Reprinted from Nature, Vol. 333, No. 6168, pp. 71-74, 5 May 1988

A computational theory for the perception of coherent visual motion

Alan L. Yuille* & Norberto M. Grzywacz†‡

* Harvard University Division of Applied Sciences, G12e Pierce Hall, Cambridge, Massachusetts 02138, USA
† Center for Biological Information Processing, Massachusetts Institute of Technology, E25-201, Cambridge, Massachusetts 02139, USA

When we see motion, our perception of how one image feature moves depends on the behaviour of other features nearby. In particular, the Gestaltists proposed the law of shared common fate[1,2], in which features tend to be perceived as moving together, that is, coherently. Recent psychophysical findings, such as the cooperativity of the motion system[3,4] and motion capture[5,6], support this law. Computationally, coherence is a sensible assumption, because if two features are close then they probably belong to the same object and thus tend to move together. Moreover, the measurement of local motion may be inaccurate[7] and so the integration of motion information over large areas may help to improve the performance. Present theories of visual motion, however, do not account fully for these coherent motion percepts. We propose here a theory that does account for these phenomena and also provides a solution to the aperture problem[8], where the local information in the image flow is insufficient to specify the motion uniquely.

The theory divides the computation of motion into two stages which we refer to as the measuring and the smoothing stages. The measuring stage estimates or restricts the velocity field from the image information. In humans, two processes have been suggested to mediate this stage, one that deals with short-range and another that deals with long-range motions[9]. Several schemes have been proposed to measure short-range motion[8,10-13] and any of them may be used as the theory's measuring stage. For long-range motions the correspondence problem arises[14]. Different types of constraint are possible to solve this problem[14,15], and each of them predicts different velocity fields. At this stage, our theory merely requires that features correspond, and the exact correspondence is unspecified.

In the smoothing stage, a velocity field is constructed over the entire visual field, even where no estimates of motion have been made. This velocity field is constrained by the restrictions found in the measuring stage and simultaneously is biased to be as smooth as possible. We call this theory the motion coherence theory. (In its present form, the theory does not incorporate mechanisms for motion transparency[16], motion segmentation[17] or motion inertia[18].)

We now present the theory formally, first for short-range motion. Let the velocity measurement at point $\vec{r}_i$ be $M(\vec{V}_i)$, where $\vec{V}_i$ is the true image velocity. For example, for contour motion, $M(\vec{V}_i)$ would correspond to the perpendicular component of the velocity field or to a rough approximation to it[8,19]. The theory proposes that the smoothing stage constructs a velocity field, $\vec{v}(\vec{r})$ that minimizes the following functional for both components of $\vec{v}$:

$$E(\vec{v}(\vec{r}), \vec{V}_i) = \sum_i (M(\vec{v}(\vec{r}_i)) - M(\vec{V}_i))^2 + \lambda \int \sum_{m=0}^{\infty} c_m |\nabla^m \vec{v}|^2 \qquad (1)$$

where $\lambda \geqslant 0$ and $c_m \geqslant 0$ are constants, and ∇^m is the $(m/2)$th power of the laplacian operator if m is even, and the gradient of the $((m-1)/2)$th power of the laplacian operator if m is odd. The sum is taken over all points where motion information is available. Minimization of the first term of the equation's right-hand side forces the computed velocity field, $\vec{v}$, to respect the measuring stage. On the other hand, the second term forces field smoothness[10,19,20]. This term is one of the Tikhonov stabilizers, a class of functionals commonly used in computational vision[21].

