# Some Insights Into Brightness Perception of Images in the Light of a New Computational Model of Figure–Ground Segregation

Kuntal Ghosh and Sankar K. Pal, *Fellow, IEEE*

***Abstract*—The excitatory–inhibitory visual receptive-field model may be looked upon as a classical structuralist approach to vision that relies upon brightness–contrast information of the image as a preliminary step toward visual representation. The corresponding mathematical operator (Laplacian) was first proposed by the empiricist Ernst Mach on the basis of the Mach band illusion. The Helmholtz's constructivist approach, on the other hand, argues that perception is the product of unconscious inference. Propagating a sort of intermediate stance between these two viewpoints led to the emergence of the Gestalt school for whom perception follows a minimum principle and is at the same time holistic, based on certain coherence criteria. In this paper, we have modeled the extraclassical receptive field through an eigenfunction-based generalization of the Gaussian derivative approach that resulted in a modification of Mach's equation, introducing a higher order isotropic derivative (Bi-Laplacian) of Gaussian and a fourth-moment operator. The proposed computational model draws its inspiration from the structuralist approach, performs figure–ground segregation in Gestalt sense, and also provides cues toward brightness perception in tune with the constructivist notion of unconscious inference.**

***Index Terms*—Brightness perception, extraclassical receptive field (ECRF), figure–ground organization.**

## I. Introduction

THE excitatory–inhibitory visual receptive-field model may be looked upon as a classical structuralist approach to vision that relies upon brightness–contrast information of the image as a preliminary step toward visual representation. The corresponding mathematical operator (Laplacian) was first proposed by the empiricist Ernst Mach [1] on the basis of the Mach band illusion. He stated:

"Let us call the intensity of illumination $u$ on a uniform mat plane where $u = f(x, y)$. Thus, the brightness sensation $v$ of the corresponding retinal point is given by

$$v = u - m(d^2u/dx^2 + d^2u/dy^2) \tag{1}$$

where $m$ is a constant. If the expression in the brackets is positive, then the sensation of brightness is reduced; in the opposite case, it is increased. Thus, $v$ is not only influenced by $u$ but also by its second differential quotients." The question that immediately arises is whether such an extraction of brightness–contrast information is indeed the fundamental step or whether there is some other precondition to such extraction. Alongside the structuralist approach, there existed Helmholtz's constructivist approach [2] which on the other hand, claimed that perception is the product of unconscious inference. Helmholtz argued that information provided by our senses is inadequate and is actually augmented by *unconscious inferences*, which add meaning to sensory information. He assumed these inferences to be unconscious because we typically have no awareness that we are making such inferences while perceiving. According to this assumption, perception is not directly given by stimulus input but occurs as the end product of the interactive influences of the presented stimulus and certain internal hypothesis, expectation, and knowledge, as well as motivational and emotional factors. The flavor of the constructivist approach has been captured in recent times by Gregory [3] who claims that perceptions are constructions "from floating fragmentary scraps of data signaled by senses and drawn from the memory banks, themselves constructions from the snippets of the past." In spite of some definite positive aspects associated with this idea of perception as an act of *unconscious inference* and likelihood principle as envisaged by Helmholtz, this approach undoubtedly dissociated itself at least to some extent from the low-level data-driven neurophysiological or computational approaches initiated by the structuralists and fell prey to nativism [4]. Therefore, considering both Mach's and Helmhotz's views on vision, one may conclude that there can indeed be some internal hypothesis, expectation, or motivational factor that may influence the extraction of brightness–contrast information from images by the human visual system (HVS). Such an attempt to synthesize the two ideas however is not completely new. Historically, in between these two opposing approaches emerged the Gestalt school of thought for whom perception follows a minimum principle and is at the same time holistic. According to the Gestalt psychologists like Koffka [5], the elementary features in a visual scene are locally bound together into coherent groups on the basis of such coherence criteria as *Proximity*, *Similarity*, *Continuity*, *Common fate*, and *Closure*. For instance, by *Similarity*, it is meant that the elements are grouped together with elements that have the same or similar features, i.e., any two parts of the same simple object probably look alike. *Continuity* means an element is a part of that object which gives

Manuscript received June 15, 2008; revised June 11, 2009. Date of publication April 12, 2010; date of current version June 16, 2010. This paper was recommended by Associate Editor J. Wu.

K. Ghosh is with the Machine Intelligence Unit and the Center for Soft Computing Research, Indian Statistical Institute, Kolkata 700 108, India (e-mail: kuntal@isical.ac.in).

S. K. Pal is with the Center for Soft Computing Research, Indian Statistical Institute, Kolkata 700 108, India (e-mail: sankar@isical.ac.in).

Digital Object Identifier 10.1109/TSMCA.2010.2044503

the most continuous border. Consequently, the next element of a set can be determined from the preceding elements. By *Closure* it is meant that parts that are not present in the picture are filled in. Thus, dots placed in a circle represent a circle. However, the Gestalt approach has its own shortcomings too. Therefore, while for the structuralists there is a commitment to some strategy of analyzing basic data or experience into its elements (atoms), the Gestalt approach sometimes goes to the extent of denying any necessity or even appropriateness of such analysis or synthesis, like the *creative synthesis* of Wundt [6]. However, it is now established that the Gestalt school indeed takes into consideration some important aspects, like grouping or figure–ground segregation. This is the aspect which is also the central theme for us in this paper. For example, it was shown by Kanizsa that convexity has a stronger influence on figural organization than other global-shape properties such as symmetry [7]. Von der Malsburg and Buhmann [8] have also shown that in addition to the local coherence criteria, as prescribed by the Gestalt school, the contribution of relevant information at a higher level may also be required to create coherent groups and to separate figures from the background more globally. However, both the neural representation of elementary features and the mechanism by which they are integrated into coherent groups and figures still remain largely unclear to date [9], although there do exist some contemporary models toward understanding the brain as an integral whole, like the recent thesis [10] proposing a hierarchical and temporal model and another very recent work by Poggio *et al.* [11] based on a Bayesian network model.

