Taking the energy out of spatio-temporal energy models of human motion processing: The Component Level Feature Model

Linda Bowns

Nottingham Visual Neuroscience, School of Psychology, The University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

1. Introduction

Humans are able to detect motion under practically any lighting conditions and use it to compute not only the movement of objects but to recover three dimensional surfaces, edges, shape, determine specific biological entities, and help to control locomotion. This information is reconstructed from the two dimensional motion projected onto the retinas. The standard basic spatio-temporal energy model for computing this two dimensional motion often has two critical stages. In the first stage, suitable filters are convolved with the pattern over time to extract information at the “component” level. Motion energy is then computed for each component. The second stage typically computes pattern velocity using the intersection of constraints rule (IOC). This paper describes a new implementation of the Component Level Feature Model (Bowns, 2002) that computes motion direction that is similar to these two stages except that it does not compute motion energy. Here the model computes direction for 200 randomly generated plaids. The output linearly matched that predicted by the IOC. The model was also able to predict the perceived direction even when it deviated from the IOC due to the following variables – speed ratio (Bowns, 1996); duration (Yo & Wilson, 1992); adaptation (Bowns & Alais, 2006). The model provides a novel explanation for each of the above and for why multiple directions can be represented for the same stimuli (Bowns & Holroyd, 1992) reverses perceived motion direction. Finally, CLFM is invariant to contrast and phase.
The following is a brief outline of some of the problems associated with a simple spatio-temporal energy model outlined above. One problem is that contrast is confounded with velocity, i.e. a high contrast moving pattern would produce more motion energy than a low contrast moving pattern. To solve this problem the use of ratios across stationary and opposing motion has been suggested (Adelson & Bergen, 1985). A second problem is that although there are neurons that have oriented Gabor filter characteristics that are narrowly tuned for a range of spatial frequencies, there is less evidence for a similar range of narrowly tuned temporal frequency filters (Perrone, 2004; Snowden & Hess, 1992). An alternative model that shares some similarities with early stages to standard energy models has solved this using a very different type of early V1 sensor that is tightly tuned for speed (Perrone & Krauzlis, 2008); also drawing attention to another problem for energy models – evidence for speed tuning found in V1 and MT (Perrone & Thiele, 2001; Pribe, Lisberger, & Movshon, 2006). Furthermore, motion energy introduces sufficient directional noise to explain some quite large perceived deviations from veridical or predicted motion, (Weiss, Simoncelli, & Adelson, 2002). There is also mounting evidence that perceived motion is affected by specific spatial frequencies not present in the amplitude spectrum of a moving pattern. These “second-order” components can only be extracted by taking account of higher order interactions among the components. A simple motion energy model of motion cannot account for these, however, additional mechanisms that could be added to the standard model have been suggested (Bowns, 2002; Chubb et al., 1994; Simoncelli & Heeger, 1998; Wilson, Ferrera, & Yo, 1992). The problems outlined here have solutions but each add more levels of complexity to the simple computation of motion energy.

The following model describes a possible alternative that takes advantage of the ample evidence for component level processing, and the use of the IOC. It is a method for computing the IOC that is relatively simple but addresses the above problems integral to the model. The idea of the model was first published by (Bowns, 2002).

2. The Component Level Feature Model of Motion (CLFM)

Here the CLFM has been elaborated upon and a new computationally explicit version has been simulated in MATLAB so that the direction of simple “plaid” patterns can be computed. The model will be described and results from the model are presented.

The basic simple idea is shown in Fig. 2. Two sinusoidal components at different orientations and the same spatial frequency are depicted on the left. The red lines indicate the mean values of each component, and the blue lines represent the direction in which they move given a positive phase shift on subsequent frames of the image to create component motion displacement. The mean values can be extracted by convolving the sinusoids with a $A^2C$ and then thresholding around the zero-crossing. The velocity space diagram shows how the vectors correspond to the phase shift magnitude and direction over time, and the lines of zero-crossings correspond to the constraint lines.