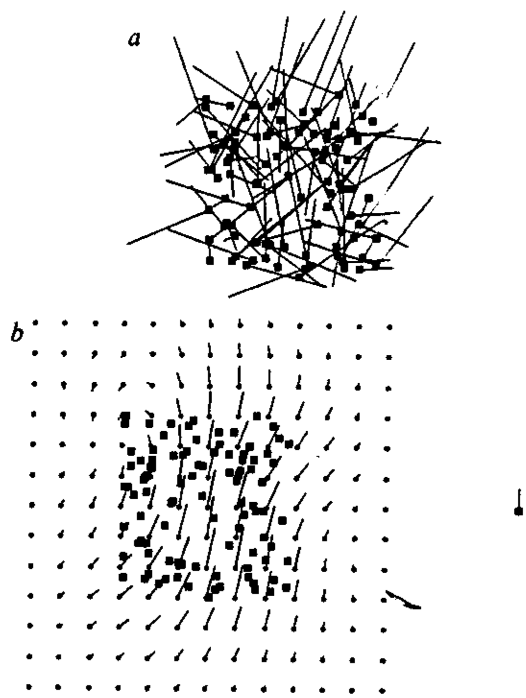

Fig. 1 Motion cooperativity. *a*, One hundred dots were randomly placed inside a square region of unit sides; the dot positions are indicated by the little solid squares. Each of these dots was assumed to be moving. The horizontal components of their velocity were obtained from a homogeneous random distribution whose values ranged from −1 to 1. The vertical components had the same distribution plus a constant bias upwards of 0.25. The velocity of each dot is shown by the lines attached to each little square. *b*, The velocity field computed by the motion coherence theory by solving equation (1) with $\lambda = 2.5$ and $\sigma = 0.6$. The computed field is shown as lines attached to little circles positioned in a grid and the position of the moving dots are also shown. (The solution is continuous, however, because it is computed analytically as in equation (3); the grid is shown for purposes of illustration.) The computed field in the region of the moving dots is more coherent than that of the input. Also, the velocity of the computed field is roughly the same as the bias given to the input data. For comparison, an isolated dot moving by the side of the field indicates the bias. This output coherence is similar to the psychophysical phenomenon of motion cooperativity[3,4]. Another desirable property of the field is that its speed decreases with distance from the moving dots, and thus objects that are not close are guaranteed to be perceived as different objects. The theory in its present form fails for close objects, however, as it does not have any mechanism for detecting motion discontinuities.

For long-range motion, any choice of feature correspondence determines a measured velocity field. To this field corresponds a new field, $\vec{v}$, for which a functional such as in equation (1) is minimum. Our theory searches among all possible correspondences the one which yields the deepest of these minima. (Not all features are necessarily matched, however, because psychophysics shows that preference may be given to those at large spatial scale[22]. At least, this preference suggests that a coarse-to-fine heuristic may be used to avoid the combinatorial explosion of considering all possible matches.) Suppose there are a number of points $\vec{r}_a$ at time t and $\vec{r}_i$ at time $t + \delta t$. A matching matrix V_{ai} is defined, so that $V_{ai} = 1$, if the ath point is matched to the ith point, and $V_{ai} = 0$ otherwise. If $V_{ai} = 1$, then the velocity at $\vec{r}_a$ is $\vec{v}(\vec{r}_a) = (\vec{r}_i - \vec{r}_a)/\delta t$. We now minimize the following energy function over the velocity field and all possible matchings that satisfy the cover principle[14].

‡ To whom correspondence should be addressed.

$$E(\vec{v}(\vec{r}),\ V_{ai})=\sum_{ai} V_{ai}\left(\vec{v}(\vec{r}_a)-\frac{(\vec{r}-\vec{r}_a)}{\delta t}\right)^2 +\lambda\int \sum_{m=0}^{\infty} c_m\,|\nabla^m \vec{v}|^2 \qquad (2)$$

Minimization of the first term of the equation's right-hand side forces compatibility with the data, while the second term forces smoothness. Ullman's minimal mapping theory[14] is an approximation of the motion coherence theory for large λ (ref. 23).