A possible reason for such a situation might be the philosophically idealist tendency to study these high-level perceptual-grouping-based approaches in isolation from the theories of low-level vision, i.e., not applying the Gestalt context in the true historical perspective of its arising out of a conflict between the structuralist and constructivist schools and the growing need to dialectically synthesize the two ideas, viz., the influence of structure on function and that of function back to structure. The work of Poggio *et al.* [11] is related to this domain. Thus, while the so-called contrast theorists on one hand, spurred by the first direct physiological evidences of lateral inhibition, tried to argue that such physiological facts render the vague ideas of Gestalists obsolete, the latter in turn tended to undermine the physiological facts, thus sparking off what, in current vocabulary, may be termed as the top-down versus bottom-up process debate. Bottom-up processing, often termed as data-driven processing, refers to those processes that take a lower level representation as input and create a higher level representation as the output. On the other hand, top-down processing, also termed as hypothesis-driven or expectation-driven processing, refers to those processes that operate in a diametrically opposite direction, taking a higher level representation as input and producing a lower level representation as output. As a matter of fact, such a tendency to study perceptual grouping in isolation from low-level vision was long contradicted by Fechner, the father of psychophysics, who proposed the theory of self-organization that leads to the principle of emergent behavior (i.e., the whole is not a mere sum of parts) [12]. Fechner's posthumously discovered theory of measuring collectives already defines probability as the scope of relative frequency. Thus, Fechner may also be considered as a forerunner of the top-down approach based on Bayesian inference or likelihood principle, which has been subsequently carried forward through the hypothesis testing of Gregory or the Bayesian inference of Rock [13]. However, unlike the constructivist school of thought, this statistical approach of Fechner cannot be separated from his proposed simple bottom-up laws of psychophysics firmly based upon experimental observations.

Our attempt in this paper is to study the influence of Gestalt perceptual grouping, particularly figure–ground segregation upon low-level visual task of brightness perception in continuation of the works of the likes of Agostini and Proffitt [14], Gilchrist and Bonato [15], Todorovic [16], and many others. The naïve intuition that vision is essentially a bottom-up process that begins with the sensory information in the retinal images and goes unilaterally upwards to perceptual and then conceptual interpretations is now no longer tenable, although many theorists still concur that the early stages of visual processing are indeed strictly bottom up [17]. Particularly, the point at which top-down processes begin to augment bottom-up processes still remains a controversial issue, but its general truth is now virtually unquestionable. Hence, the influence of perceptual groupings on brightness perception is now already a subject of intense study, but computational models to this end are still not very much in vogue. A good deal of rationale for a theoretical framework of the role of bottom-up processes in higher level perceptual phenomena, not in the sense of naïve unilateralism, was provided quite some time back by D. Marr and his group at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology [17], based on which visual perception was decomposed at the algorithmic level into four major stages beyond the retinal image itself, viz., the image-based, surface-based, object-based, and category-based stages of perception. The works of Marr and his group established that the initial registration in the two eyes is not the sole representation based on a 2-D retinal organization—on the contrary, there are additional representations and processes in this image-based stage, including image processing operations like detection of local edges, linking these edges more globally, and defining 2-D regions in the image. These 2-D features characterize their structure and organization before being interpreted as properties of the 3-D visual scenes. The question that we would like to address in this paper is whether we can find a computational model, in continuation of the works of Marr and his group, for one such higher level perceptual grouping, namely, figure–ground segregation that in turn influences a low-level process like brightness perception.

## II. Proposed Approach

The structuralist approach already mentioned, which has been championed by contrast theorists, emanates from Mach's mathematical model given by (1) in the Introduction. Two points which appear significant regarding this equation are the following: 1) proposal of an isotropic derivative operator as filtering function operating on the intensity distribution that is combined with the original distribution at the stage of low-level vision and 2) the constant $m$, i.e., how $m$ may actually be fixed or adjusted, indicating a possible role of higher level unconscious inference.

Mach's mathematical model [(1)], where he first used the Laplacian operator as a low-level filter for information processing, was further extended much later by Marr and Hildreth [18] who proposed the Laplacian of Gaussian (LOG) function in searching edge information. The 2-D LOG function proposed by Marr and Hildreth is given by

$$\nabla^2 g(r,\sigma) = -\frac{1}{\pi\sigma^2}\left[1-\frac{r^2}{2\sigma^2}\right]e^{-\frac{r^2}{2\sigma^2}}. \quad (2)$$

Here

$$g(r,\sigma) = \frac{1}{2\pi\sigma^2}e^{-\frac{r^2}{2\sigma^2}} \quad (3)$$

$\sigma$ being the standard deviation of the Gaussian function $g$.

When a 2-D intensity distribution is convoluted with this function, i.e., (2), zero crossings are formed at positions of intensity discontinuity. Detection of these zero crossings thus provide an edge map of the image. The function LOG in fact, structurally also models the center-surround receptive-field organization of many neurons, starting from the retinal ganglion cells and the cells of the lateral geniculate nucleus to many others in visual cortex. This center-surround receptive-field organization was independently already modeled by the physiologists as a difference of two Gaussians (DOG)—one with a smaller variance representing the center and the other with a wider variance representing the surround. Mathematically, this famous lateral inhibition model in a representative form in 1-D is given as

$$\mathrm{DOG}(\sigma_1,\sigma_2) = A_1\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{x^2}{2\sigma_1^2}} - A_2\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{x^2}{2\sigma_2^2}}. \quad (4)$$

Here, $\sigma_1$ and $\sigma_2$ represent the scales of the center and the surround, respectively, and $A_1$ and $A_2$ represent the corresponding amplitudes. Marr–Hildreth also demonstrated an empirical equivalence of the LOG and DOG.

The LOG-based spatially filtered information is presumably sent to the brain via optic nerves, and the zero crossings may be detected in the primary visual cortex according to this scheme, which has some, although not conclusive, support from experiments [19], [20]. However, it has been established by physiologists time and again that the actual receptive field for the neurons is much wider than was previously thought of at both retinal and cortical levels that has led to the notion of the extraclassical receptive field (ECRF) [21]–[23]. This ECRF has recently been modeled by a linear combination of three Gaussians, the one with the widest scale representing an extended disinhibitory surround [24], [25]. Mathematically

$$\mathrm{ECRF}(\sigma_1,\sigma_2,\sigma_3) = A_1\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{x^2}{2\sigma_1^2}} - A_2\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{x^2}{2\sigma_2^2}} + A_3\frac{1}{\sqrt{2\pi}\sigma_3}e^{-\frac{x^2}{2\sigma_3^2}} \quad (5)$$

where ECRF represents the response function for the extended classical receptive field again in a representative 1-D form, $\sigma_1$, $\sigma_2$, and $\sigma_3$ represent the scales of the center, the antagonistic surround, and the extended disinhibitory surround, respectively, and $A_1$, $A_2$, and $A_3$ represent the corresponding amplitudes. It has also been shown that this ECRF can be approximated by a similar but new function [24], [26], [27]. In two dimensions, this is given by

$$\mu\delta(r) + \nabla^2 g(r,\sigma). \quad (6)$$

Here, $\mu$ is an amplitude scale, while $\delta(r)$ represents the 2-D Dirac delta function. It has been shown that this function is more advantageous compared with the LOG in that it can decipher the image structure in the form of light and shade information, even from an edge map [26] and even if that image, and consequently the map, is noisy [27]. In other words, the zero crossings of a noisy image convolved with this function may carry more information as compared with that using the LOG operator. There is another important distinction between these two functions which we are now going to discuss.