Fig. 3 depicts each operation of the CLFM (see Appendix A for the corresponding equations). The input stimulus shown in Fig. 3 comprises two simple sinusoidal components added together to form a plaid. The plaid is moved by shifting the underlying sinusoidal components through a specific phase shift in time as indicated by the arrows. The length of the arrows has been exaggerated for clarity (NB these are perpendicular to the orientation of the sinusoid, where the contrast is at its maximum). The movement of these underlying components causes the plaid to move in its unique veridical direction predicted by the intersection of constraints, in this case $337^\circ$. A bank of oriented Gabor filters are convolved with the stimulus at five different temporal intervals (1) (in this case the intervals corresponded to a phase shift). The temporal sampling at this point in the model is assumed to be a fixed interval. Provided the phase ratios of the components are fixed as a function of time, the point at which the moving pattern is sampled relative to the phase would not affect the outcome of the model. The bank of filters represents orientations around the clock separated by $15^\circ$, and spatial frequencies that span four octaves. So far this is similar to the spatial filtering of the standard model corresponding to physiological responses reported in area V1 of the visual cortex. The main difference is that the temporal frequency is not necessary or encoded at this stage. Examples of the filters that would respond maximally to the stimulus are shown together with the filter responses (2) and (3). Instead of extracting temporal frequency and computing motion energy at this point, the two filters that respond maximally are selected for further processing. As can be seen from Fig. 3, these correspond to the components used to construct the plaid as ex-

---

1 The type of filters and their characteristics are not critical provided they extract appropriate contrast, spatial frequency and orientation information from the image.
expected. These two filter response outputs are convolved with a $\Delta G$ filter and thresholded zero-crossings are extracted (4) and (5). These thresholded zero-crossings are a convenient way to extract the thresholded mean values of a sinusoidal response function, so this convolution may not be biologically explicit provided neurons can respond to the mean firing rate of the Gabor filter responses. Similarly, to Fig. 2, the red lines superimposed upon the zero-crossings have the same orientation as the constraint lines in the intersection of constraints rule, and the blue arrows indicate the direction and phase shift of the zero-crossing lines. Thus extracting any intersections (6) and tracking them over time (7) will correspond to velocity as predicted by the IOC. The intersections and how they accumulate over time are also shown at the lower left of Fig. 3. Interestingly these accumulated zero-crossings appear similar to “motion streaks”, the orientation of which corresponds to the direction of the plaid, and their length corresponds to the displacement of the plaid. Rather than ‘motion streaks’ a more accurate description might be ‘displacement lines’. A convenient way to determine direction is to apply a Hough transform to the accumulated zero-crossings (8). A Hough transform reveals the dominant line that the points go through where theta is the line and rho is the distance between the line and the origin. The output from the Hough transform shows the first few peaks in the Hough transform and indicates a clear orientation (9). This is the direction predicted by the intersection of constraints rule up to a rotation. A biological system could use the topological arrangements (a type of grid arrangement) of the neurons that respond to the mean firing rates of individual Gabor response outputs to achieve this, and to determine the quadrant direction in which they are accumulating, and thereby determine the unique direction. Neurons responding to the grid would be tuned for responses that occur along specific orientations within the grid of responses; the extent or length of the aligned spatial distribution of responses – ‘displacement-lines’. Although not depicted in Fig. 3, the model fits lines to the accumulated zero-crossings to determine their length explicitly.

The above describes how the model determines motion direction and motion displacement and is sufficient for testing the model for its capacity to predict perceived motion direction. However, the following offers several suggestions as to how speed can be obtained. It is clear that speed can be computed when the length of the displacement-lines are divided by the time taken to create them (10). This is where motion energy and the CLFM look quite different. Division however does not seem biologically plausible. Another possibility is that the neurons responding to the ‘displacement lines’ are not only tuned for orientation and length but also temporal frequency. The length and temporal frequency would be coupled such that there would be an inverse relationship between temporal frequency and line length, i.e. the shorter the line the higher the temporal frequency. There is evidence supporting speed tuning in this context (Bowns, 2001; Reisbeck & Gegenfurtner, 1999) in addition to that described in the introduction. Such neurons could then provide a response tuned for velocity corresponding exactly to that predicted by the intersection of constraints rule. These temporal frequency narrowly tuned line detectors are assumed to be in area MT; there is ample evidence that neurons in this area respond to pattern direction, e.g. (Britten et al., 1992; Dubner & Zeki, 1971). One problem for CLFM is the issue of why V1 neurons have any temporal frequency tuning. It is possible that the tuning of V1 neurones maybe more related to temporal sampling than computing spatio-temporal energy per se. This remains an unexplained problem for CLFM, and one of the reasons why in the earlier non-simulated model CLFM was combined with spatio-temporal energy. This of course may still be a good possibility.

