LETTER **Communicated by John Perrone**

A Neural Circuit for Robust Time-to-Contact Estimation Based on Primate MST

N. Andrew Browning
buk@bu.edu
Center for Computational Neuroscience and Neural Technology, Boston University, Boston, MA 02215

Time-to-contact (TTC) estimation is beneficial for visual navigation. It can be estimated from an image projection, either in a camera or on the retina, by looking at the rate of expansion of an object. When expansion rate *(E)* is properly defined, *TTC* = *1/E*. Primate dorsal MST cells have receptive field structures suited to the estimation of expansion and TTC. However, the role of MST cells in TTC estimation has been discounted because of large receptive fields, the fact that neither they nor preceding brain areas appear to decompose the motion field to estimate divergence, and a lack of experimental data. This letter demonstrates mathematically that template models of dorsal MST cells can be constructed such that the output of the template match provides an accurate and robust estimate of TTC. The template match extracts the relevant components of the motion field and scales them such that the output of each component of the template match is an estimate of expansion. It then combines these component estimates to provide a mean estimate of expansion across the object. The output of model MST provides a direct measure of TTC. The ViSTARS model of primate visual navigation was updated to incorporate the modified templates. In ViSTARS and in primates, speed is represented as a population code in V1 and MT. A population code for speed complicates TTC estimation from a template match. Results presented in this letter demonstrate that the updated template model of MST accurately codes TTC across a population of model MST cells. We conclude that the updated template model of dorsal MST simultaneously and accurately codes TTC and heading regardless of receptive field size, object size, or motion representation. It is possible that a subpopulation of MST cells in primates represents expansion in this way.

1 Introduction

Time-to-contact (TTC) estimation is potentially important for animals when planning obstacle avoidance. TTC is defined mathematically as the time until an object crosses an infinite plane parallel to the image plane at the position of the focal point (see Figure 1). Mathematically it does not imply

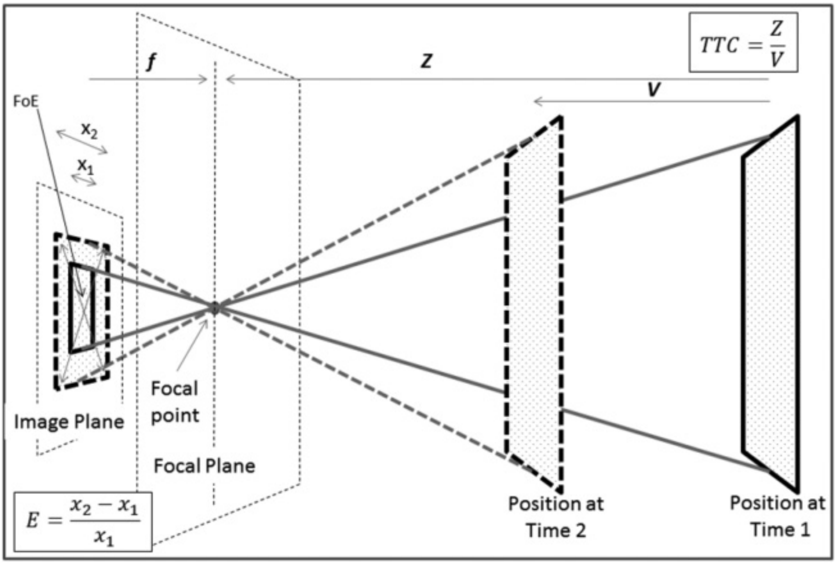

Figure 1: Schematic diagram showing the projection of a moving planar object in space (right) on to an image plane (far left). The object plane is parallel to the image plane and starts at distance Z from the focal plane. It moves with velocity V along the Z direction. TTC is the time until the object crosses the focal plane. In this diagram, TTC is also the time to collision, but the object could be shifted vertically such that TTC remains the same but no collision occurs. The size of the object, X, does not change during its approach, but the projection of X on to the image plane, x, expands from the FoE as the object gets closer. Expansion rate, E, can be defined independent of object size and camera properties such as focal length, resolution, and pixel pitch, as shown. All planes in the figure have the same surface normal (are parallel) but are drawn with perspective for clarity.

a collision between the observer and the object. TTC can be calculated from visual information by looking at the expansion of an object in the visual field. If expansion (E) is defined as the rate of growth of the object (e.g., 20% growth = 0.2) per unit time then the TTC (T) can be derived from expansion (see appendix A):

$$T \equiv \frac{1}{E}. \tag{1.1}$$

Equation 1.1 is usually represented as $x/\dot{x}$, where x is the size of the object on the image plane (Horn, Fang, & Masaki, 2007; Lee, 1976), or by using the ratio of the distance from the focus of expansion (FoE) with motion vectors in the image, in which case x is the distance from the FoE (Byrne & Taylor, 2009; Longuet-Higgins & Prazdny, 1980). Tau, and more recent variants, is the dominant model used for TTC estimation in neurophysiological and psychophysical experiments and is defined in terms of visual angle, θ, $\tau = \theta/\dot{\theta}$ (Hoyle, 1957; Lee, 1976). Tau is not equal to TTC, although it is a reasonable approximation for small angles when $\tan\theta \cong \theta$. TTC can also be measured from a decomposition of the optic flow field via the divergence component (Koenderink & van Doorn, 1976).