When an observer mentally reconstructs a visual scenario in his brain from the available intensity distribution, he is required to do so at different scales. Hence, the edge information should also be extracted in low-level vision at different scales. This is possible according to Marr–Hildreth's scheme [18] by convolving the original intensity distribution with LOG at various $\sigma$. Edges obtained at these different scales then need to be integrated together to form a complete edge map for the observer. The new operator, mentioned earlier [expression (6)], apart from scale, has another variable parameter $\mu$. It has previously been shown that in a zero-crossing map obtained through this operator, $\mu$ behaves as threshold [26]. We shall now try to see if this property can be used to produce effects similar to image thresholding toward isolating a figure from background, in general, provided of course that there is some other operator that performs such grouping. In other words, even if there is no zero-crossing detection, whether in biological vision or in machine vision, we are interested to know if this model of ECRF can produce a figure–ground segregation similar to those obtained through thresholding-like operation. Furthermore, we attempt to unveil if such a computational model for figure–ground segregation can in turn provide cue to brightness perception or, in other words, if such preliminary grouping operation influences brightness–contrast information processing in early vision as claimed by many psychologists [14]–[16]. To this end, we shall first derive a new and more generalized form of ECRF model that leads to such grouping. It has already been shown [26] that the ECRF, which includes both the mean-increasing and mean-decreasing nonlinear subunits [21], can also be beautifully modeled by the function

$$\mu\delta(r) + \nabla^4 g(r,\sigma). \quad (7)$$

The similarity between (6) and (7) in terms of the use of Gaussian derivatives (GDs) tempt us to try and uncover if there can be a generalized approach to such modeling of the ECRF in the light of the well-known linear harmonic oscillator model that may perform our targeted grouping.

Thus, if

$$g(x) = \pi^{-\frac{1}{4}} e^{-\frac{x^2}{2}} \tag{8}$$

represents the standard ground-state eigenfunction of a linear harmonic oscillator Hamiltonian operator, for example, $H$, then other eigenfunctions of $H$ are given by the higher order Hermite functions

$$\psi_n(x) = \pi^{-\frac{1}{4}} 2^{-\frac{n}{2}} (n!)^{-\frac{1}{2}} \left(x - \frac{d}{dx}\right)^n e^{-\frac{x^2}{2}} \tag{9}$$

which satisfy the eigenvalue equation

$$H\psi_n(x) = 2n\psi_n(x). \tag{10}$$

It is also well known that $\{\psi_n(x); n \in N\}$ forms an orthonormal basis for $L^2(\Re)$. If we leave alone the normalization part in (9), then for $n = 2$

$$\left(x - \frac{d}{dx}\right)^n e^{-\frac{x^2}{2}} = x^2 e^{-\frac{x^2}{2}} - 2x\frac{d}{dx}\left(e^{-\frac{x^2}{2}}\right) + \frac{d^2}{dx^2}e^{-\frac{x^2}{2}}. \tag{11}$$

It has also been shown [24], [26] that the HVS is capable of constructing only even-order derivatives by a linear combination of any even smoothing function, for example, $e^{-(x^2/2)}$, as in the present case. In addition, presently, we are not concerned with the fact whether the odd-order derivatives, like the first order one in (11), are computed or not as a linear combination of some odd smoothing function as explained in [26] or otherwise. Thus, for all practical purposes, the middle term on the right-hand side of (11) vanishes so that the eigenfunction without the normalization factor becomes

$$\psi_2(x) = x^2 e^{-\frac{x^2}{2}} + \frac{d^2}{dx^2}e^{-\frac{x^2}{2}} \tag{12}$$

Following the same procedure, without the normalization factor and by neglecting the odd-order derivative terms

$$\psi_4(x) = x^4 e^{-\frac{x^2}{2}} + 6x^2\frac{d^2}{dx^2}e^{-\frac{x^2}{2}} + \frac{d^4}{dx^4}e^{-\frac{x^2}{2}}. \tag{13}$$

Here, an inner scale $\sigma$ may be introduced by substituting $x$ with $x/\sigma$, whereby one may reproduce similar expressions for a GD family. Young [28], [29] has extensively used the GD family for modeling the receptive-field structures in HVS.

It has also been argued elsewhere [24], [26] that for computing the right-hand side in (12), for example, after introduction of scale, one need not consider the scales for the two terms to be of the same order. Therefore, if the Gaussian for the first term is assumed to be very local and that for the second is averaging over a comparatively larger area, then the first term, in the limiting case, would tend to a Dirac delta function. Thus, in two dimension, as $\sigma \to 0$, $\psi_2(r, \sigma) \to \{\text{expression (6)}\}$. Once, $\psi_2(r)$ has been calculated by computing the LOG, one may now assume the Gaussian for the middle term in (13) to be the local (it has also been shown previously [24], [26] that the LOG can indeed be approximated in the limit by a delta function) so that for the computation of $\psi_4(r)$, one arrives at a more generalized version of the ECRF model compared with (7), i.e., in two dimensions

$$\mu\delta(r) + \nabla^4 g(r, \sigma) + 2\pi\sigma^6\eta r^4 g(r, \sigma). \tag{14}$$

Here, $2\pi\sigma^6$ is the normalization factor for $\nabla^4 g(r, \sigma)$. Therefore, the higher order eigenfunction that involves a higher order GD (bi-Laplacian of Gaussian), according to this generalized approach, brings with it a new parameter $2\pi\sigma^6\eta = \nu$ (for example) and a new moment term (the third term) as compared with (7).

