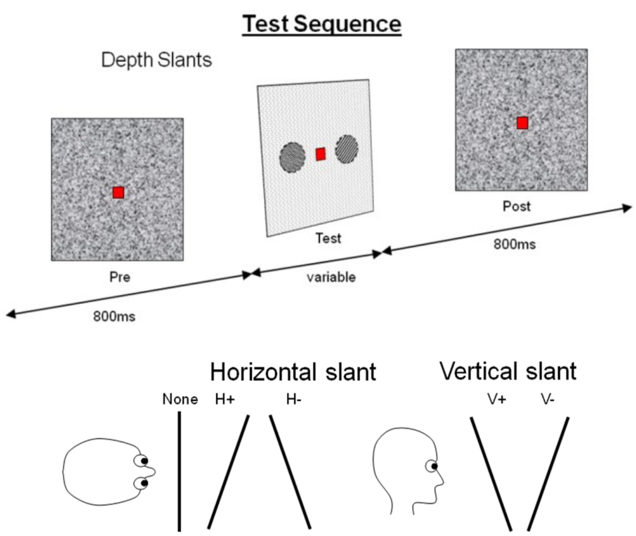

Fig. 2. Stimulus arrangement for the visual 3D reconstruction task (see text).

Each disk in the pair had a 50% probability of appearing, such that on any given trial there could be two, either one, or no disks. The subject's task was to indicate whether the presentation on the two sides was the same (two disks or none), or different (left or right disk). The main experimental manipulation was whether the disks were embedded in a plane of the same 3D slant (defined in the field of correlated noise dots) or in a 'snowstorm' of binocularly uncorrelated noise dots. The task was performed for a range of disk presentations from 0.05 to 1 sec.

The SAME/DIFFERENT task was unusual in that the decision function had different shapes for different outcomes. Assuming left/right symmetry, there were three different ways in which the patches could become different from the surround during the trial, in each of which the decision could be right or wrong, making six qualitatively different outcomes. The three cases were those in which Both Patches, One Patch or No Patches appeared.

The probabilities of a correct response when No, One or Both targets were present are given by

$$P_0 = (1-q)^2 + q^2$$
$$P_1 = p(1-q) + (1-p)q$$
$$P_2 = p^2 + (1-p)^2$$

6. RESULTS

The experiment was run of 4 subjects with normal stereoscopic vision. The performance in the decision task was tracked as a function of the presentation duration of the test and surround stimuli (Fig. 3). Unusually for such a decision task, the probabilities of a correct decision differed widely across the 3 cases.

The No Patch case was correct a far higher proportion of the time than the Both Patches cases, implying a perceptual bias toward no presentation. The most unusual function was the One Patch case, which showed a non-monotonic function with a dip at intermediate durations and a continued lower sensitivity than the other two cases. A similar pattern of behavior was seen for both the correlated noise surrounds containing the planar 3D slant information and the uncorrelated noise surrounds. This feature was only captured by the fully parametrized model, which accounted for 87.8% of the variance, as opposed to 77.8% in the case of the classic Signal Detection Theory model.

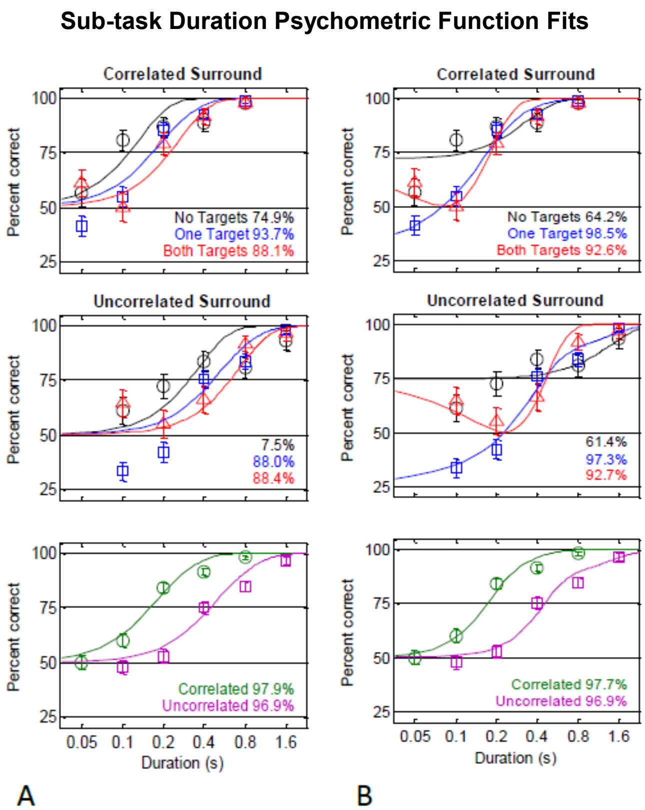

Fig. 3. Comparison of model fit to an example of human performance in the two-patch 3Ddiscrimination task. Upper panels: Duration psychometric functions for Both, One or No patches decision tasks with Correlated and Uncorrelated Surrounds. Bottom panels: Combined duration functions for the two surround types. A. Fits of the classic Signal Detection model. B. Fits of the optimal criterion model with an upper limit. Numbers indicate variance accounted for by each function fit.

