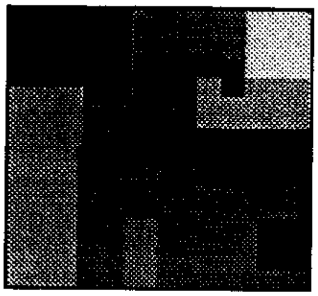
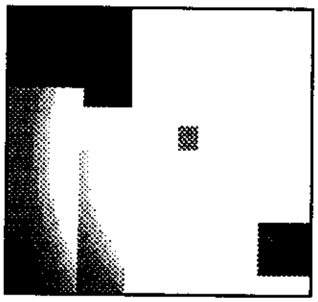
Computer Science Department Technical Report
Machine Perception Laboratory
University of California
Los Angeles, CA 90024-1596

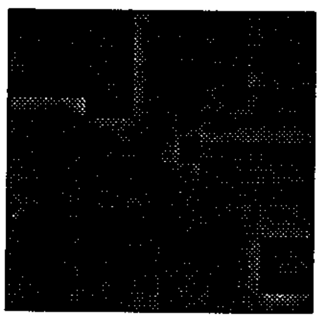
LIGHTNESS CONSTANCY ARCHITECTURE - THEORY AND SIMULATION

Josef Skrzypek

December 1989
CSD-890067







MPL

Machine Perception Lab

UCLA Computer Science Department

Lightness Constancy Architecture - Theory and Simulation

Josef Skrzypek

TR 89-5

Grant ATI-PMTC#15/03-C768

MPL

Machine Perception Lab

top-down

bottom-up

Lightness Constancy Architecture - Theory and Simulation

Josef Skrzypek

Machine Perception Laboratory
Computer Science Department
University of California
Los Angeles, CA 90024

March 31, 1989

Contents

1 Introduction **3**

1.1 What is lightness constancy ? 3

1.2 How does Biological Vision System Achieve Lightness Constancy? 4

1.3 Current Vision Systems' Shortcomings 5

2 Computational Algorithms for Achieving Lightness Constancy **6**

2.1 Retinex Theory 6

2.1.1 Land and McCann's 1-D approach 6

2.1.2 Horn's 2-D approach 7

2.1.3 Blake modified Horn approach 9

2.1.4 Summary of Retinex Theory Approach and their Major Problems 10

2.2 Intensity Dependent Model 10

2.3 Grossberg's BCS and FCS 12

2.4 UCLA-MPL Approach 14

2.4.1 Introduction 14

2.4.2 Architecture Description 17

2.4.3 Simulations in SFINX 19

2.4.4 Discussions 22

3 Summary **23**

1 Introduction

As autonomous mobile robots maneuver through unknown terrain under unpredictable and complicated illumination condition, their visual sensory requirements must be much more sophisticated than the visual sensory of industrial robots. For example, a cloud blocking the sun will cast a shadow over the scenery. The changes in illumination can be many orders of magnitude variations in intensity and these changes happen in seconds. Currently, vision systems used for manufacturing environment are limited in their ability to operate over 1 or 2 orders of magnitude of light intensity. Often in manufacturing environments, huge amount of money is spent on the light structure to ensure that the scene is evenly illuminated and within the operating range of the imaging sensors. However, in the natural environments it is impossible to control the lighting condition which is changing continuously. By contrast, humans can function in normal environment without a priori knowledge of the lighting condition of their surroundings and in general are able to perceive almost the same scene in daylight as at night. Vertebrate retina is able to operate over changing lighting intensity of 10 orders of magnitude. This ability of maintaining high sensitivity regardless of local or global ambient light level can be viewed as part of the lightness constancy function.

1.1 What is lightness constancy ?

Lightness constancy is often refered to as brightness constancy. Strictly speaking, the two are different phenomena. Brightness refers to labeling surfaces as dim or bright and as such, it involves high level visual functions which use memory and attentive process. For example, we perceive a piece of white paper in a sunlight as bright and in a shadow as dim . On the other hand, lightness constancy allows us to consistently assign the label white to the same piece of paper regardless of the amount of ambient light. Unfortunately, often the two terms are used interchangeably in the literature.