TTC estimation is not strictly necessary for obstacle avoidance. Several species have been shown to perform avoidance maneuvers when a stimulus reaches a certain angular size (Nakagawa & Hongjian, 2010). However, Lee (1976) demonstrated that humans use a measure similar to TTC for playing ball games and driving a car. In the psychophysics literature, the term *looming* is often used to describe an object that expands on the image plane. Cells that respond to looming have been found in a variety of animals (Hayes & Saiff, 1967; Judge & Rind, 1997; Schiff, Caviness, & Gibson, 1962; Wang & Frost, 1992). Looming responses may be more behaviorally relevant than literal TTC responses since an expansion-sensitive cell will become more active as the threat becomes greater rather than decreasing as a function of threat. Note that a theoretical TTC proportional response cell would have a high tonic activation and become less active as the object approaches. Cells may also be considered TTC responsive if activation rises to a peak and then decreases proportionally to TTC. In both cases, for the range of TTC that is being accurately encoded, activation decreases as TTC gets shorter. However, in the latter case, it has also been shown that peak timing can signify important behavioral information. For example, bullfrogs appear to have cells tuned for a peak response at a particular TTC and initiate avoidance maneuvers when those cells fire (Nakagawa & Hongjian, 2010). Pigeons have been shown to have cells that respond to both expansion and TTC (Wang & Frost, 1992).

In primates, there is no strong evidence for a TTC response in cellular data; however, in dorsal MST (MSTd), some cells respond to global patterns of expansion motion (Duffy & Wurtz, 1991a, 1991b; Tanaka, Fukada, & Saito, 1989). These cells generally have large receptive fields, on average over 1000 deg^2 (Raiguel et al., 1997). These MST cell responses are consistent with inputs from a large number of smaller receptive field MT cells, and the organization of the cells' selectivity for the MT inputs seems to define the pattern preference of the cells. If preferred directions of MT cells are arranged radially, then the resulting MST cell prefers expansion or contraction (Tanaka et al., 1989). MST cells that respond to expansive motion patterns can determine the focus of expansion (FoE) of the motion pattern, which often coincides with the observer's heading (Gibson, 1955). As a result, expansion-sensitive MSTd cells are often characterized as coding *heading.*

Perrone (1992) and Perrone and Stone (1994) demonstrated that MSTd could be modeled by a number of templates describing behaviorally important patterns of motion, primarily coding different headings. Each template defines the organization of inputs from MT that the MST cell responds optimally to. Template models of MSTd are able to explain human heading perception data in static environments (Browning, Grossberg, & Mingolla, 2009a; Perrone & Stone, 1994, 1998) and in the presence of independently moving objects (Layton, Mingolla, & Browning, 2012). Template models of MSTd can explain rotation data through the use of extraretinal signals to remove the effects of rotation before the template is applied (Beintema & van den Berg, 1998; Elder, Grossberg, & Mingolla, 2009). Template models thereby provide a functionally accurate model of primate MSTd expansion cells. Template models have been demonstrated to code relative depth, which is proportional to TTC in a static world (Browning et al., 2009a; Grossberg, Mingolla, & Pack, 1999; Perrone, 1992; Perrone & Stone, 1994). However, the output of these models is independent of neither the receptive field size of the cell nor the size of the object. For the purposes of producing depth maps of the environment, where the receptive field sizes of the cells are either known or constant, this may be sufficient. However, these template definitions are insufficient for the estimation TTC for approaching objects that do not fill the receptive field of the cell, in environments where the receptive field size is unknown, or for the precise estimation of TTC. Lappe (2004) further argues that the large receptive field sizes of MST cells are inconsistent with TTC estimation. However, if we assume that in general, TTC cells are not object specific, then to obtain an object-size independent TTC response, a cell will require a large receptive field to account for a range of objects throughout their approach trajectories.

The work described in this letter analyzes a general template model of MSTd to determine how TTC can be coded in MSTd regardless of the receptive field size of the cell, independent of the size of the object. This analysis is used to update the ViSTARS model (Browning, Grossberg, & Mingolla, 2009a, 2009b) to demonstrate how V1 and MT encode the required information to enable TTC estimation in MSTd.

2 Time-to-Contact Estimation in a Template Model of MSTd

A template model of MSTd consists of a number of templates, each corresponding to a particular motion pattern. In general, each motion pattern characterizes the expected motion for a particular heading direction (see Figure 2). In the model, a template is represented as a multidimensional array across space and motion. When using a standard 2D motion representation, there is a 2D vector (u, v) representing the expected motion at each spatial position across the input space. For a model consisting of an N-D representation of direction of motion, there is an N-D motion vector at each spatial location. Motion vectors, estimated from the image measurements, are compared against the template via the inner product between the template and the estimated motion. In its simplest form, the template with the highest inner product is considered the best match, and the motion