3. Experiment 1: to test the capacity of the CLFM to compute the IOC direction

The orientation of each of the two components was randomly generated for 200 plaids. The spatial frequency was set at 4 cpi, contrast at 100%, and size 100 × 100 pixels. There were five frames comprising the movie, the first frame was stationary, and both components were shifted through a phase of 40° on subsequent frames. There were no free parameters in the model, and the whole process was fully automated.

The results are shown in Fig. 4. The model output direction is plotted against the direction predicted by the intersection of constraints rule. The plaids used in this simulation were all Type 1 plaids where the intersection of constraints rule predicts a similar result to that of the vector average. The results are linear indicating the reasons why in the earlier non-simulated model CLFM was combined with spatio-temporal energy. This of course may still be a good possibility.
that the CLFM has accurately computed the IOC direction. There are two outliers, these are where the orientations have an insufficient angle difference to compute the IOC, and therefore would not be perceived in the IOC direction.

4. Experiment 2: to test the idiosyncratic perception of Type II plaids at three different speed ratios

Type II plaids are defined as plaids where the IOC direction falls to one side of the components in velocity space (Wilson, Ferrera, & Yo, 1992). Type II plaids have the advantage of separating out predictions from the vector average and the IOC. It has been shown that the speed ratio of Type II plaids determines whether or not the plaid is perceived to move in the IOC direction or closer to the vector average direction (Bowns, 1996; Bowns & Alais, 2006). The three plaids used in the Bowns and Alais study were examined because they represent the main affects of speed ratio on perceived direction. The first was a Type I plaid and had a speed ratio of 1:1; the second was a Type II plaid with a speed ratio of 1:0.75 and is perceived closer to the vector average at short durations; the third was a Type II plaid with speed ratio 1:0.45 and is perceived in the IOC at short durations. CLFM output was computed exactly as described previously. Fig. 5a shows a bar graph comparing the perceived direction (Bowns & Alais, 2006) with the CLFM output, the IOC, and the vector average for each of the three plaids. The first two peaks in the Hough transform represent the direction plotted. The CLFM output more accurately predicts the perceived direction than the IOC or the vector average.

5. Experiment 3: to test the model to see if it predicts a shift from vector average to IOC as a function of duration

Some Type II plaids are initially perceived near to the direction predicted by the vector average, and over time perception of the plaid direction is shifted to the IOC direction (Bowns, 1996; Yo & Wilson, 1992). An example of this is the Type II plaid with speed ratio 1:0.75 described above. The plaids used had orientation 202° and 225°; a spatial frequency of 4 cpi; and 100% contrast. The phase was shifted to create the motion at a speed ratio of 1:0.75, 1 = 40°. The first frame has 0 phase and each subsequent frame was shifted by a phase shift of 40°. To test CLFM the output direction was computed over an increasing number of frames to mimic the increasing duration.

6. Experiment 4: to see if the model provides an explanation for why there appears to be multiple representations of direction for some Type II plaids

The IOC rule and the vector average are mutually exclusive hypotheses, so it was surprising to find that if the perceived direction of a plaid perceived in one or other of these directions is adapted out, it is then perceived in the other direction (Bowns & Alais, 2006). This suggests that multiple directions are represented corresponding to both the IOC and the vector average. It was therefore hypothesized that if the number of peaks in the Hough transform was increased to see other values, then they would occur in both the IOC and vector average for these plaids used in the Bowns
and Alais study. In addition, it was predicted that there would be more peaks in the direction perceived before adaptation.

