A Model of Motion Computation in Primates

H. Taichi Wang and Bimal P. Mathur
Science Center, Rockwell International, Thousand Oaks, CA 91360
and
Christof Koch
Division of Biology, California Institute of Technology, Pasadena, CA 91125

Abstract

A neural network model of motion field computation in the visual system of the primates is proposed. The model assumes the local velocities are computed first in the primary visual cortex (V1) using the information available from the sustained and the transient pathways originating in the retina. A variation of Horn and Schunck's optical flow algorithm (1981) and the direction-selective representation of velocities found in the primary visual cortex and middle temporal area (MT) are used to solve the aperture problem. The resulting network has its connection matrix independent of measured data. We adopt Hubel and Wiesel's model for orientation selectivity, and generalize Marr and Ullman's XYX model of direction selectivity to compute local velocities. Preliminary fusion of edge and motion information is also proposed. Results of computer simulation, including the Adelson and Movshon experiment and the phenomena of motion capture, are presented.

1. Introduction

Any natural environment contains moving objects. They may be preys for which we need to track their motion. Or they may be the predators for which we need to know whether they are approaching or not. We also use information about relative motion to infer the three-dimensional structure of objects and their relative distances from us. Thus, extracting motion is crucial among early vision processes.

There are two fundamentally different methods for computing motion, the token-based methods and the image-intensity-based optical flow methods. The token-based methods rely on unambiguous identification of stable features, such as corners, before matching the features from one instance to the next. The methods of optical flow operate on the spatial and temporal variation of the values of or quantities derived from the time-varying image brightness. Psychophysical evidence suggests that both methods are used in primates (An extensive review can be found in [1].). This article concerns itself with a model based on optical flow.

The key problem in computing optical flow from changing intensity values is the **Aperture Problem**: For certain images, e.g. sine waves or straight edges, no local measurement process can extract the correct velocity field; instead, only the components of velocity parallel to local gradients can be measured (Figure 1(a)). Additional *a priori* constraints are required to recover the motion field (the true velocity field).

To solve this problem, Horn and Schunck [2] imposed a smoothness constraint: the computed motion field should be (1) compatible with the locally measured data, and (2) smoothest in a certain sense. This leads to a quadratic energy functional which needs to be minimized. It has been shown that this energy functional can be minimized using a simple resistive network [3], which one of the authors (C.K.) is currently implementing in subthreshold analog CMOS VLSI circuits [4]. This algorithm leads to the correct motion field for the class of rigid objects translating in space. The "velocity space" method of Adelson and Movshon [5] can be considered as a special case of Horn and Schunck's more general method. In fact, it has been recently realized that the problems arise in early vision are *ill-posed* in nature. They can best be formulated in the general mathematical framework of the theory of regularization [6] by introducing suitable stabilizing functionals.

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 28,2025 at 00:20:29 UTC from IEEE Xplore. Restrictions apply.

The Horn and Schunck algorithm works without computing local velocities explicitly. The resulting resistive network has some of its connection matrix elements dependent on the measured data. Existing psychophysical and physiological evidence, however, suggest that in primates motion field are computed in two stages. Firstly, the local velocities are computed in the primary visual cortex (area V1) by taking inputs from the sustained and transient pathways from retinal ganglion cells through LGN [7]. Then the motion field is computed by solving the aperture problem, which is thought to take place in the extrastriate cortical area MT (middle temporal) [8][9]. Recently, we have proposed a neuronal network model of motion computation in area MT, based on the information available to the primary visual cortex [10]. In this article, the details of the model and the algorithm used are presented. The algorithm is a variation of Horn and Schunck's. Because of the particular representation chosen for the local velocity field and the motion field, the resulting connection matrix of the network is independent of measured data. This makes it appealing from both a biological and a technical point of view.

In the next section, the representation and algorithm used for computing optical flow are discussed. We then describe how the required information can be extracted from the retinal pathway in area V1. We adopt Hubel and Wiesel's model for orientation selectivity [11] to calculate the local edge strengths, and generalize Marr and Ullman's model of direction selectivity [12] to compute local velocities. As we shall see, additional information is needed; such as the direction of local velocities at each position, and whether the local edges are strong enough (the meaning and purpose of each term will become clear as we go along). Results of computer simulation are presented next. Finally, some conclusion and future extensions are discussed.