In analyzing the problem of perception, Helmholtz described the brightness constancy phenomenon as "...A grey sheet of paper exposed to sunlight may look brighter than a white sheet in the shade; and yet the former looks grey and the latter white, simply because we **know** very well that if the white paper were in the sunlight, it would be much brighter than the grey paper which happens to be there at the time. ... Colors have their **greatest** significance for us insofar as they are properties of bodies and can be used as marks of identification of bodies. Hence in our observations with the sense of

[1] *Acknowledgments:* Support for the UCLA Machine Perception Laboratory environment is provided in part by generous equipment grants from IBM and Hewlett Packard. We sincerely acknowledge support by ARCO-UCLA Grant #1, MICRO-TI grant #541122-19907, MICRO-Hughes Research Laboratories grant #541122-57442, ONR grant #N00014-86-K-0395, and ARO grant #444251-25519.

vision we always start out by forming a judgment about the colors of bodies, eliminating the differences of illumination by which a body is revealed to us" [9]. This suggests that the human visual system does not measure brightness of an object, but the surface reflectance of the object, which is invariant under different illumination conditions. Some learning process might be involved in the determination of the color or brightness of a surface. Hering also thought that experience plays an important role in color or brightness labeling, "...all objects that are already known to us from experience or that we regard as familiar by their color, we see through the spectacles of memory color, and on that account quite differently from the way we would otherwise see them " [10]. The Gestalt psychologists, like Kohler, Gibson, and Wallach subscribe to the theory that human visual systems perceive an invariant property of objects [13,4,23]. Wallach suggested that luminance ratio is an invariant attribute of any scene. When the overall level of illumination on an array of surfaces is changed, the luminance of a particular surface is indeed altered, but since all luminances are altered proportionately, the luminance ratio of any one surface relative to any other in the total array remains constant.

The luminance ratios are the relative reflectances of the objects surfaces. Every surface has a reflectance, or the albedo coefficient. It is the reflectance of a surface that conveys the information about that surface. However, our eyes cannot measure this reflectance directly. The surfaces have to be illuminated and the reflected light carries the reflectance information about the surface to our eyes. Thus, lightness constancy is essential for human visual systems to perceive a stable world under different illumination conditions.

From physics, the light intensity at a point in the image is the product of the reflectance at the corresponding object point and the intensity of illumination at that point, aside from a constant factor that depends on the optical arrangement. The term lightness is psychophysical correspondence of reflectance. An important distinction between the two components is that they differ in their spatial distribution. Incident light intensity or illumination will usually vary smoothly, with no discontinuities, while reflectance or lightness will have sharp discontinuities at edges where objects adjoin. The reflectance is relatively constant between such edges [14].

1.2 How does Biological Vision System Achieve Lightness Constancy?

In terms of physiology, significant aspect of lightness constancy can be explained as a result of lateral inhibition in retinal cells. In all vertebrates, the retina is constructed by five basic types of cells [24]. Within the retina, information is carried by major pathways. The main path starts with photoreceptors (rods and cones) that transduce light to electrical signals. Photoreceptors synapse onto bipolar cells which in turn send their input to ganglion cells. These cells output their signals through the bundle of optic nerve fibers that connect retina to visual cortex in the brain.

The horizontal cells and amacrine cells provide the lateral pathway for interactions with the main path. Horizontal cells synapse onto bipolars, photoreceptors and other horizontal cells. Amacrine cells, another lateral pathway, form synaptic connections with bipolars, ganglion cells and other amacrine cells. It is known that Necturus (amphibian) cones have response range of at most 3.5 log units [15]. More recently it was shown that cones were able to shift their response ranges along the intensity axis when presented with different levels of surround illuminations [19]. It is this shifting characteristic of the receptors which resulted in the effective operating intensity domain of the retina to span over 10 log units. It has been suggested that perhaps horizontal cells modulated the response range of the receptors to have maximum contrast sensitivity at the prevailing light conditions [22]. Similar shifting characteristic is also evident in the bipolar cells' responses [24]. As a result, in the retina the response range of the cells is adjusted to fit the ambient light intensity. This makes the retina a high contrast, wide operating domain sensory array that measures relative light intensity. In the process of measuring relative light intensity, the retina "discounts the illuminant" in real time and thus achieves one aspect of lightness constancy.