There are many possible forms of smoothing, but two restrictions apply. To ensure that speed decreases with distance (Fig. 1), that is, that the interaction falls off at infinity, c_0 must be non-zero[23]. Also, c_m must be non-zero for some $m > 1$, so that solutions may exist for sparse data[24]. Previous theories that integrated motion over space[10,19] used only first-order derivatives and did not satisfy these criteria. The higher-order derivatives affect the rate of fall of the interaction at infinity. For this paper's simulations, we chose the smoothing such that the interaction is a gaussian. This form of interaction was chosen for four reasons: (1) It meets the criteria above; (2) it generates analytical solutions; (3) it has a natural spatial scale; and (4) it approximates several processes in vision[20]. Moreover, variations of the higher-order terms, c_m for $m > 2$, seems to have little effect on the interaction, which thus almost always resembles a gaussian[20]. This suggests that although analytical solutions may not be possible for other choices of smoothing, good approximations with lower terms are possible. Thus, these choices are probably not computationally prohibitive. To obtain a gaussian interaction we set $c_m = \sigma^{2m}/(m!2^m)$ (ref. 23). In this case, the solution of equation (1), obtained by standard calculus of variations, is

$$\vec{v}(\vec{r})=\sum_i \frac{\vec{\beta}_i}{2\pi\sigma^2}\exp\frac{-|\vec{r}-\vec{r}_i|^2}{2\sigma^2} \qquad (3)$$

where the $\vec{\beta}_i$ may be computed analytically. This solution is the superposition of gaussians centred in points where motion information is available.

This paper's first example (Fig. 1) shows that the motion coherence theory may account for motion cooperativity[3,4]. When dots move randomly, but slightly biased, a coherent motion percept in the bias direction occurs. The theory predicts this behaviour, because the smaller variance of the computed velocity field makes the mean motion more perceptible.

In another example, it is shown in Fig. 2 that the motion coherence theory provides a solution to the aperture problem[8], which can be considered a problem of coherent motion[16]. In this particular example, the problem is to find the motion of a contour, because local motion detectors can only detect motion

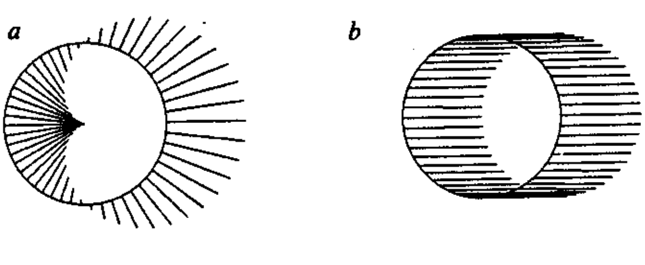

Fig. 2 Aperture problem. A circle of unit radius was assumed to be moving to the right with a velocity of 1.0. If the motion is measured with local detectors, then only the components of the motion perpendicular to the contour can be measured. This is an example of the aperture problem, and in, *a*, these perpendicular components are shown. The velocity field along the contour, as predicted by the motion coherence theory, is shown in *b*. This field was computed by solving equation (1) with $\lambda = 0.001$ and $\sigma = 3$. The isolated dot moving at the bottom of *b* has the correct velocity of the circle, implying that the aperture problem has been solved. The average speed of the computed velocity field along the contour was 0.996, which is incorrect by only 0.4%.