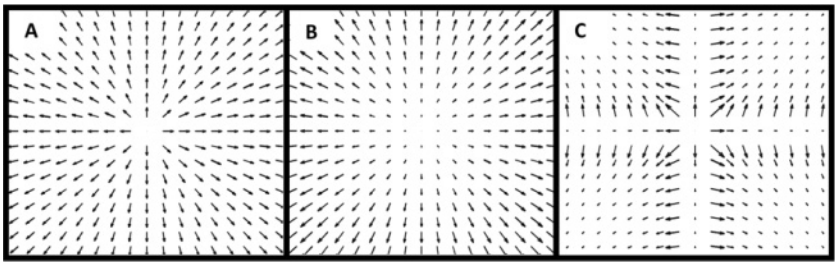

Figure 2: Pictorial representation of templates corresponding to a heading toward the center of a frontal plane. (A) Unit vector template. The direction of motion is assessed with respect to FoE. (B) Veridical vector template. The magnitude of each vector is scaled with distance from the FoE. This is the same as a motion pattern that occurs when approaching a frontal plane normal to the approach vector. (C) TTC template. The direction of the motion vector is not preserved, but there is still a singularity at the FoE. Note that the direction of motion within each dimension is preserved: all horizontal vectors point to the midline, all vertical vectors point to the midline, and the midlines cross at the FoE. When panel C is multiplied by a motion estimate toward a frontal plane (e.g., panel B) using an inner product, the result is the expansion rate of the plane.

pattern corresponding to the template is considered to be veridical. For example, the heading corresponding to the template with the best match is considered to be the current heading.

If we consider a planar object moving in a straight line at constant velocity toward an observer with the same surface normal as the image plane but at a distance Z from the observer, the motion pattern on the image plane will be a pure expansion pattern from the FoE (see Figure 1). The image motion vectors will have a smaller magnitude near to the FoE than they have farther from the FoE. More specifically,

$$\boldsymbol{v} \equiv \dot{y} \equiv \frac{y}{T} \equiv E\boldsymbol{y}, \tag{2.1}$$

where $\boldsymbol{v}$ is the image motion vector at a particular location, and $\boldsymbol{y}$ is a motion vector defined by the N-D distance of that location from the FoE. For example, to obtain a 2D motion vector $\boldsymbol{v}$, $\boldsymbol{y}$ is a 2D vector defined independently by the distance to the FoE in the two spatial dimensions. T is time to contact, and E is expansion rate. Since T and E are constant at any point in time, $\boldsymbol{v}$ scales linearly with distance from the FoE, as defined by the vectors in $\boldsymbol{y}$. In all equations, we use bold notation to define vectors, lowercase represents a single vector, and bold uppercase represents a set of vectors across spatial location.

We define $\boldsymbol{Y}$ and $\boldsymbol{V}$ to contain one vector, $\boldsymbol{y}$ and $\boldsymbol{v}$, for every spatial location represented by the cell (see Figure 2).

If templates are constructed to incorporate speed, that is, the vector magnitudes (see Figure 2B) and there is no noise in the motion vector estimates, then the motion estimate and the template are the same; both are equal to $E\mathbf{Y}$, and the template match, defined by the inner product ($\boldsymbol{M}$), will be

$$\boldsymbol{M} = E^2(\boldsymbol{Y}^T\boldsymbol{Y}), \tag{2.2}$$

which is dependent on both the number of elements in $\boldsymbol{Y}$ (the number of spatial locations represented in $\boldsymbol{Y}$) and the maximum value of $\boldsymbol{Y}$ defined by the template (i.e., the maximum distance from the origin represented in $\boldsymbol{Y}$, which defines the receptive field size of the template). Cells with inputs from more spatial locations (more vectors represented in $\boldsymbol{Y}$) have a larger response. Cells with larger receptive field sizes have a larger response.

If the templates are constructed such that the vectors in $\boldsymbol{Y}$ are represented with unit magnitude, preserving the angle but not the magnitude of the distance from the FoE (see Figure 2A), then the same properties are evident, but the inner product scales linearly rather than as a square of receptive field size, making the response more robust to noise. This was the template definition used in the ViSTARS model (Browning et al., 2009a, 2009b).

Since construction of the template is somewhat arbitrary, we can update the template so that rather than being defined as $E\boldsymbol{Y}$, it is defined as

$$\boldsymbol{L} = \frac{1}{N\boldsymbol{Y}}, \tag{2.3}$$

where $\boldsymbol{L}$ is the template and N is the number of elements in $\boldsymbol{Y}$ (see Figure 2C). The singularity defined by the list of vectors in $\boldsymbol{L}$ still corresponds to the heading, but now the magnitude of vectors in the template decreases, independently in each dimension, as a function of distance from the FoE and is normalized by the number of elements in the template. We verified that a template model defined in this way has a maximum activation in response to the heading coincident with the singularity of the template in the same way as a unit or veridical template. The inner product ($\boldsymbol{M}$) between the template in equation 2.3 and a noiseless motion pattern (E$\boldsymbol{Y}$) toward a planar surface is

$$\boldsymbol{M} = \boldsymbol{L}E\boldsymbol{Y} = \frac{E}{N}\left(\frac{1}{\boldsymbol{Y}}\right)^T \boldsymbol{Y} = E. \tag{2.4}$$

In this case, the template match, $\boldsymbol{M}$, is exactly equal to the expansion rate of the planar surface and, by extension, provides a direct measure of TTC (see equation 1.1). Note that the inner product of $1/\boldsymbol{Y}$ and $\boldsymbol{Y}$ is equal to N. The template match calculates the mean of the component values (one component for each spatial location), each of which provides an estimate of the expansion of the surface. The analysis thereby generalizes to angled and nonplanar surfaces, which do not have constant TTC, by providing a mean estimate of the expansion of the object. The number of elements, N, defines how robust the expansion estimate is to noise. Equation 2.4 demonstrates that a template model can be arbitrarily robust to any randomly distributed noise in the motion estimates by manipulating the number of motion vectors included in the template match.