1.3 Current Vision Systems' Shortcomings

A typical machine vision system uses CCD (CID) camera. The camera basically controls the amount of light that is coming through and focuses it onto the CCD array. The CCD array then produces a logarithmic (maybe linear) response proportional to the incoming light intensity. There are inherent physical limitations of this type of system. First, the aperture and shutter of a camera control only the global light intensity over the entire image. In contrast, the vertebrate retina is able to have local light intensity adjustment. Another limitation of CCD devices is a limited response range which in turn limits the domain of intensities. Current CCD can only operate in 2 log units and extendible by one or two log units with the adjustment of aperture and shutter. The biological systems took care of this by local adjustment of photoreceptor's operating point. For example, in a single scene, two areas in bright sunlight and dark shadow may span 3-5 log units of intensity. Clearly, ganglion cells which normally span only 0.1 log units of intensity must be continuously shifted in intensity domain to be incorrespondence with prevailing ambient light level. Thus, the key difference between biological receptors and CCD is the ability of biological receptors to have individual biasing points while the CCD must be biased globally for the entire array of sensors during the time of fabrication.

2 Computational Algorithms for Achieving Lightness Constancy

2.1 Retinex Theory

Using Mondrian image (a random arrangement of rectangular patches of different gray levels) as a stimulus for psychophysical experiments, Land developed the retinex theory about color and lightness. The basic assumption is that there are three independent cone systems, each starting with a set of receptors most sensitive to, respectively, the long-, middle-, and short-wavelength regions of the visible spectrum. Each system forms a separate image of the world in terms of lightness that shows a strong correlation with reflectance within its particular band of wavelengths [14]. Land has pointed out that changes due to illuminant are on the whole gradual, appearing usually as smooth illumination gradients, whereas those due to changes in reflectance tend to be sharp. This important observation forms the basis of Land's One-Dimensional and Horn's Two-Dimensional algorithms for lightness computation.

2.1.1 Land and McCann's 1-D approach

Land and McCann [14] in their "two square and a happening" experiment discovered that the change of luminance at the junction between areas both constituted an edge and also led to the visual difference between the whole two areas. This led them to rediscover that our visual sensors measure the ratio between any two adjacent points; Wallach in 1948 had suggested the same ratio principle. In fact, the simple procedure of taking the ratio between two adjacent points can both detect an edge and eliminate the effect of nonuniform illumination.

The mathematical model is as follow. For point a, the luminance is $I_a \times R_a$, where I_a is illumination and R_a is the reflectance.. The ratio between two points a and b then becomes $\frac{(I_a \times R_a)}{(I_b \times R_b)}$. Assuming the illumination changes on two adjacent points are minimal, i.e. I_a equals I_b, then the ratio of luminance reduces to the ratio for reflectance. Therefore, in a Mondrian world, regardless of illumination condition, only detectors along the edges will produce a ratio that is not unity. Land and McCann in their Retinex algorithm, chose a random path across the Mondrian world along which they compute relative reflectance of the patches. This is done by multiplying the ratios along the path and since the ratios within a patch is near unity, the multiplication simplified to products of edge ratios. However, this model only computes reflectance ratios, a reference must be chosen so that an unique value can be assigned to each patch in the Mondrian world.

In order to correctly label all patches with respect to some reference value, Land and

McCann propose to normalize the image function. They assume that the highest value of lightness corresponds to white. To compute lightness of the Mondrian world they use a sufficient number of randomly chosen paths across an image to cover all locations adequately. In summary, the central problem of lightness constancy is to separate the two components in an image, namely illumination and reflectance. Land invented a simple method in one dimension by taking ratios of luminance on any two adjacent points. Since light intensity of an image is a product of illumination and reflectance, the first step is to take logarithms to convert the product into a sum. The second step is differentiation, which is the equivalent of division in continuous case. The results are sum of the derivatives of the two components. The derivative of illumination should be a smooth continuous curve, whereas the derivative of reflectance should result in sharp pulses. These sharp pulses correspond to intensity discontinuities, or edges.

The next step is to filter out the smooth continuous curve, which is produced by illumination, using some threshold method. Here, Land and McCann suggested that in physiological model, synaptic thresholds take care of this problem. What is left after the thresholding is the derivative of lightness. To obtain the original lightness up to a constant, an integration is needed. To fully recover the original lightness, a normalization is needed to take care of the constant introduced by integration. Since the zero spatial frequency term has been lost in the differentiation, only an approximation of the original lightness can be reconstructed. A simple scheme of normalization is just to assume that the highest value of lightness corresponds to white. However, the algorithm would fail if light sources are present in the field of view. The same is true for fluorescent colors and specular reflections.