To quantify the effect of the surround slant on the discrimination of the slanted test patches, the overall percent correct was cumulated over the three cases to form discrimination functions. The intersection of these functions with the 75% correct level provides a measure of the decision time of the perceptual decision process (Fig. 3, bottom panels). In the example for one observer, it can be seen that the functions are separated by about a factor of 3, with the discrimination much easier when the surround is correlated than when it is an uncorrelated binocular snowstorm (i.e., depth noise rather than a defined depth plane.).

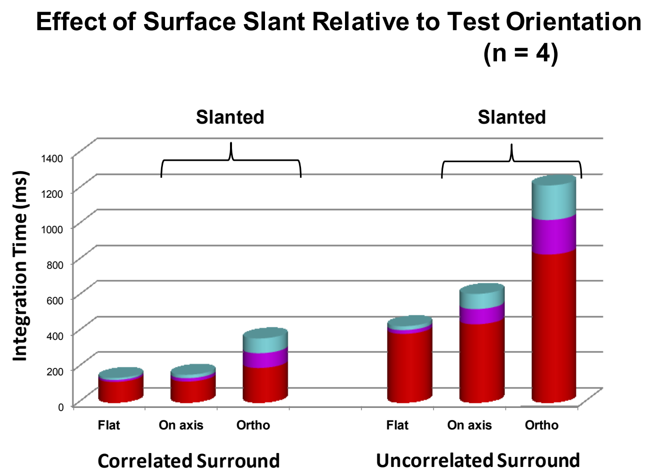

Fig. 4. Mean decision times (±1 SEM) from the model fits for the 4 subjects are graphed as a function of surround correlation and slant type. Note lack of significant on-axis slant effect but large effects of orthogonal axis and type of surround correlation.

The task involved four slanted surround planes in addition to the flat plane (to which the test patches were also coplanar with matching slants). The five slant conditions were interleaved at random in an unpredictable series. The slants can be categorized into two classes – the cases where the slant is around the axis of the two patches and those where the slant axis is orthogonal to that axis (which are therefore rotated out of the plane of fixation). The mean decision times for a group of subjects are shown in Fig. 4.

The results show a main effect of surround correlation, with the decision times about three times longer with the binocularly uncorrelated surrounds than in the presence of the simultaneously correlated surround planes. Slant around the axis through the patches has no significant effect on decision time (relative to that for the flat planes), but orthogonal slants approximately double the decision times relative to the other slant conditions.

7. DISCUSSION

We report the development of a self-biasing, or leaky, drift diffusion model of decision processes that has applicability for all levels biological processing from single neuron firing to individual and even societal political decisions. This model overcomes a largely unrecognized deficiency in classic drift diffusion models that they assume noise processes with linearly increasing variance, implying a recent starting point for the prevailing noise processes, whereas in fact virtually all biological information processing systems have noise with stable variance, or steady-state noise. We regard the development of a drift diffusion model with stable variance as a major contribution to the field of decision theory, here applied to the analysis of context effects on human decision behavior in a 3D reconstruction task.

To isolate the 3D cue from the 2D spatial information that carries it, we use the Juleszian paradigm of cyclopean dynamic noise fields. The depth form of the patches was defined solely by binocular disparity cues within the dynamic noise fields. The task required solving the binocular correspondence problem of which pairs of dots in the two eyes go together before the form of the patches can be perceived. Even though the 3D reconstruction is limited to identification of slanted patches, it is still a challenging task for human or computational visual systems.

The decision task of determining the number of patches was modeled in the framework of Signal Detection Theory. Here another innovation was in the application of a biased (non-Ideal) criterion to the separate analysis of each type of response with respect to the presentation of the two patches. Incorporating the free parameter of an adjustable decision bias allowed a better fit to the human decision performance and captured the unexpected feature of a dip in the discrimination function for the case of Both Patches.

In general, an important role for the 3D context of the patches was established for this stereoscopic reconstruction task, despite the fact that they were presented simultaneously with, rather than prior to, the test patches. The decision time required to detect the presence of the patches was on average three times shorter in the presence of the stereoscopic surround than without it, regardless of the slant condition. Thus, the human visual system was able to make use of the surround information to reduce the decision time with respect to the state of the patches, even though that required processing more region of the image than if attention had been paid to the patches alone. This capability implies effective use of the parallel processing capabilities of the neural architecture early during the target/ surround presentation period, together with actives decision processes for optimize the criteria for distinguishing signal from noise in the context of solving the binocular correspondence problem. Although we view the world as mostly static, our own activities turn it into a dynamic presentation on the retinas of the two eyes, especially in fast-paced situations with which we challenge ourselves in many sports activities. Understanding the nature of the visual decision processes is therefore of particular value in these challenging environments.

8. ACKNOWLEDGMENTS

Supported by AFOSR grant FA9550-09-1-0678

9. REFERENCES

[1] Tyler, C.W. The role of midlevel surface representation in 3D object encoding. In *Computer Vision: From Surfaces to 3D Objects*. Tyler, C.W. (Ed) Chapman Hall: New York, p. vii-xxiii, 2011.

[2] Bogacz, R., Usher, M., Zhang, J., McClelland, J.L. Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. *Philos Trans R Soc Lond B Biol Sci.* 362(1485):1655-70, 2007.

[3] Finch, S. R. "Pólya's Random Walk Constant." §5.9 in *Mathematical Constants.* Cambridge, England: Cambridge University Press, p. 322-331, 2003.

[4] Usher, M. & McClelland, J. L. On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108*, 550-592, 2001.