This analysis demonstrates that large receptive field template models of MSTd can be constructed to provide accurate expansion/TTC estimates without decomposing the optic flow field into component parts. The response of such model cells is independent of receptive field size and the number of elements in the template.

However, equation 2.5 requires that there be a motion estimate at every spatial location represented by the template. In practice, when processing natural (and even most unnatural) stimuli, image-based motion estimates will be sparse across the input space. In order to make equation 2.4 independent of the distribution of motion estimates, N must be redefined as the number of nonzero components in the motion estimate,

$$N = \sum(\boldsymbol{V} \neq 0), \tag{2.5}$$

where $\boldsymbol{V}$ is the motion estimate across the visual field. With this definition we define a template model MSTd cell as

$$\boldsymbol{M} = \boldsymbol{L}\boldsymbol{V} \tag{2.6}$$

where $\boldsymbol{L}$ is the template defined in equation 2.3 using the definition of N given in equation 2.5 and $\boldsymbol{V}$ is the list of N-D motion vectors derived from visual input that can be either dense or sparse in space. The activation of $\boldsymbol{M}$ is exactly equal to the expansion rate (E) in noiseless environments and provides a mean estimate across all component estimates in $\boldsymbol{V}$. From inspection, assuming noise is zero mean and N is large, M is highly robust to noise and provides a veridical TTC estimate regardless of receptive field size or the distribution of image motion vectors.

3 Integration of Time-to-Contact into the ViSTARS Model

We integrated the updated template into a difference-equation version of the ViSTARS model (dViSTARS) to demonstrate that a distributed representation of motion, such as that found in primate V1 and MT, could support the estimation TTC in MSTd. The dViSTARS model is based on the dynamical systems models described in Browning et al. (2009a, 2009b) and Layton et al. (2012).

3.1 The dViSTARS Model. Input to dViSTARS comes from a 2D array describing grayscale values in the scene; values in the array are normalized between 0 and 1. All dViSTARS variables are defined across the 2D spatial locations represented in the input. For notational clarity, we do not include indices for spatial location except where it affects the computation. Model retina converts the input into an on and an off channel and assesses the change since the last array was presented:

$$p_{on}(t) = I(t), \quad p_{off}(t) = 1 - I(t), \tag{3.1}$$

$$z(t) = p(t) - y(t-1), \tag{3.2}$$

$$y(t) = (1-\alpha)p(t) + \alpha y(t-1), \tag{3.3}$$

where I is the input array to the system, p represents the array for the on and off channels, z is a model transient response cell, and y accumulates the input signal at any given position (initialized at 0). α is a parameter defining the rate of accumulation. Variables z and y are calculated for both on and off channels. Figure 3(top) shows the temporal response curve of model retina cells.

Directionally selective cells in model V1 are defined using the same mechanisms in ViSTARS and as described in Chey, Grossberg, and Mingolla (1998). Note that indices are dropped for clarity when all variables share the same values:

$$c^d_{xy} = ([z_{xy} - \beta z_{XY}]^+)^2, \tag{3.4}$$

$$e^d = ([c^d - c^D]^+)^2, \tag{3.5}$$

$$f = \frac{e_{on} + e_{off}}{2}, \tag{3.6}$$

where c is a directionally selective interneuron with direction preference d, z is defined in equation 3.2, β is a parameter defining the magnitude of difference required to signal motion, xy indicates the spatial position in the array, XY indicates the spatial position offset by 1 in the direction d, and $[]^+$ denotes half-wave rectification $(\max(x,0))$. Variable e is a directionally selective cell, D denotes the opposite direction preference, and f is the combined directional response of the on and off channels. Note that directional selectivity is computed using a spatial comparison of the variable z. Variable z accumulates over time and decays according to equations 3.2 and 3.3. Motion in direction D at position xy is detected if there is no activation in z at position XY. Since z accumulates over time and decays slowly, it retains memory of the stimulus position in the recent past; this is a form of nulling

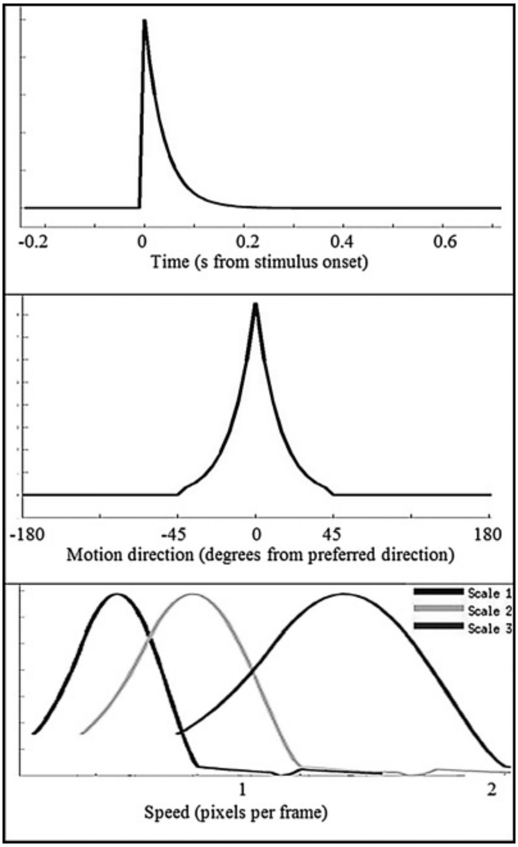

Figure 3: (Top) Model transient cells output a peak response immediately following stimulus onset that decays back to baseline after around 150 ms using the parameters described in section 3.2. (Middle) Model V1 cells are sharply tuned for motion direction. (Bottom) Model V1 cell speed tuning is a function of scale; larger scales prefer faster speeds and have a broader speed tuning. From left to right, the lines correspond to scales 1, 2, and 3, respectively. Note that model MT cells inherit their direction and speed tuning from V1.

inhibition. Figure 3 (middle and bottom) illustrates the direction and speed tuning curves obtained from these model V1 cells.