2.1.2 Horn's 2-D approach

Horn [11] extended the Land and McCann's one dimensional algorithm to two dimensions. Horn's approach consists of differentiation, thresholding, and integration. In differentiation, Horn introduces a two dimensional center-surround form of differentiation operator, namely the Laplacian. Here, the principle of lateral inhibition in biological system is used in the algorithm. His model starts with

$$p'(x,y) = s'(x,y) \times r'(x,y) \tag{1}$$

where $p'(x,y)$ is the luminance at point (x,y) and s' is the illumination and r' the reflectance. After a log transform on (1), the product becomes a sum.

$$p(x,y) = s(x,y) + r(x,y) \tag{2}$$

where

$$p(x,y) = \log(p'(x,y)), s(x,y) = \log(s'(x,y)), \quad and \quad r(x,y) = \log(r'(x,y)) \tag{3}$$

Next, a Laplacian operator is applied to equation (2) to perform a 2D differentiation.

$$d(x,y) = L(p(x,y)) = L(s(x,y)) + L(r(x,y)) \tag{4}$$

The derivative of $s(x,y)$ is ≈ 0 because the illumination varies smoothly. Therefore, a threshold operator, T, can be applied to (4)

$$T(d(x,y)) = T(L(p(x,y))) \approx L(r(x,y)) \tag{5}$$

One unresolved problem is the method for selecting the threshold. The same is true for Land's one dimensional case. Horn establishes some bounds which are related to the smallest intensity step between regions in an image, noise of the optical system, and the spacing of the sensor cells.

$$e < g, \quad e > (a \times h) <, \quad e > 3(2) - 1/2s \tag{6}$$

where e is threshold, $g = g' \min(1, \frac{h}{2m})$ with g' defined as the smallest step in the log of reflectance in scene, h is the spacing of the sensor cells, m is the radius of the point-spread function of the optical system, s is the root-mean-square noise amplitude, a is the largest slope due to illumination gradient. Among these parameters that determines a threshold value, both g' and a require global comparison in the entire scene. These parameters are rather arbitrary, but technically reasonable. Finally, the reconstruction or integration operator, G, is performed to recover the original lightness.

$$l(x,y) = G(T(L(p(x,y))) = r(x,y) + C \tag{7}$$

where $l(x,y)$ is the computed lightness and G is a convolution with a Green's function. Here, Green's function is chosen to be an inverse laplacian to a partial differential equation of the form $L(p(x,y)) = d(x,y)$ inside a bounded region. Again, assigning the highest numerical value in the integrated output to correspond to white is used as the normalization scheme.

2.1.3 Blake modified Horn approach

Another shortcoming of Horn's method, as pointed out by Blake [1], is that he did not have an exact equivalent 2-D algorithm to that of Land's. Horn made two wrong assumptions. First, his algorithm only imposes that within a patch in Mondrian, only the laplacian of the intensity need to be zero. Since there might be intensity gradient in a patch, a stronger constraint is needed; first derivative of intensity in any patch must be zero, not just the laplacian. The second assumption that the entire Mondrian is within the view of the retina and it has a constant reflectance border is unnecessarily restrictive. Granted that Horn needs this assumption to guarantee an unique solution for reconstructing the reflectance, but Blake proposed a modified algorithm that will also guarantee an unique solution, but without the constant reflectance border. Blake first gets the gradient field of the Mondrian in study.

$$F = \nabla[p(x)] \tag{8}$$

where F is the gradient field and p the luminance. The gradient vector field is then passed through an threshold function which compares the square of the norm of gradient with a threshold value.

$$E = T(F) \tag{9}$$

where

$$T(F) = t((\mid F \mid \times \mid F \mid - [\lambda] \times [\lambda])F) \;\; with \;\; t(z) = 1 \;\; if \;\; z > 0. \tag{10}$$

This threshold scheme serves the same purpose as Horn's, that is to discard the illumination gradient. However, Blake does this after the vector field and Horn does it after the laplacian. Thresholding the gradient field, Blake makes sure that the correct constraint is enforced, so that no gradient is allowed within a patch. The thresholded gradient is then differentiated once more to set up a Poisson equation with well defined boundary conditions:

$$L(x,y) = E(x,y) \tag{11}$$

where L(x,y) is reflectance with a Neumann boundary condition. The solution to the Poisson equation can be obtained by iterative convolution in a similar way to Horn's method.