The same expression could also have been derived alternatively even without directly approximating the scaled middle term in (13) to a delta function. From (13)

$$\psi_4(x) = x^4 e^{-\frac{x^2}{2}} + 6x^2(x^2 - 1)e^{-\frac{x^2}{2}} + \frac{d^4}{dx^4}e^{-\frac{x^2}{2}}$$

$$\psi_4(x) = 7x^4 e^{-\frac{x^2}{2}} - 6x^2 e^{-\frac{x^2}{2}} + \frac{d^4}{dx^4}e^{-\frac{x^2}{2}}. \tag{15}$$

Now, applying the same approximation that was applied to the scaled equation (12), one may, in two dimensions, easily arrive at the same form of (14), only with slightly different coefficients.

In essence, the Dirac delta function here undergoes a linear combination with the fourth-order derivative and the fourth moment of the Gaussian function. This is our proposal for a new computational model of figure–ground segregation whose efficacy in figure–ground extraction will be presently examined. It is also to be seen how such a model of figure–ground segregation is indeed capable of predicting brightness perception in early vision.

## III. Results and Discussion

The results have all been obtained by carrying out convolution operation of the proposed spatial filter [(14)] on different images. The convolution operation involves the following steps.

1) Take an image $I(x, y)$.
2) Take the filter function $f(x, y)$.
3) Make the operation: $P(x, y) = \iint f(x - \alpha, y - \beta)I(\alpha, \beta)d\alpha d\beta$ in discrete domain to obtain the processed image $P(x, y)$.

For all the examples shown in the following (Figs. 1–9), the standard deviation of the Gaussian function has been chosen as $\sigma = 1$, and the proposed function has throughout been sampled between $\pm 3\sigma$ at a sampling rate of 0.1. The value of $\mu$ has been kept constant throughout at $10^4$ and that of $\nu$ at seven. What is to be noted is that no additional thresholding is required to arrive at the desired figure–ground segregation. This has first been demonstrated with two standard images in Fig. 1. The proposed model is found to identify many visually distinct regions in the images. For example, in the stapler-on-table image [Fig. 1(a)], we find that our figure–ground separation algorithm has managed to separately identify many such regions in the resultant binary image [Fig. 1(c)]. The results are even more

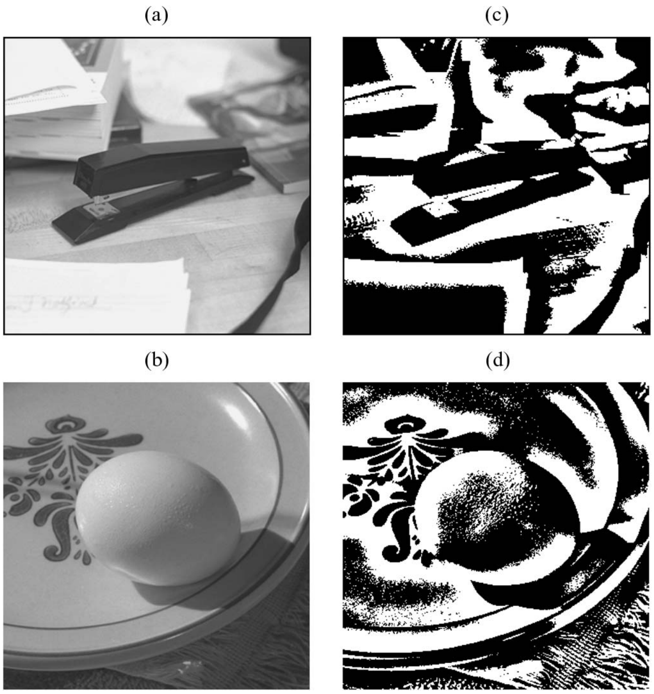



Fig. 1. (a) and (b) Images of stapler-on-table and egg-on-plate. (c) and (d) Figure–ground segregation has been performed by convoluting the images with the proposed filter. The proposed model is found to identify many visually distinct regions in the images.


Fig. 2. (a) SBC illusion where, of the two equiluminant gray patches, the one on lighter background appears darker than the other. (b) The proposed filter produces a figure–ground grouping that isolates both the patches and as perceived visually; the darker one on white background and the brighter one on dark.

prominent for the egg-on-plate image [Fig. 1(b)]. Here, not only are the important features, like the carpet beneath the plate identified, but also even the main figure (egg), which is difficult to identify from the background (plate) because of the absence of any sharp transition in intensity levels at many locations, has gotten clearly segregated, although the designs on the plate have gotten slightly messy at places [Fig. 1(d)]. Thus, in a sense, the model is a valid bottom-up approach to figure–ground segregation. Therefore, the target is now to see whether the proposed computational model of figure–ground grouping, which is supposed to be a comparatively higher level phenomenon, is also capable of influencing a supposedly lower level phenomenon of brightness perception. To this end, some brightness-perception-based illusions have been used as stimuli. The first among these is the simultaneous brightness–contrast (SBC) illusion (Fig. 2), which has been generally explained by lateral inhibition alone. Here, two equiluminant test patches appear darker and brighter, respectively, depending upon whether the background is bright or dark.

The results show that the test patches have not only been identified as figures with respect to their backgrounds but also predictions have been made about their perceived brightness by the same computational model. Thus, the test patch on white background appears black upon convolution, while that on dark appears bright. Even when the SBC illusion is presented in a little more complicated form, as in Fig. 3, the computational model is equally successful in predicting brightness perception while performing figure–ground segregation.

Next, similar results have been reproduced for another illusion of this class commonly explained through lateral inhibition, viz., the grating induction (Fig. 4), where a test-patch equiluminant throughout appears to have been inducted by a sinusoidal grating effect from its background, although in spatially opposite phase. Such a perceptual effect has been captured by our model in the convoluted image in the sense that the sinusoidal grating has been represented as a square grating whose effect has been inducted into the luminance patch and indeed in opposite phase, corroborating with our perceptual experience with respect to brightness.

The next example is of a real square grating which is in fact a far more complex example of brightness perception, where brightness assimilation dominates over brightness contrast [30], [31]. This is the White-effect illusion [32] (Fig. 5) where the left-hand patch appears darker in spite of being flanked by a greater amount of darker neighborhood—a perception in opposition to the SBC, i.e., lateral-inhibition effect. Here, brightness appears to have assimilated from surround rather than an induction in contrast mode. It is found that our proposed computational model manages to locate the test patches from the background and have, at the same time, captured the perceptual signature of the reverse direction of brightness induction.