Model MT takes the motion vector estimation in equation 3.6 as input, and at each spatial location, the direction with maximum activation stays active and all other directions are suppressed:

$$h^d = \begin{cases} f^d & \text{for } d = \operatorname{argmax} f^d \\ 0 & \text{all other } d \end{cases}. \tag{3.7}$$

A bilinear resizing operation on the array reduces the size of the array (increases the receptive field size of each location) by a factor of 4 in both x and y dimensions, reducing the area (increasing the receptive field) by a factor of 16. This array reduction is a simplification of a large spatial filter that pools motion information across space. A temporal filter is applied to smooth the activations,

$$k^d(t) = (1 - \gamma_{MT})h^d(t) + \gamma_{MT}k^d(t-1), \tag{3.8}$$

where γ_{MT} is a temporal smoothing parameter. Finally, a spatial nearest-neighbor non-max suppression is applied to find the dominant direction activation at each spatial location:

$$V^d = \begin{cases} k^d & \text{for } d = \operatorname{argmax} k^d \\ 0 & \text{all other } d \end{cases} \tag{3.9}$$

Model MSTd templates (L) are constructed by equations 2.3 and 2.5, and the template match (M) is defined by equation 3.1, where $\boldsymbol{V}$ is defined by equation 3.9. MSTd activation is temporally filtered by

$$R(t) = (1 - \gamma_{MST})\boldsymbol{M}(t) + \gamma_{MST}R(t-1), \tag{3.10}$$

where γ_{MST} is a temporal smoothing parameter. Finally, the output of model MSTd is multiplied by a scale factor σ to bring the cell activation into an absolute value range consistent with the estimation of expansion.

3.2 Simulation Performance. A 256 × 256 pixel random dot input was generated simulating an approach to a frontal plane from 10 s time to contact to 1 s time to contact with a frame rate of 100 fps. Random dots were generated using the Matlab *rand()* function with 1% of the pixels given a value of 1. All other pixels had a value of zero. To ensure sufficient dots in the image projection toward the end of the simulation, the center 65 × 65 pixels were replaced with a random dot array where 10% of the pixels had a value of 1. To provide a realistic image projection, we conceptualized the distance of the frontal plane at the start of the simulation as 10 m, the velocity of the camera at 1 m/s, and the focal length of the camera to 0.1 m. These units and values are arbitrary, provided they maintain the same ratios.

The model was configured with $\alpha = 0.775$, $\beta = 1.5$, $\gamma_{MT} = 0.9$, $\gamma_{MST} = 0.99$. Three scales were implemented: scale 1 processed the stimulus at its native resolution, scale 2 processed the stimulus with x and y dimensions reduced to 0.75 that of the original stimulus, and scale 3 processed the stimulus with x and y dimensions reduced to 0.5 that of the original stimulus, with $\sigma_1 = 3000$, $\sigma_2 = 5000$, $\sigma_3 = 14000$. Larger scales correspond to larger cell receptive field sizes in model V1 and MT. The receptive field sizes of

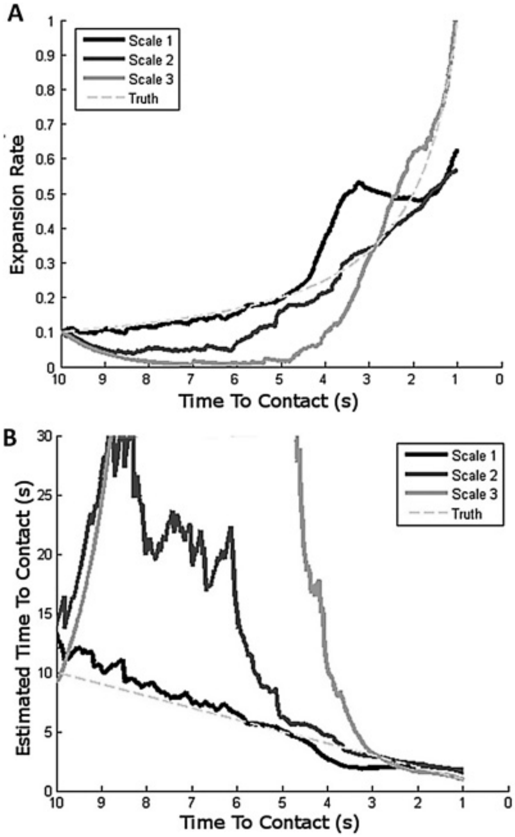

Figure 4: (A) Model MSTd outputs are shown for each of the three implemented scales for a single full field random dot sequence. The stimulus started with a TTC of 10 s and moved at constant speed toward the camera until it had 1 s TTC. MSTd outputs approximate the expansion associated with each TTC; the true expansion is shown by the dashed line. Scale 1, which has a preference for the slowest speed, has better approximations for low expansion rates, Scale 2 has better approximations for medium expansion rates, and scale 3 codes better approximations for high expansion rates. (B) Model MSTd outputs can be converted into TTC through a $1/E$ operation. Results are shown for all three scales. The true TTC is shown by the dashed line. As with the expansion estimates, each scale has a domain in which it produces a better approximation than the other scales.

model MSTd cells were the same for all scales and covered the whole input space.