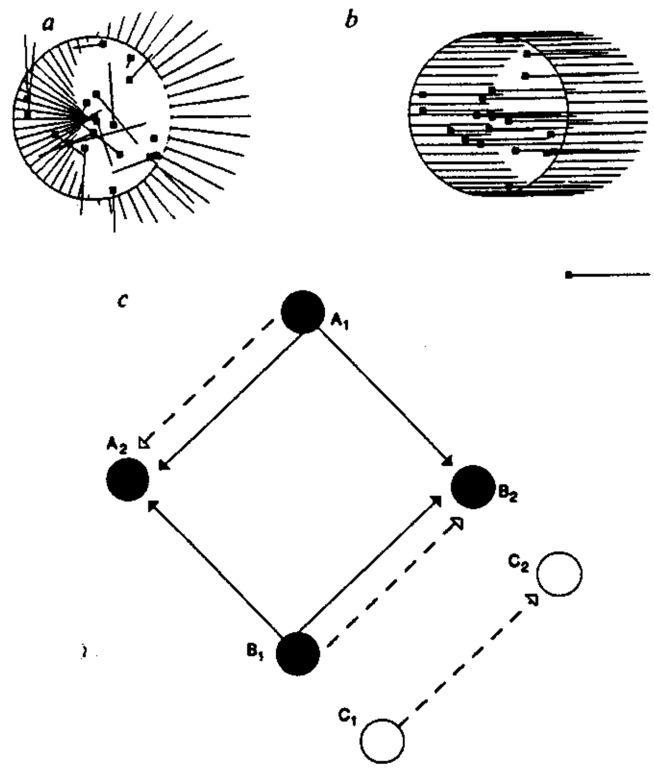

Fig. 3 Motion capture. In *a*, twenty randomly moving dots were placed in random positions inside the circle of Fig. 2. Their motion was random with the same distribution as in Fig. 1, but without the bias. Next, equation (1) was solved with the identical values of λ and σ as in Fig. 2. The result is shown in *b*. As for Fig. 2, the aperture problem was solved for the circle's motion, but this time this solution captured the motion of the internal dots. This result is similar to what humans perceive under the same conditions[5,6]. The capture predicted by the motion coherence theory can be strong; in the example the average speed of the solution along the contour and over the internal dots was 1.009. The theory also predicts capture for situations where only a few dots move. An example of this is in *c*. Suppose that an apparent motion display is created by alternating two dots between positions A_1 and B_1, and A_2 and B_2. If these positions lie on the corners of a square, then an ambiguous motion is perceived[30]; for example, the motion from B_1 is half the time towards B_2 and half the time towards A_2. This is represented by the solid arrows. Now, suppose that a third dot is shown in C_1 when the original dots are in A_1 and B_1, and in C_2 when the original dots are in A_2 and B_2. According to the theory, due to the spatial proximity of C_1 and B_1, the motion from C_1 to C_2 captures the motion starting from B_1, and forces this motion to go to B_2. This is represented by the open arrows. (From the cover principle[14] a motion from A_1 to A_2 is also induced.) We experimented with this paradigm and found that humans experience these illusions[23].

that is perpendicular to the contour. We apply smoothness to the velocity field in two-dimensional space, in contrast to Hildreth[19] who applies smoothness along the contour. (There is evidence that in humans, however, the motion system prefers integration of information along contours[25].) The solution has the same form of equation (3), but with an integral along the contour replacing the sum. The new method can provide a good solution to the aperture problem; in the example, the estimated speed is only 0.4% off the true speed. For the gaussian interaction, however, the computed speed is never exactly correct, being always smaller than the true one. This is due to the term c_0 of equation (1), which forces the overall speed down, but is necessary to ensure that the spatial interactions fall off smoothly at infinity[23]. (Without this term, it can be proven that the correct solution may be obtained for a certain class of stimuli[23] with similar results holding for Hildreth's method[26].) The 'incorrect' computed speed may explain the apparent non-rigidity of certain translating smooth planar curves[25].

A final example (Fig. 3) shows a possible explanation for motion capture[5,6]. In these phenomena, randomly moving dots are captured and move coherently with a superimposed grating or a surrounding contour. The motion coherence theory simultaneously solves the aperture problem of the surrounding contour and captures the internal dots. The theory also predicts that capture may happen when only a few dots move. An example of this occurs when a central ambiguous motion is captured by a peripheral unambiguous motion (Fig. 3*c*).

In conclusion, a theory for motion coherence that deals simultaneously with the aperture problem and the phenomena of motion cooperativity and motion capture is proposed. The theory differs from other recent works that deal with similar problems[27,28]. While these works are based on neural considerations, we provide a computational theory expressed as an optimization process. Theories derived in this way can directly incorporate the natural constraints of the visual world[29]. In spite of the computational approach, the theory's elements can be interpreted in terms of neural processes.