2. The Network for Motion Computation

Our original intention was to find a network for motion computation whose connection matrix is independent of input data. It turns out that the *direction selective representation* found in the hypercolumns in area V1 and MT is particularly convenient [7][9]. In this representation, the state of a neuron V(i, j, k) represents the velocity at position (i, j) in the (discrete) direction k, with speed (non-negative) V. The direction selective representation is very different from the vector decomposition representation used in the resistive network, in which two neurons are used to encode the x component and y component of the velocity field at each pixel. We shall use the direction selective representation to represent both the local velocities U(i, j, k) and the final motion field V(i, j, k) with Nd of directions distributed between 0 and 360 degree.

The concept of a constraint line associated with each local velocity measurement is illustrated in Figure 1(b). Every point on the constraint line has the same projection along the direction of measured local velocity and hence it is consistent with that local measurement. If the moving object is rigid, moving in the plane normal to the plane of sight, and all the measurements are noise-free, all the constraint lines associate with the object will intercept each other at a point V, which is the true velocity of the object (Figure 1(b)) [5].

In the presence of a moving edge, not only the local velocity neuron U(i, j, k) parallel to the edge gradient respond. In fact the direction selective tuning curve is cosine-like centered around the one normal to the edge at that position. It is easy to see that all the constraint lines at that position are *wrong* except for the one associated with the local velocity normal to the local edge; i.e. the one with largest local velocity measurement. A direction selection neuron $U_\theta(i, j, k)$, therefore, is called for. At the moment, suffice it to say that at each position (i , j), $U_\theta(i, j, k)$ are all zero except for the direction whose local velocity measurement is largest among the k directions; in which case, it is equal to unity. The terms corresponding to invalid constraint lines are thus shunted out by the direction selection function. It is well known that shunting, or silent, inhibition together with depolarizing excitation and the hyperpolarizing inhibition are the most fundamental interactions between neurons [13]. More detail discussion on the direction selection function can be found in section 3.

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 28,2025 at 00:20:29 UTC from IEEE Xplore. Restrictions apply.

Because of the unavoidable noise, the concept of constraint lines can best be formulated in terms of an energy functional, L_0 , which needs to be minimized. It can be written as, taking into account that only one constraint line at each position is valid:

$$L_0 = \sum_{i,j,k,k'} [U(i, j, k') - \cos(\theta_k - \theta_{k'}) V(i, j, k)]^2 U_\theta(i, j, k') \quad (1)$$

where is the angle of the kth direction with respect to the x axis. Equation (1) constrains the final solution to be as close as possible to the measured local velocity. Additional constraints are required, however, to resolve the aperture problem. Horn and Schunck [2] suggested the use of the smoothness functional, L_1, of the following form:

$$L_1 \sim \int \left[\left(\frac{\partial Vx}{\partial x}\right)^2 + \left(\frac{\partial Vx}{\partial y}\right)^2 + \left(\frac{\partial Vy}{\partial x}\right)^2 + \left(\frac{\partial Vy}{\partial y}\right)^2 \right] dx\, dy \quad (2)$$

where Vx and Vy are the x and y components of the velocity field, respectively.

We can rewrite the smoothness constraint (2) in terms of the direction selective representation as follows. Let's assume that there exists a winner-take-all (WTA) among the directions at each position; that is, only the direction with the largest V retains its value, all other directions at that position are set to zero. In other words, we assume very highly tuned direction selective cells. The smoothness functional, L_1, after discretization, can now be written as

$$L_1 = \sum_{i,j,k,k'} [\, 4V(i,j,k) - V(i-1,j,k') - V(i+1,j,k') - V(i,j-1,k') - V(i,j+1,k') \,] \cdot \cos(\theta_k - \theta_{k'}) V(i,j,k)) \quad (3)$$

Because a winner-take-all circuit has been assumed in formulating the smoothness functional (5), we have to add an appropriate energy functional, L_2, that will enforce WTA. It can be written as

$$L_2 = \sum_{i,j,k,k'} [1 - \cos(\theta_k - \theta_{k'})] V(i, j, k') V(i, j, k) \quad (4)$$

The total energy functional of the motion network, L, now consists of three terms:

$$L = L_0 + C_1 L_1 + C_2 L_2 \quad (5)$$

where C_1 and C_2 are coefficients for the smoothness term and the WTA term, respectively. Other terms, such as that for line processes for dealing with the motion discontinuities [4], and fusion of information with stereo vision can be easily incorporated in the energy functional framework.