The lateral-inhibition model was first utilized by Marr and Hildreth [18] to extract edges. The present model also produces important edge information and through the next example, the importance of such edges in figure–ground segregated outputs toward making prediction on brightness perception is demonstrated. To serve this purpose, an image has been chosen, whose horizontal line profile at any position is nothing but a staircase function (Fig. 6). It is so understood that the edges are the lines of demarcation at positions of strong intensity discontinuities. Here, it is found that in the convoluted image, the edges that appear not only demarcate the individual grayscale segments but also border them to an increasingly larger extent as the intensity shifts toward the brighter side, with the brightest one being finally bordered on all sides.

The importance of such borders will become more prominent in the following example (Fig. 7) of Todorovic effect [16]. This may be looked upon as the SBC illusion made complex by occluding the test patch on white background with four dark squares, and vice versa. For this reason, the separation of the object (i.e., the patch) from the background in a binary output becomes even more difficult as compared with SBC. Here, again, one notices the importance of the edgelike borders in the output image, which not only isolate the patches (objects) from the complex background but also provide clue to brightness perception. Thus, the black border may be interpreted as a signifier to dark perception and the white border to bright perception,

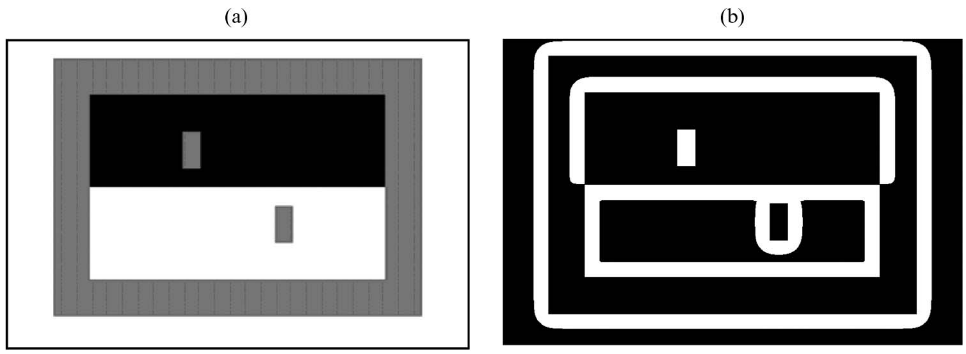



Fig. 3. (a) This is a more complicated version of the SBC illusion. (b) The proposed filter produces a binary image that shows both the patches segregated from background and as perceived visually with respect to their brightness.

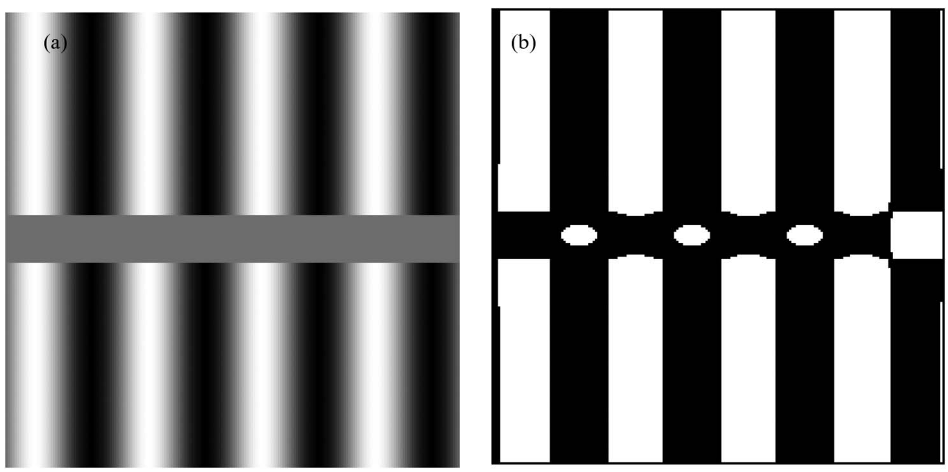



Fig. 4. (a) Grating-induction illusion where a gray patch equiluminant throughout is inducted with the sinusoidal grating effect of its background and in opposite phase. (b) The proposed filter produces a binary image that mimics the brightness induction into the test patch as in a square grating.

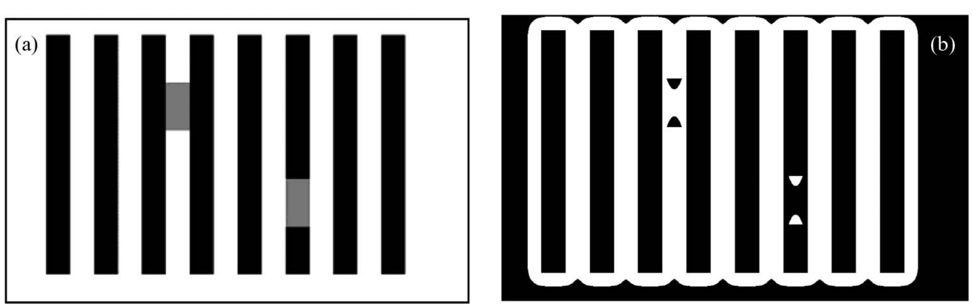



Fig. 5. (a) White-effect illusion on a square grating where, of the two equiluminant gray patches, the one flanked by larger amount of lighter background appears brighter than the other which is in opposition to SBC perception. (b) The proposed filter produces a binary image that shows both the patches segregated; the direction of brightness induction being as perceived visually, with the darker one with dark flanking bars and the brighter one with bright flanking bars.

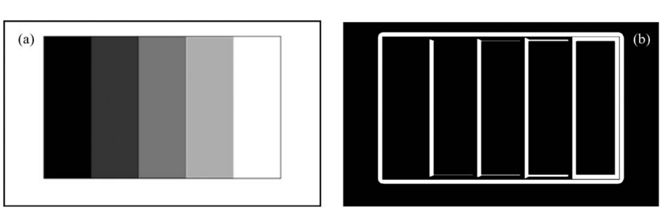



Fig. 6. (a) An image with five distinct gray levels with any horizontal profile representing a staircase function. (b) The proposed filter produces a binary image that shows the five different segments in black, separated by white band-like edges. If we investigate the nature of these edges, we find that they carry a clear-cut signature of the increase in gray level from left to right through an increase in the number and extent of these separating white borders, with the brightest segment being finally bordered on all sides.