Figure 4A shows MSTd outputs for each scale in response to the stimulus. No single scale has an output that accurately codes the expansion rate of the

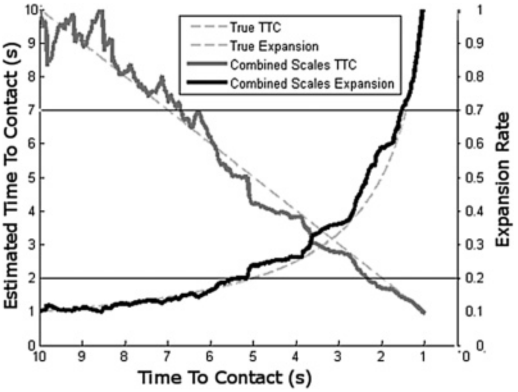

Figure 5: Population of MSTd cells codes expansion/TTC accurately across the full range of TTCs tested. Horizontal lines show the regions in which each scale was dominant. Scale 1 was dominant below expansion rates of 0.2, scale 2 was dominant between 0.2 and 0.7, and scale 3 was dominant above 0.7. True expansion and TTC are shown by dashed lines.

stimulus. However, each scale appears to provide a good approximation of expansion for some range of values. Figure 4B shows how MSTd outputs relate to an estimate of TTC.

In order to demonstrate that the population of MSTd cells represents expansion/TTC accurately, we replaced the scale-independent temporal accumulation in model MSTd with a temporal filter that combined the outputs of each scale. If the output from scale 3 was above 0.7, scale 3 was input to the temporal filter; if scale 3 was below 0.7 and scale 2 was above 0.2, scale 2 was input to the temporal filter; otherwise scale 1 was input to the temporal filter. These cut-off values were chosen empirically to provide a reasonable TTC estimate. The motivation here is to demonstrate that the population accurately codes TTC. It is not intended to reflect how the population response is decoded in the primate brain. Figure 5 shows how this combined estimate produces an accurate expansion/TTC estimate.

Simulations were repeated multiple times with different random dot stimuli, and results were qualitatively the same in each case. To simulate different-sized objects, we repeated the analysis but removed stimulus components toward the periphery of the input space; only the center 65 × 65 pixels contained dots at the start of the stimulus stream. Smaller cell receptive fields were simulated by zeroing out 20 pixels from each edge of the stimulus. For both smaller objects and smaller receptive fields, results were qualitatively the same as are shown above but generally displayed more noise in the MSTd outputs at small expansion rates (higher TTCs). We also tested the model using stimuli based on geometric shapes and on one video

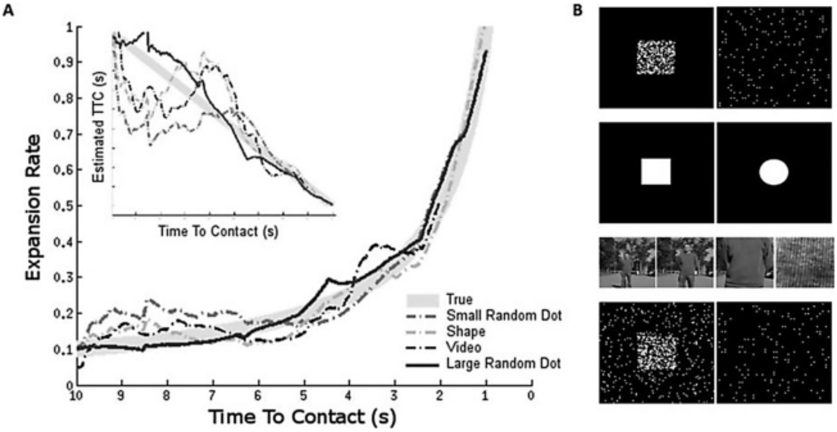

Figure 6: The model was tested with (B) 1000 large, random dot sequences—bottom row: left, first frame; right, last frame; 1000 small, random dot sequences—top row: left, first frame; right, last frame; 2000 geometric shape sequences—second row: left, example of square; right: example of circle. Shapes were random orientations, and sizes of squares, circles, and compound shapes made up of two triangles and a video sequence. Third row: frames 1, 200, 400, and 600 are shown. Model parameters were held the same for all simulations except for the video sequence where $\sigma_1 = 2000$, $\sigma_2 = 5000$, $\sigma_3 = 5500$ to account for differences in the frame rate and average pixel intensity of the natural versus generated sequences. Mean results for all four sets of data are shown in panel A. Expansion is shown in the main figure, and TTC is shown in the inset. Model performance was robust to all of the tested stimuli.

of an approach to a human; our results in all cases were qualitatively the same. The geometric shapes were chosen to produce different variations of the aperture problem in V1: circles do not produce an aperture problem; squares and compound triangle shapes had oriented edges that were not orthogonal to the direction of motion and so introduced an aperture problem. Shapes were generated in Matlab to contain no texture in their interior so that the model motion estimates were forced to be sparse around the boundaries of the object. With the exception of the video, which had a different frame rate and pixel intensity distribution, the model parameters were exactly the same for all stimuli. The video was recorded at 60 fps using a ContourHD camera mounted on a John Deere Gator and manually driven away from a human (we then process the video backward). Accurate speed and position information were unavailable. True TTC was defined as the actual time between the current image frame and the camera touching the human. Results for all stimuli and the modified parameters used for the video are shown in Figure 6.