The network equation can be obtained simply as

$$\begin{aligned} \partial V(i, j, k)/\partial t &= -\partial L/\partial V(i, j, k) \\ &= \sum_{k'} [\, [U(i, j, k') - \cos(\theta_k - \theta_{k'}) V(i, j, k)] \cos(\theta_k - \theta_{k'}) U(i, j, k') \\ &\quad + C_1 \, [V(i - 1, j, k') + V(i + 1, j, k') + V(i, j - 1, k') + V(i, j + 1, k') \\ &\quad\quad - 4V(i, j, k')] \cdot \cos(\theta_k - \theta_{k'}) \\ &\quad - C_2 \, [1 - \cos(\theta_k - \theta_{k'})] V(i, j, k')] \end{aligned} \quad (6)$$

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 28,2025 at 00:20:29 UTC from IEEE Xplore. Restrictions apply.

3. From Retina to V1

For primates, the inputs from retina through LGN to the primary visual cortex consist mainly of two types: the sustained X channel, and the transient Y channel. Their receptive fields are circularly symmetrical. According to Marr and Hildreth's theory of edge detection, the X channel can be modeled by the convolution of the input image, I(i, j, t), with the Laplacian-of-Gaussian filter, $X(i, j) = \nabla^2(G_x * I)$ [14]. The Y channel is thought to carry the temporal derivative of the Laplacian-of-Gaussian convolved images, $Y(i, j) = (\partial / \partial t)[\nabla^2(G_y * I)]$ [12]. Orientation selectivity and direction selectivity first appears in area V1 [7].

There exist several models of orientation selectivity [15]. We adopt Hubel and Wiesel's model for its simplicity, and because the performance of the motion network does not depend on the details of the underlying circuitry (It may become relevant once we consider the dynamics of motion computation.). The Hubel and Wiesel model amounts to computing the gradient of the X channel input along different directions with a kernel of finite size. In our simulation, we use a kernel size of 2 x 5. The result, (non-negative) E(i, j, k), signal edges at (i, j) in orientation k:

$$E(i, j, \mathbf{k}) \sim [\, \nabla X(i, j) \,] \bullet \mathbf{k} \tag{7}$$

where **k**, ranging from 0 to 360 degree, is the unit vector in the direction of kth orientation.

In order to obtain a quantitative measure of local velocity, we generalize Marr and Ullman's model of direction selectivity [12]. The local velocities at each position (i, j) in the direction k, U(i, j, k) can be computed as

$$U(i, j, \mathbf{k}) = - \frac{Y(i, j)\, E(i, j, \mathbf{k})}{[\, |E(i, j)|^2 + \varepsilon]} \tag{8}$$

Notice that the normalization in the denominator is necessary to normalize out the intensity of the image. Because the orientation tuning curve is not a perfect cosine function, the $|E(i, j)|$ is calculated as

$$|E(i, j)|^2 = \frac{4}{Nd} \bullet \sum_{\mathbf{k}} E^2(i, j, \mathbf{k})$$

Since the local velocity measurements tend to give erroneous results at locations where no significant edge is present, a parameter, ε, is added to the denominator.

The above definition of local velocity (8) and constraint line equation (1) point to the need of preliminary data fusion between motion field and edge information. When an edge is moving in the direction normal to the edge gradient, the local velocity at that point is zero. There should be a term in the constraint line equation (1) corresponding to a constraint line going through the origin in the velocity plane, in the direction normal to the edge gradient. In contrast, if there is no edge present in a neighborhood, there should not be a constraint line at that position, even though the local velocity (8) in both cases is zero (assuming the input images are noise-free). We have to, therefore, shunt out those constraint lines which come from weak or non-existent edges.

With the basic edge and local velocity information in place, we can now construct the auxiliary quantities we need for the motion network. They are the edge selection function $E_\theta(i, j, k)$ and direction selection function $U_\theta(i, j, k)$. Their direction indices range from 0 to 360 degree.