Fig. 7. (a) Todorovic illusion. Here, two equally illuminant gray test patches are occluded by four black and four white squares on the left and right panels, respectively. (b) The proposed filter produces a binary image that separates the test patches from the background through borders whose intensities indicate the brightness perception in the test patches—a black border implies darker perception, and a white border implies a brighter perception.

as has been reported by psychophysical observations [31]. This phenomenon too, like the White effect, appears to be the consequence of brightness assimilation from the occluding


Fig. 8. (a) Complex version of Todorovic effect. Here, the two equally illuminant gray test patches have been occluded by the four black and four white squares on the left and right panels, respectively, in such a way that the visual system identifies a cross each on the black and white windows placed on opposite backgrounds and perceived darker and brighter, respectively. (b) The proposed filter produces a binary image that identifies the crosses, with the shortages being filled up by borders whose intensities indicate the brightness perception of the crosses; a black fill in implies dark perception, and a white fill in implies bright perception.


Fig. 9. (a) Another complex version of Todorovic effect. Here, the two equally illuminant crosses extend beyond the occluding black and white squares on the left and right panels, in such a way that the visual system identifies a cross each on black and white windows placed on opposite backgrounds and perceived darker and brighter, respectively. (b) The proposed filter produces a binary image that identifies the crosses, with the shortages being filled up by borders whose intensities indicate the brightness perception of the crosses; a black fill in implies dark perception, and a white fill in implies bright perception.

squares. Therefore, a dark border may be interpreted as signifier of darkness assimilation and bright borders as signifier of brightness assimilation.

Finally, the results of the two very complex versions of the Todorovic effect which are directly related to higher level grouping are shown. These stimuli are assumed to be beyond the scope of any spatial-filtering-based approach even by those researchers working with mathematical-model-based explanations to brightness-perception illusions [31].

Here, the test patches in Figs. 8(a) and 9(a) can be perceived as cross, coplanar with and wedged between four abutting squares. In the first of these, the arms of the test patches exactly match the perimeters of the superimposed squares, while in the second, the arms extend beyond the superimposed squares. Applying the proposed computational model to extract figure from ground, it is found that although the test patches have again been successfully isolated from the background, in both the cases they appear shorter than actually perceived. This again is similar to the appearance of the supposedly meaningful fictitious borders that we saw in Figs. 6 and 7. Therefore, continuing with a similar line of argument, we conclude that when the shortages in the four ends of the crosses in both the filtered outputs [Figs. 8(b) and 9(b)] are filled with dark border, it implies darker perception, and when filled with brighter border, it implies brighter perception. Such an interpretation corroborates with observations from the psychophysical experiments [31] regarding actual perception.

Finally, we provide a 1-D profile of the modified ECRF function as given in (14) by means of which we have performed all the convolutions whose results have been presented previously. This has been shown in Fig. 10(b) along with the 1-D plot of the classical receptive field in Fig. 10(a).

## IV. Conclusion

It is true that the structuralist school, in spite of it forming the basis of materialistic and scientific viewpoints, suffers from the shortcoming of completely ignoring the role of emergent behaviors of complex systems so that "part" alone is assumed important and the "whole" often loses its legitimate role. Gestalt and the constructivist views attempt in their own way to overcome this shortcoming but often without any firm footing on actual bottom-up computational models, thus leading to undesirable mysticisms. This paper may be considered to be a humble attempt toward setting up a dialogue between the positive aspects imbibed in all these views toward strengthening the scientific materialist viewpoint. An attempt has been made to understand some of the aspects of visual signal processing through a model that tries to incorporate the views of micro (edges, brightness induction) as well as macro levels (grouping, figure–ground segregation). We had tried to emphasize upon the point that any unconscious inference or perceptual grouping at high level should also have a low-level computational basis and in turn, should be integrated top-down with the bottom-up processes only through such computational linkages. Although this is nothing new, sometimes it is worth iterating the views of someone like Barlow [33]:

"A description of that activity of a single nerve cell, which is transformed to and influences other nerve cells, and of a nerve cell's response to such influences from other cells, is a complete enough description for functional understanding of the nervous system. There is nothing else "looking at" or controlling this activity, which must therefore provide a basis for understanding how the brain controls behavior" [33]—which means that there is no "soul" sitting anywhere and interpreting things from the neuronal outputs, but rather, it is a collective step-by-step synchronization of the outputs at various stages in the eye and the brain, no matter how complex that process is, that ultimately creates the visual perception of the world around us. Using a new computational model based on the concept of the ECRF and generalizing it by an eigenfunction-based approach, it has been shown in this paper that through such a possible computational model, not only can one achieve figure–ground segregation but, as a follow-up top-down action, can also predict the direction of brightness induction; a comparatively more fundamental task according to some. While not directly refuting the Gestalt viewpoint on perceptual grouping, this paper can at least claim to have put forth a computational model that can complement the Gestalt viewpoint. In this sense, this paper attempts to establish a computational basis even to those cases of complex stimuli where the Gestalt grouping factors, viz., *Proximity*, *Good Continuation*, *Common Fate*, and *Similarity*, along with other grouping factors, including *Coplanarity* [34], T-junctions, and X-junctions [35], all supposedly become important. The results on real photographs, as shown in Fig. 1, encourage further investigation in the direction of exploring the efficacy of such an algorithm in the domain of image segmentation in general or particular [36], by incorporating these type of visual cues [37]. However, the actual motivation behind this paper has been to strengthen the school of thought where recurrent interplay between bottom-up

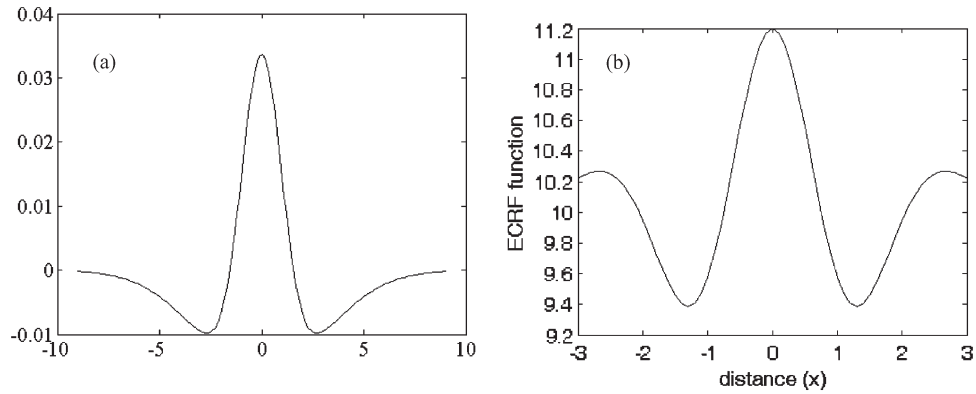



Fig. 10. (a) One-dimensional profile of the well-known classical receptive field represented by the DOG function [(4)]. The LOG function [(2)] is also of the same nature. (b) One-dimensional profile of the modified ECRF function as given in (14) by means of which all the convolutions in Figs. 1–9 have been performed.

and top-down processing holds the key to visual perception [38]. While the LOG model of Marr–Hildreth has been extended here in the light of the extended receptive field, future investigation may be carried out to assess the implication of still higher order derivatives and/or moments toward higher level perceptual groupings and consequent integration into more complete pictures of brightness induction, edge linking, and so on, many of which may have potential applications like, for example, the concept of edge brightness dealt with in this paper that has already been used in visual anticamouflage of moving objects [39].