Velocity space diagrams for three plaids are shown in the upper part of Fig. 6, these include: a Type I plaid where the IOC = vector average and perceived in the IOC/VA direction; a Type II plaid perceived in the vector average direction, and a Type II plaid perceived in the IOC direction. In the lower part of Fig. 6 are the corresponding Hough transforms with the peaks indicated by small red squares. The first 20 peaks of the transform are shown revealing that most of the information is in the expected directions. With this number of peaks allowed it is not surprising that several extraneous peaks have emerged. For the Type I plaid there is a spread of peaks centered around the IOC/VA direction. Adapting out the IOC = VA direction would therefore have little effect, as reported, because there are nearby peaks available. The Type II plaid perceived in the vector average direction has most peaks in the vector average direction but also has some peaks in the IOC direction; so if the vector average was adapted out the IOC would be the next dominant direction. The Type II plaid perceived in the IOC direction on the other hand has the most peaks in the IOC direction but also has some peaks in the vector average direction, therefore if the IOC was adapted out the dominant direction would be the vector average. These results are consistent with those reported, and provide a novel explanation for the dramatic swings in direction reported.

7. Summary and discussion

A novel and explicit simulation of the CLFM has been described that is able to compute two dimensional pattern direction and displacement consistent with the intersection of constraints rule without computing motion energy. The model avoids the confound of motion energy and velocity because the information used to compute the intersection of constraints is invariant with respect to the amplitude of contrast because it uses the mean response output. In doing so, this also avoids much of the noise associated with motion vectors derived from motion energy. There will of course still be some noise arising from the computation of the thresholded zero-crossings and their intersections, but as can be seen from the results of Experiment 1 the IOC direction is computed accurately with minimal noise. Using the mean response also provides an alternative explanation for a number of second-order results that implicate spatial frequencies not present in the Fourier spectrum. The lines of zero-crossings have double the frequency of the underlying components, thus frequency doubling is an inherent part of the model. CLFM simulations of plaids where the direction becomes ambiguous or reverses show similar behavior. The reason is because the intersecting zero-crossings for these plaids have nearest neighbor matches that are similar for both directions or are nearer for the reversed direction (Bowns, 2001). The property of being invariant to contrast is of course very beneficial for computing motion – the idea that speed or direction changes when the contrast changes is clearly not acceptable. However, there are a number of studies showing that contrast can affect perceived direction and speed for some types of plaids – see Champion, Hammett, and Thompson (2007) for a clear review and update of this literature. They argue that the best explanation of the data is based on the movement of different features in the two dimensional pattern as described by Bowns (1996) and therefore may involve higher level features than those described here. Currently, CLFM uses only the intersections to extract the IOC information, but it is possible that a parallel system could make more use of the zero-crossing lines extracted in the earlier stages that also correspond to these features rather than extracting them from the two dimensional pattern. The CLFM was partly developed to avoid such two-dimensional feature tracking mechanisms that are computationally very complex. The author is also currently exploring possible filter characteristics that may account for these effects of contrast, e.g. the range of spatial frequencies used.

Results from Experiment 2 show that the output from CLFM deviates from the IOC when perceived direction also deviates, i.e. when the speed ratio of components varies. Furthermore, it is more effective at predicting this than the IOC or the vector average. Experiment 3 shows how CLFM shifts from the vector average to the IOC with increasing duration in a similar way to that reported for observers. It also provides a novel explanation for why we can represent multiple directions and why one dominates over the other. Although investigators have tried to incorporate some of these results it has meant quite complex additions to the simple
energy model outlined in the introduction, and even then they are
not always able to predict the exact nature of the above results.

Appendix A

\[ F_{O1}(p,t^n) = \max\{F_{O1}(p)\} \]

\[ F_{O2}(p,t^n) = \max\{F_{O2}(p) - \Delta F\} \]

\[ Z_1(x^n, y^n, t^n) = 0 \pm 0.001(FO_1(x^n)) \times \nabla^2 G \]

\[ Z_2(x^n, y^n, t^n) = 0 \pm 0.001(FO_2(x^n)) \times \nabla^2 G \]

\[ Z_{int} = Z_{int}(t^n) \]

\[ Hough_{ints} = Hough(Z_{ints}) \]

\[ Direction = \text{mode}(\text{peaks}(Hough_{ints})) \]

\[ Speed(t^n) = \text{length}(\text{Adjacent}(\text{mode}(\text{peaks}(Hough_{ints}))))/t^n \]

\[ \theta^n = [0, 15, 30, 45, 60, 75, 90, 105, 120, 135, 150, 165] \]

\[ \theta^n_{cycles/image} = [1, 2, 4, 8] \]

\[ t^n = [2, 3, 4] \]

References