Since an edge of a given orientation can give rise to local velocity in parallel or anti-parallel to the edge gradient, we define the edge selection function as follows. If the maximum edge strength at the location (i, j) is in direction k and is above a threshold θ, $E_\theta(i, j, k)$ and $E_\theta(i, j, -k)$ are set to 1, while the remaining directions are set to 0, where direction -k is the direction opposite to direction k. Otherwise,

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 28,2025 at 00:20:29 UTC from IEEE Xplore. Restrictions apply.

$E_\theta(i, j, k)$ are set to zero for all directions. The edge selection function, is essentially the contrast-independent oriented edge detector.

The local velocity selection function, $U_\theta(i, j, k)$, can be defined as the binarized product of the edge selection function E_θ and the local velocities,

$$U_\theta(i, j, k) = \begin{cases} 1 & \text{if} \quad E_\theta(i, j, k)\, U(i, j, k) \neq 0 \\ 0 & \text{if} \quad E_\theta(i, j, k)\, U(i, j, k) = 0 \end{cases} \tag{10}$$

since only one of either the k or the -k direction of U at (i, j) can be non-zero.

4. Simulation Results

We have tested the network's ability to resolve the aperture problem with both acquired and synthesized images. In both cases, the network arrived at the correct solutions. The network not only solves the aperture problem correctly, it also shows the same perception as the human do in psychophysical experiments of Adelson and Movshon [5] and motion capture of Ramachandran and Inada [16]. These are the consequence of the smoothness constraint used in the algorithm. In the following simulations, typical parameters used are: Nd = 16, the Gaussian width of both X channel and Y channel S = 2, C_1 = 50, and C_2 = 5.

In the Adelson and Movshon experiment, the images, acquired through camera, consist of two orthogonal gratings, each moving in the direction perpendicular to its edges, as shown in Figure 2(a). A human subject perceives the pattern as moving as a coherent right plaid pattern in the horizontal direction. The network also perceives it in the same way, as shown in Figure 2(b).

When a human subject is presented a screen of randomly moving random-dots pattern, the human subject perceives the random motion of the dots. On the other hand, if bars moving in the same direction are superimposed on the randomly moving dots pattern, the human subject perceives the random dots as moving coherently with the bars. The phenomena is called motion capture [16]. We have simulated the experiment with the patterns shown in Figure 3(a). As shown in Figure 3(b), the random dots are seen by the network as moving incoherently. As two vertical bars moving horizontally are added to the test pattern, as shown in Figure 3(c), the network perceives the randomly moving dots as moving coherently in the horizontal direction, as shown in Figure 3(d).

5. Conclusion

A model of motion field computation in primates has been developed. This model assumes only the information available to the primary visual cortex, such as orientation and direction selectivity in V1. The resulting neuronal network has a connection matrix independent of the input data. This model relies on existing models for the sustained and transient pathways, the orientation selectivity, and direction selectivity. The simulation results show that it is capable of correctly solving the aperture problem for a certain types of stimulus patterns. There are certainly a lot of biological details which have not been taken into account, such as speed selectivity for the direction selective cells [9][17][18]. It is possible that an elaborated version the model forms the basis of motion computation in area MT.

Even though Horn and Schunck's smoothness constraint is used because of its simplicity, it is certainly possible that more complicated smoothness constraints are used in primates. For example, it has been suggested that the motion field is better modeled by second-order polynomials [19]. In that case the appropriate smoothness constraint may contain combinations of third order derivatives [20]. The network model can easily be modified with the method outlined in section 2.

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 28,2025 at 00:20:29 UTC from IEEE Xplore. Restrictions apply.

One of the key problems affecting the class of algorithms which use any version of the smoothness constraint is that they tend to smooth over object boundaries. Their performance can be greatly improved by introducing analog or binary variables sensitive to discontinuities either in magnitude or direction or both of the velocity within their receptive fields [4][21]. These processes have great similarities with cells first reported by Allman, Miezin, and McGuinness which respond optimally to shearing motions [18]. Such line processes are currently being implemented in our model.

References

[1] E. C. Hildreth and C. Koch, "THE ANALYSIS OF VISUAL MOTION: From Computational Theory to Neuronal Mechanisms", *Ann. Rev. Neurosci.*, **10**, 477 - 533, 1987.