We thank E. C. Hildreth and E. H. Adelson for useful comments on the manuscript. N.M.G. was supported by ONR, and the Sloan and Fairchild foundations. A.L.Y was supported by USARO.

Received 29 December 1987; accepted 17 March 1988.

1. Ternus, J. *Psychol. Forsch.* **7,** 81-136 (1926).
2. Koffka, K. *Principles of Gestalt Psychology* (Harcourt, Brace & World, New York, 1935).
3. Williams, D. W. & Sekuler, R. *Vision Res.* **24,** 55-62 (1984).
4. Williams, D. W., Philips, G. & Sekuler, R. *Nature* **324,** 253-255 (1986).
5. Mackay, D. M. *Nature* **192,** 739-740 (1961).
6. Ramachandran, V. S. & Anstis, S. M. *Vision Res.* **23,** 1719-1724 (1983).
7. Verri, A. & Poggio, T. *Artif. Intell. Memo* **917,** (MIT, Cambridge, 1986).
8. Marr, D. & Ullman, S. *Proc. R. Soc.* **B211,** 151-180 (1981).
9. Braddick, O. J. *Phil. Trans. R. Soc. Lond.* **B290,** 137-151 (1980).
10. Horn, B. K. P. & Schunk, B. G. *Artif. Intell.* **17,** 185-203 (1981).
11. Hassenstein, B. & Reichardt, W. E. *Z. Naturforsch.* **11b,** 513-524 (1956).
12. van Santen, J. P. H. & Sperling, G. *J. opt. Soc. Am.* **A1,** 451-473 (1984).
13. Adelson, E. H. & Bergen, J. R. *J. opt. Soc. Am.* **A2,** 284-299 (1985).
14. Ullman, S. *The Interpretation of Visual Motion* (MIT Press, Cambridge, 1979).
15. Grzywacz, N. M. & Yuille, A. L. *AIP Conf. Proc.* **151,** (ed. Denker, J. S.) 200-205 (American Institute of Physics, New York, 1986).
16. Adelson, E. H. & Movshon, J. A. *Nature* **300,** 523-525 (1982).
17. Anstis, S. M. *Vision Res.* **10,** 1411-1430 (1970).
18. Ramachandran, V. S. & Anstis, S. M. *Vision Res.* **23,** 83-85 (1983).
19. Hildreth, E. C. *The Measurement of Visual Motion* (MIT Press, Cambridge, 1983).
20. Poggio, T., Voorhees, H. & Yuille, A. L. *Artif. Intell. Memo* **776,** (MIT, Cambridge, 1984).
21. Bertero, M., Poggio, T. & Torre, V. *Artif. Intell. Memo* **924,** (MIT, Cambridge, 1987).
22. Ramachandran, V. S. & Inada, V. *Spatial Vision* **1,** 57-67 (1985).
23. Yuille, A. L. & Grzywacz, N. M. *Artif. Intell. Memo* **1002,** (MIT, Cambridge, 1987).
24. Duchon, J. *Lecture Notes in Math.* **571,** (eds Schempp, W. & Zeller, K.) 85-100 (Springer-Verlag, Berlin, 1979).
25. Nakayama, K. & Silverman, G. H. *Invest. Ophthalmol. Vis. Sci.* **26,** 188 (1985).
26. Yuille, A. L. *Artif. Intell. Memo* **724,** (MIT, Cambridge, 1983)
27. Little, J., Bulthoff, H. & Poggio, T. *Proc. Image Understanding Workshop* (ed. Baumann, L.) 915-920 (Science Applications International Corporation, McLean, Virginia, 1987).
28. Heeger, D. *J. opt. Soc. Am.* **A4,** 1455-1471 (1987).
29. Marr, D. *Vision* (Freeman, New York, 1982).
30. von Schiler, P. *Psych. Forsch.* **17,** 179-214 (1933).