## ACKNOWLEDGMENT

The first author would like to acknowledge the influence of the thoughts and works of K. Marx, F. Engels, V. I. Lenin, J. Stalin, and Mao Tse Tung on the basic idea presented in this paper. Their understanding of dialectical materialism in nature and society (particularly Lenin's criticism of E. Mach's empiricist philosophy, discussed in [1], in his book *Materialism and Empirio-criticism*) has led to realizing the shortcomings of both the bottom-up (mechanical materialist) and the top-down (idealist) viewpoints, and brought about the urge to work toward developing a dialectical understanding of the brain, in general, and the visual system, in particular. This paper was done when the second author S. K. Pal was a J.C. Bose Fellow of the Government of India.

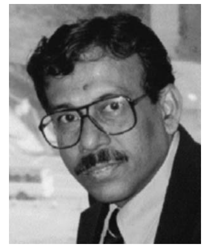
## REFERENCES

[1] F. Ratliff, *Mach Bands: Quantitative Studies on Neural Network in the Retina*. San Francisco, CA: Holden-Day, 1965.
[2] H. von Helmholtz, *Helmholtz's Treatise on Physiological Optics*. New York: Opt. Soc. Amer., 1962.
[3] R. L. Gregory, *The Intelligent Eye*. New York: Weidenfeld and Nicolson, 1970.
[4] J. R. Anderson, *Cognitive Psychology and Its Implications*, 6th ed. New York: Worth Publishers, 2005.
[5] K. Koffka, *Principles of Gestalt Psychology*. Orlando, FL: Harcourt Brace Jovanovich, 1935.
[6] W. N. Dember, *Visual Perception: The Nineteenth Century*. Hoboken, NJ: Wiley, 1964.
[7] G. Kanizsa, *Organization in Vision*. New York: Praeger, 1979.
[8] C. von der Malsburg and J. Buhmann, "Sensory segmentation with coupled neural oscillators," *Biol. Cybern.*, vol. 67, no. 3, pp. 233–242, 1992.
[9] S. Zeki and A. Bartels, "Toward a theory of visual consciousness," *Conscious. Cogn.*, vol. 8, no. 2, pp. 225–259, Jun. 1999.
[10] D. George, "How the brain might work: A hierarchical and temporal model for learning and recognition," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, 2008.
[11] T. Poggio, T. Serre, C. Tan, and S. Chikkerur, "An integrated model of visual attention using shape-based features," MIT, Cambridge, MA, CBCL-278, MIT-CSAIL-TR-2009-029, 2009.
[12] M. Heidelberger, *Nature From Within: Gustav Theodor Fechner and His Psychophysical Worldview*. Pittsburgh, PA: Univ. Pittsburgh Press, 2004.
[13] I. Rock, *Indirect Perception*. Cambridge, MA: MIT Press, 1997.
[14] T. Agostini and D. R. Proffitt, "Perceptual organization evokes simultaneous lightness contrast," *Perception*, vol. 22, no. 3, pp. 263–272, 1993.
[15] A. Gilchrist and F. Bonato, "Anchoring of lightness values in center/surround displays," *J. Exp. Psychol.: Human Perception Perform.*, vol. 21, no. 6, pp. 1427–1440, 1995.
[16] D. Todorovic, "Lightness and junctions," *Perception*, vol. 26, no. 4, pp. 379–395, 1997.
[17] D. Marr, *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. New York: Freeman, 1982.
[18] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. Lond. B*, vol. 207, no. 1167, pp. 187–217, Feb. 1980.
[19] S. Hochstein and H. Spitzer, "Zero-crossing detectors in primary visual cortex," *Biol. Cybern.*, vol. 51, no. 3, pp. 195–199, Dec. 1984.
[20] J. Richter and S. Ullman, "Non-linearities in cortical simple cells and the possible detection of zero-crossings," *Biol. Cybern.*, vol. 53, no. 3, pp. 195–202, Jan. 1986.
[21] C. L. Passaglia, C. Enroth-Cugell, and J. B. Troy, "Effects of remote stimulation on the mean firing rate of cat retinal ganglion cell," *J. Neurosci.*, vol. 21, no. 15, pp. 5794–5803, Aug. 2001.
[22] C. Sun, X. Chen, L. Huang, and T. Shou, "Orientation bias of the extraclassical receptive field of the relay cells in the cat's dorsal lateral geniculate nucleus," *Neuroscience*, vol. 125, no. 2, pp. 495–505, 2004.
[23] A. Angelucci and P. C. Bressloff, "Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons," *Progr. Brain Res.*, vol. 154, pp. 93–120, 2006.
[24] K. Ghosh, S. Sarkar, and K. Bhaumik, "A possible mechanism of zero-crossing detection using the concept of extended classical receptive field model of retinal ganglion cells," *Biol. Cybern.*, vol. 93, no. 1, pp. 1–5, Jul. 2005.
[25] K. Ghosh, K. Bhaumik, and S. Sarkar, "Retinomorphic image processing," *Progr. Brain Res.*, vol. 168, pp. 175–191, 2008.
[26] K. Ghosh, S. Sarkar, and K. Bhaumik, "Understanding image structure from a new multi-scale representation of higher order derivative filters," *Image Vis. Comput.*, vol. 25, no. 8, pp. 1228–1238, Aug. 2007.
[27] K. Ghosh, S. Sarkar, and K. Bhaumik, "A possible mechanism of stochastic resonance in the light of an extra-classical receptive field model of retinal ganglion cells," *Biol. Cybern.*, vol. 100, no. 5, pp. 351–359, May 2009.
[28] R. A. Young, "The Gaussian derivative model for spatial vision: I. Retinal mechanisms," *Spat. Vis.*, vol. 2, no. 4, pp. 273–293, 1987.