[2] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow", *Artif. Intell.* **17**, 141 - 184, 1981.

[3] C. Koch, J. Marroquin, and A. Yuille, "Analog 'neuronal' networks in early vision", *Proc. Natl. Accad. Sci. USA*, **83**, 4263 - 4267, June 1986.

[4] J. Hutchinson, C. Koch, J. Luo, and C. Mead, "Computing Motion Using Analog and Binary Resistive Network", *IEEE Computer*, **21** (3), 52 - 63, March 1988.

[5] E. H. Adelson and J. A. Movshon, "Phenomenal coherence of moving visual patterns", *Nature*, **300**, 523 - 525, 1982.

[6] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory", *Nature*, **317**, 314 - 319, 1985.

[7] D. H. Hubel and T. N. Wiesel, "Functional architecture of macaque monkey visual cortex", *Proc. R. Soc. Lond.* **B 198**, 1- 59, 1977.

[8] J. A. Movshon, E. H. Adelson, M. S. Gizzi, and W. T. Newsome, "The Analysis of Moving Visual Pattern", In *Pattern Recognition Mechanisms*, ed. C. Chagas, R. Gattas, C. G. Gross. Rome: Vatican Press, 1985.

[9] T. D. Albright, "Direction and Orientation Selectivity of Neurons in Visual Area MT of the Macaque", *Journal of Neurophysiology*, **52** (6), 1106 - 1130, 1984.

[10] H. T. Wang, B. P. Mathur, and C. Koch, "Motion field computation in the mammalian cortex", in Snowbird Meeting on Neural Networks for Computing, 1988.

[11] D. H. Hubel, and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", *J. Physiol. (Lond.)*, **166**, 106 - 154, 1962.

[12] D. Marr, and S. Ullman, "Directional selectivity and its use in early visual processing", *Proc. R. Soc. Lond.* **B 211**, 151 - 180, 1981.

[13] E. R. Kandel, and J. H. Schwartz ed., *Principles of Neural Science*, 2nd ed., Elsevier, 1985.

[14] D. Marr, and E. Hildreth, "Theory of edge detection", *Proc. R. Soc. Lond.* **B 207**, 187 - 217, 1980.

[15] D. Ferster, and C. Koch, "Neuronal connections underlying orientation selectivity in cat visual cortex", *Trends NeuroSci*, **10** (12), 487 - 492, 1987.

[16] V. S. Ramachandran, and V. Inada, "Spatial phase and frequency in motion capture of random-dots pattern", *Spatial Vision*, **1**, 57 - 67, 1985.

[17] J. H. R. Maunsell, and D. C. Van Essen, "Functional properties of neurons in middle temporal visual area of the macaque monkey. I. selectivity for stimulus direction, speed, and orientation", *Journal of Neurophysiology*, **49** (5), 1127 - 1147, 1983.

[18] J. Allman, F. Miezin, and E. McGuinness, "Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT)", *Perception*, **14**, 105 - 126, 1985.

[19] A. M. Waxman, and K. Wohn, "Image flow theory: a framework for 3D inference from time-varying imagery", ed. C. Brown, in *Advances in Computer Vision*, vol. **1**, 165 - 224, 1988.

[20] A. M. Waxman, private communication.

[21] S. Geman, and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Trans. Pattern Anal. Machine Intell.*, **PAMI-5**, 721 - 741, 1984.

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 28,2025 at 00:20:29 UTC from IEEE Xplore. Restrictions apply.