4 Discussion

The mathematical analysis and insights described in this letter come directly from the representation of time-to-contact as a function of expansion rate, equation 1.1. This representation clearly describes how expansion (looming/eta) relates to TTC (tau). Equation 2.3 demonstrates a template model of a heading-sensitive cell that responds proportionally to the expansion rate of the input stimulus. This elegant solution to the estimation of TTC indicates that a simple neural circuit, as documented in primate dorsal MST, is capable of accurately estimating TTC directly from an optic flow field while concurrently estimating heading. Provided that motion estimates are accurate, this method provides a robust way of incorporating multiple motion vectors into a single object-based measure of time to contact. The technique could be expanded further to differentially weight motion vectors through manipulation of the parameter N. For example, if N were constructed to incorporate a normalized confidence metric for each of the motion estimates, then the template match equation, 2.6, could perform a weighted mean of all of the component TTC estimates. This could help remove outliers and reduce the reliance on a large value of N to provide accurate results in noisy data. Temporal filtering of equation 2.6, as we did in dViSTARS, further reduces the effect of noise and relaxes the constraint on a large N for any given image frame provided N is large over some subset of image frames.

The primary limitation of our model is that it requires that objects be segmented from each other and the background. If they are not, then the resulting TTC estimate will be a mixture of the TTCs for all objects within the receptive field. Furthermore, the model assumes that objects are roughly planar. Any object that has a large, protruding element or highly irregular surface will produce a biased TTC estimate. These limitations are shared by any method that attempts to estimate a single TTC value for an object of arbitrary 3D shape. Note that in our video example, we did not explicitly segment the object, but the background was sufficiently far away to introduce few, if any, motion signals.

Section 3 shows that when motion direction and speed are represented across a population of units, as it is in primate V1 and MT, the template model accurately represents TTC across the population of templates. Based on these results, we claim that our proposed template definition is sufficient for accurate TTC estimation from a distributed representation of motion, such as that found in primate V1 and MT.

In our analysis, we combined the outputs of different scales after the template match. This is not the optimal method of scale integration and is not sufficient for the robust estimation of TTC across large slanted planar surfaces or highly nonplanar objects. In general, this method will tend toward the smallest TTC (largest expansion rate) component of a given object. In any expanding object, the motion vectors closer to the FoE are smaller than the motion vectors far from the FoE. Scale integration before the template match would allow for this distribution of speeds, all corresponding to the same TTC, to be captured by a single template match. This should improve reliability by increasing the number of valid motion estimates in the template match. Whether or how this may be implemented in the brain is unknown. For computational applications, this is irrelevant since speed is represented by the magnitude of a vector rather than by a population code, and as a result, no combination of scales is required.

Neurophysiological data from MST cells in response to stimuli designed to elicit a TTC response are inconclusive. The analysis shown here may provide insights into how to process the neural data to find TTC responses across a population of cells. In primates, speed is represented across a population of neurons. We show that TTC could be coded across a population of neurons. The experimenter should therefore look for cells that respond proportionally to a small range of TTC values and demonstrate that the population codes a behaviorally beneficial range of TTC values. We predict that TTC is coded in MSTd and is represented by expansion rather than time. However, analyzing individual neurons over small ranges of TTC and neuronal populations over large ranges of TTC will allow for investigation of TTC in any brain area. Moreover, the work described here makes a specific prediction that TTC and heading estimation are performed by the same circuits. Human and primate researchers could investigate this through stimuli designed to show that heading estimation is not affected by TTC and, by construction of contrived inputs, that TTC estimation, or cell activity, does not require a consistent FoE in the stimulus space.

In summary, TTC estimates from our proposed template model of MSTd are accurate regardless of the receptive field of the cell, the object size, or whether motion is coded as a vector or distributed across a population of cells.

Appendix: TTC/Expansion Equivalence

Let f be the focal length of the camera, X be the width of an object, Z is the distance between the lens (pinhole) and the object along the normal to the focal plane, at time t_1, and V is the distance traveled by the object toward the camera between t_1 and t_2. Let x_1 be the projected image size of the object at time t_1 and x_2 be the projected image size of the object at time t_2 (see Figure 1). Let the unit of time be $dt = t_2 - t_1$. Then TTC at time t_1 is

$$T_1 = \frac{Z}{V}. \tag{A.1}$$

Therefore,

$$Z = T_1 V. \tag{A.2}$$

From the pinhole camera model,

$$x_1 = f\frac{X}{Z}, \quad x_2 = f\frac{X}{Z - V}. \tag{A.3}$$

Let expansion rate,

$$E = \frac{x_2}{x_1} - 1. \tag{A.4}$$

Substituting from equation A.3, then equation A.2, and simplifying,

$$E = -\frac{Z}{V - Z} - 1 = \frac{1}{T_1 - 1}. \tag{A.5}$$

In general, E is measurable from the image plane and T (in standard units) is what we want to know:

$$T_1 = \left(1 + \frac{1}{E}\right) dt, \tag{A.6}$$

$$T_2 = \frac{1}{E} dt, \tag{A.7}$$

where T_2 is the TTC at time t_2. Note that if we drop the dt, time to contact is defined in units of dt rather than in standard units. Also note that we could incorporate the dt directly in the definition of E.