[29] R. A. Young, R. M. Lesperance, and W. W. Meyer, "The Gaussian derivative model for spatial-temporal vision: I. Cortical model," *Spat. Vis.*, vol. 14, no. 3/4, pp. 261–319, 2001.
[30] B. Blakeslee and M. E. McCourt, "Similar mechanisms underlie simultaneous brightness contrast and grating induction," *Vision Res.*, vol. 37, no. 20, pp. 2849–2869, Oct. 1997.
[31] B. Blakeslee and M. E. McCourt, "A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction," *Vision Res.*, vol. 39, no. 26, pp. 4361–4377, Oct. 1999.
[32] M. White, "The effect of the nature of the surround on the perceived lightness of gray bars within square-wave test gratings," *Perception*, vol. 10, no. 2, pp. 215–230, 1981.
[33] H. B. Barlow, "Single units and sensation: A neuron doctrine for perceptual psychology?" *Perception*, vol. 1, no. 4, pp. 371–394, 1972.
[34] A. Gilchrist, "When does perceived lightness depend on perceived spatial arrangement," *Percept. Psychophys.*, vol. 28, no. 6, pp. 527–538, Dec. 1980.
[35] A. Gilchrist, C. Kossifidis, F. Bonato, T. Agostini, J. Cataliotti, X. Li, B. Spehar, V. Annan, and E. Economou, "An anchoring theory of lightness perception," *Psychol. Rev.*, vol. 106, no. 4, pp. 795–834, Oct. 1999.
[36] C. C.-F. Juang, C. C.-M. Chang, J. J.-R. Wu, and D. D. Lee, "Computer vision-based human body segmentation and posture estimation," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 1, pp. 119–133, Jan. 2009.
[37] P. L. Rosin, "Shape partitioning by convexity," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 30, no. 2, pp. 202–210, Mar. 2000.
[38] S. Hochstein and M. Ahissar, "View from the top: Hierarchies and reverse hierarchies in the visual system," *Neuron*, vol. 36, no. 5, pp. 791–804, Dec. 2002.
[39] H. Lu, X. Wang, S. Liu, M. Shi, and A. Guo, "The possible mechanism underlying visual anticamouflage: A model and its real-time simulation," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 29, no. 3, pp. 314–318, May 1999.

**Kuntal Ghosh** received the B.Sc. and M.Sc. degrees in physics from the University of Calcutta through the Presidency College, Kolkata, India, and Ph.D. degree in science from Jadavpur University, Kolkata, while he was with the Microelectronics Division, Saha Institute of Nuclear Physics, Kolkata, India.

He is currently an Assistant Professor at the Machine Intelligence Unit and also a deputed faculty at the Center for Soft Computing Research in Indian Statistical Institute, Kolkata. His area of interest is visual perception and brain sciences. He has published a few papers in this domain in international journals and conferences and has coedited a book in Bengali language on the history of science.



**Sankar K. Pal** (M'81–SM'84–F'93) received the Ph.D. degree in radio physics and electronics from the University of Calcutta, West Bengal, India, in 1979, and the Ph.D. degree in electrical engineering, along with DIC, from Imperial College, University of London, London, U.K., in 1982.

He was with the University of California, Berkeley and the University of Maryland, College Park, in 1986–1987; the NASA Johnson Space Center, Houston, TX, in 1990–1992 and 1994; and with the U.S. Naval Research Laboratory, Washington DC, in 2004. He is currently the Director and a Distinguished Scientist with the Indian Statistical Institute, Kolkata, India. He is also currently a J.C. Bose Fellow of the Government of India. He founded the Machine Intelligence Unit and the Center for Soft Computing Research: a national facility in the Institute in Kolkata. He is a coauthor of fourteen books and more than three hundred research publications in the areas of pattern recognition and machine learning, image processing, data mining and web intelligence, soft computing, neural nets, genetic algorithms, fuzzy sets, rough sets and bioinformatics.

Prof. Pal is a fellow of the Academy of Sciences for the Developing World (TWAS), Italy, International Association for Pattern Recognition, U.S., International Association of Fuzzy Systems, U.S., and all the four National Academies for Science/Engineering in India. Since 1997 he has been serving as a Distinguished Visitor of IEEE Computer Society (U.S.) for the Asia–Pacific Region, and held several visiting positions in Hong Kong and Australian universities. He was the recipient of the *1990 S.S. Bhatnagar Prize* (which is the most coveted award for a scientist in India) and many other prestigious awards in India and abroad including the 1999 G.D. Birla Award, 1998 Om Bhasin Award, 1993 Jawaharlal Nehru Fellowship, 2000 Khwarizmi International Award from the Islamic Republic of Iran, 2000–2001 FICCI Award, 1993 Vikram Sarabhai Research Award, 1993 NASA Tech Brief Award (U.S.), 1994 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award (U.S.), 1995 NASA Patent Application Award (U.S.), 1997 IETE-R.L. Wadhwa Gold Medal, the 2001 INSA-S.H. Zaheer Medal, 2005–2006 ISC-P.C. Mahalanobis Birth Centenary Award (Gold Medal) for Lifetime Achievement, 2007 J.C. Bose Fellowship of the Government of India, and the 2008 Vigyan Ratna Award from Science and Culture Organization, West Bengal. He is/was an Associate Editor of IEEE TRANSACTION ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (2002–2006), IEEE TRANSACTION ON NEURAL NETWORKS (1994–1998 and 2003–2006), *Neurocomputing* (1995–2005), *Pattern Recognition Letters*, *International Journal of Pattern Recognition and Artificial Intelligence*, *Applied Intelligence, Information Sciences, Fuzzy Sets and Systems*, *Fundamenta Informaticae*, *LNCS Transactions on Rough Sets*, *International Journal of Computational Intelligenceand Applications*, *IET Image Processing*, *Journal of Intelligent Information Systems*, and *Proceedings of INSA-A*. He is Editor-in-Chief for the *International Journal of Signal Processing*, *Image Processing and Pattern Recognition*; a Book Series Editor for *Frontiers in Artificial Intelligence and Applications* (Amsterdam, The Netherlands: IOS Press), and *Statistical Science and Interdisciplinary Research* (Singapore: World Scientific). He is a member of the Executive Advisory Editorial Board of the IEEE TRANSACTIONS ON FUZZY SYSTEMS, *International Journal of Image and Graphics* and *International Journal of Approximate Reasoning*; and a Guest Editor of *IEEE Computer*.