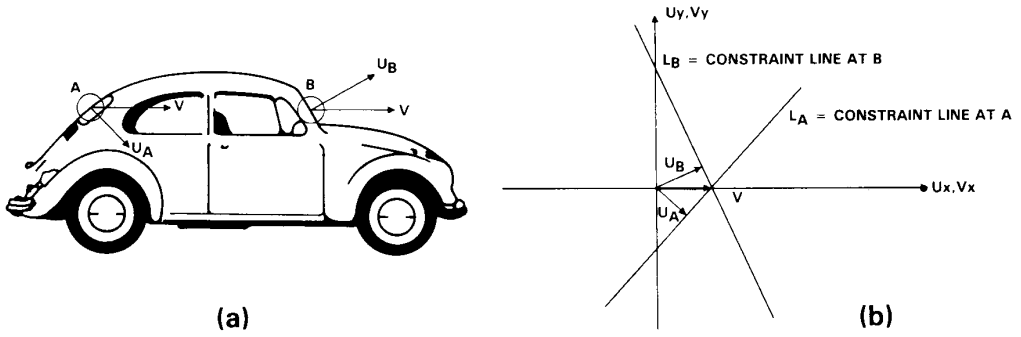

Figure 1 The aperture problem. (a) Local velocity measurements can only yield the components normal to the local edges. (b) Each local measurement give rise to a constraint line in the velocity space, which is perpendicular to the local velocity measured. If the data is noise-free, and all the data are taken from the same object, the intersection point of the constraint lines is the object's true velocity.

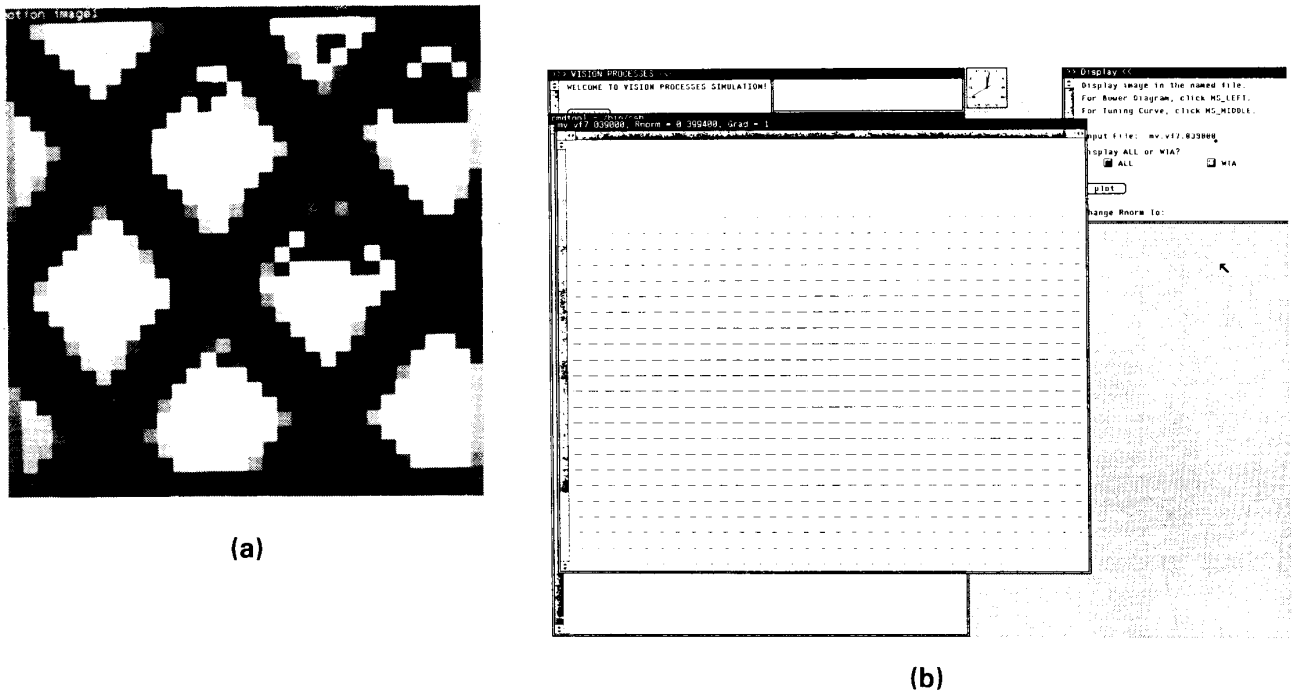

Figure 2 The Adelson and Movshon experiment. The input patterns consist of two orthogonal plaid patterns each moving perpendicular to it edges (a), and (b) the network perceives the pattern as moving as a whole horizontally. The result is shown with the needle diagram in which the length and direction of the needle represent the speed and direction of the velocity vector at that pixel position.

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 28,2025 at 00:20:29 UTC from IEEE Xplore. Restrictions apply.