Acknowledgments

I thank the three anonymous reviewers for their thoughtful and helpful feedback. This work has been supported in part by the Office of Naval Research (ONR N00014-11-1-0535).

References

Beintema, J. A., & van den Berg, A. V. (1998). Heading detection using motion templates and eye velocity gain fields. *Vision Research*, *38*(14), 2155–2179. doi:10.1016/S0042-6989(97)00428-8

Browning, N. A., Grossberg, S., & Mingolla, E. (2009a). A neural model of how the brain computes heading from optic flow in realistic scenes. *Cognitive Psychology*, *59*(4), 320–356.

Browning, N. A., Grossberg, S., & Mingolla, E. (2009b). Cortical dynamics of navigation and steering in natural scenes: Motion-based object segmentation, heading, and obstacle avoidance. *Neural Networks*, *22*(10), 1383–1398.

Byrne, J., & Taylor, C. J. (2009). Expansion segmentation for visual collision detection and estimation. In *Proceedings of the IEEE International Conference on Robotics and Automation, 2009* (pp. 875–882). Piscataway, NJ: IEEE.

Chey, J., Grossberg, S., & Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research*, *38*(18), 2769–2786.

Duffy, C. J., & Wurtz, R. H. (1991a). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, *65*(6), 1329–1345.

Duffy, C. J., & Wurtz, R. H. (1991b). Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *Journal of Neurophysiology*, *65*(6), 1346–1359.

Elder, D. M., Grossberg, S., & Mingolla, E. (2009). A neural model of visually guided steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1501.

Gibson, J. (1955). The optical expansion-pattern in aerial locomotion. *American Journal of Psychology*, *68*(3), 480–484.

Grossberg, S., Mingolla, E., & Pack, C. (1999). A neural model of motion processing and visual navigation by cortical area MST. *Cerebral Cortex*, *9*, 878–895.

Hayes, W. N., & Saiff, E. I. (1967). Visual alarm reactions in turtles. *Animal Behaviour*, *15*(1), 102–106.

Horn, B. K., Fang, Y., & Masaki, I. (2007). Time to contact relative to a planar surface. In *Proceedings of the 2007 IEEE Intelligent Vehicles Symposium* (pp. 68–74). Piscataway, NJ: IEEE. doi:10.1109/IVS.2007.4290093

Hoyle, F. (1957). *The black cloud*. London: Penguin.

Judge, S., & Rind, F. (1997). The locust DCMD, a movement-detecting neurone tightly tuned to collision trajectories. *Journal of Experimental Biology*, *200*(16), 2209–2216.

Koenderink, J. J., & van Doorn, A. J. (1976). Local structure of movement parallax of the plane. *Journal of the Optical Society of America*, *66*(7), 717–723. doi:10.1364/JOSA.66.000717

Lappe, M. (2004). Building blocks for time-to-contact estimation by the brain. *Advances in Psychology*, *135*, 39–52.

Layton, O., Mingolla, E., & Browning, N. A. (2012). A motion pooling model of visually guided navigation explains human behavior in the presence of independently moving objects. *Journal of Vision*, *12*(1), 1–19. doi:10.1167/12.1.20

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception*, *5*(4), 437–459. doi:10.1068/p050437

Longuet-Higgins, H. C., & Prazdny, K. (1980). The Interpretation of a moving retinal image. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *208*(1173), 385–397. doi:10.1098/rspb.1980.0057

Nakagawa, H., & Hongjian, K. (2010). Collision-sensitive neurons in the optic tectum of the bullfrog, *Rana catesbeiana*. *J. Neurophysiol.*, *104*, 2487–2499.

Perrone, J. A. (1992). Model for the computation of self-motion in biological systems. *Journal of the Optical Society of America, A*, *9*, 177–194.

Perrone, J. A., & Stone, L. S. (1994). A model of self-motion estimation within primate extrastriate visual cortex. *Vision Research*, *34*(21), 2917–2938.

Perrone, J. A., & Stone, L. S. (1998). Emulating the visual receptive-field properties of MST neurons with a template model of heading estimation. *Journal of Neuroscience*, *18*(15), 5958–5975.

Raiguel, S., Van Hulle, M. M., Xiao, D., Marcar, V. L., Lagae, L., & Orban, G. A. (1997). Size and shape of receptive fields in the medial superior temporal area (MST) of the macaque. *Neuroreport*, *8*(12), 2803–2808.

Schiff, W., Caviness, J., & Gibson . (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of "looming." *Science*, *136*, 982–983.

Tanaka, K., Fukada, Y., & Saito, H. (1989). Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.*, *62*, 642–656.

Wang, Y., & Frost, B. J. (1992). Time to collision is signalled by neurons in the nucleus rotundus of pigeons. *Nature*, *356*(6366), 236–238.

Received December 28, 2011; accepted May 5, 2012.