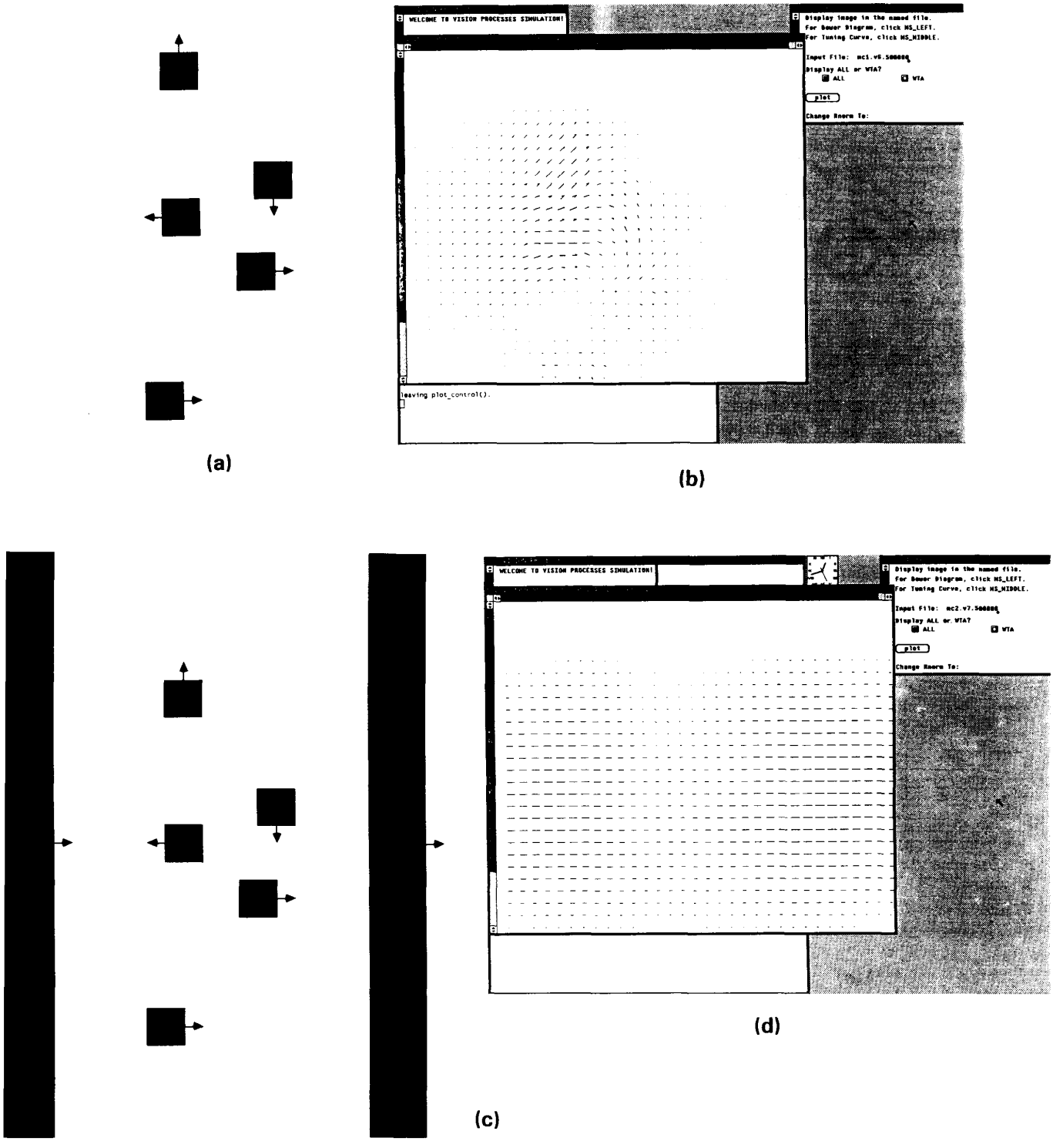

Figure 3 The experiment of motion capture. When the input images consist of only randomly moving dots, (a), the network perceives a random motion field. Once two bars, which are moving horizontally in the same direction are included (c), the network perceives the randomly moving dots as moving coherently in the horizontal direction (d), just as human subjects do.

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on October 28,2025 at 00:20:29 UTC from IEEE Xplore. Restrictions apply.