2.1.4 Summary of Retinex Theory Approach and their Major Problems

In summary, Land, Horn and Blake all take very similar approach which involves a logarithmic transform of intensity, followed by a differentiation plus a threshold, and reconstruction of reflectance by integration. Two major problems arise in this approach when applied to real images. One is that the integration process only produce a reflectance solution up to a constant. How can we recover that constant to get the original reflectance? (So far, no good solution has been proposed.) Another problem is the selection of the threshold value. Since this threshold is global, meaning the same threshold values is applied to all pixels in an image, how should the threshold be chosen such that shallow edges caused by shadows will not be threshold out. Horn proposed a threshold that is bounded above by a function that depends on the smallest step of reflectance, the radius of point-spread-function of the optical system and spacing of sensor cells and bounded below by maximum illumination gradient and root-mean-square noise amplitude. However, both the smallest step of reflectance and maximum illumination gradient are not apriori knowledge and might vary from scene to scene. Horn proposed that a histogram of the differentiated image be taken to eliminate the values around zero, which are due to illumination gradient and noise. But this method is relatively slow and prone to error when a highly textured image is being analyzed. The following sections present methods that have variable setting mechanism for thresholds to preserve shallow edges.

2.2 Intensity Dependent Model

Intensity Dependent Spatial Summation (IDS) model for the photoreceptors [3] rests on the principle of constant volume. Each receptor gives rise to a nonnegative point-spread function which has a conical shape. The center height of this function is directly proportional to the intensity of the input image at that receptor and its volume is constant. Hence, the area of the base (that is, the volume divided by the center height) is inversely proportional to the input intensity. The image at the outputs of Photoreceptors is therefore a sum of these point-spread functions.

The IDS model produces well known phenomenon of Mach Band output when presented a step intensity. Such output has been well described for lateral inhibition circuits. Mach Band effect is known to result in enhancing edges. This enhancement is a property of constant volume modes, be it lateral inhibition or IDS. Lateral inhibition model is a constant volume model whose sombrero-shaped point spread function has a constant volume of Cornsweet. The IDS model of Cornsweet and Yellot is interesting because it demonstrates Mach Band effect in a nonlinear system but without the inhibitory mechanisms. This model however does not apply to cone Photoreceptors which were shown to have inhibitory feedback synapse from horizontal cells [19]. Another feature of the constant volume models is that because of its inherent nonlinearity it does not obey the principle of superposition. Thus input intensity 2I and I both produce the same outputs instead of former output being twice the amplitude of the later one. As a consequence, the effect of an image on the IDS system is not to change its output but rather to redistribute that output in space. In other words, constant volume model has zero sensitivity to spatial frequency zero.

The illumination in natural scenes can vary over as much as 10 log units over the course of a day. Because of shadows, local illumination variation in a scene can span more than 3 log units. Currently, no imaging medium can readily accommodate such wide dynamic range. The usual solutions of using filters or amplifier gain changes to deal with this problem has two fundamental objections. First, sensitivity to local illumination variation, for example those caused by shadows, is compromised because using an iris or a filter reduces the effective luminance of the entire scene by a common factor, which can reduce the signal level in shadowed regions down into the range of the system noise. The other objection is that contrast sensitivity can always be improved by increasing the quantum catch. Thus any gain-control mechanism that simply enforces a fixed quantum catch, as the use of an iris or a filter does, is bound to become increasingly inefficient as the illumination level rises.

The IDS mechanism overcomes these two objections by compressing all input intensities into an output range extending from zero up to around twice the value of the constant point-spread volume. The mechanism made efficient use of every photon. As the image plane illuminance increases, the extra photons serve to decrease the size of the spatial-summation area, improving spatial resolution while maintaining a fixed reliability of contrast detection. And this effect occurs locally within a single image, so that in every region the size of the summation area is matched to the illumination falling upon objects in that portion of the scene. In another word, the IDS can locally adapt to illumination variation to improve spatial resolution and contrast detection.

However, one limitation of Cornsweet model is that when the intensity of the input field is too high, such that the summation area of each receptor is smaller than one receptor size, the whole system is saturated. This is perfectly all right for Cornsweet because his model seeks to explain scotopic vision and not photopic vision. In scotopic vision, operating mainly with rod photoreceptors, spatial summation is needed to improve sensitivity. But in photopic vision, the receptor's operating range must shift with the background illumination to preserve contrast sensitivity over wider and higher range of illumination.

2.3 Grossberg's BCS and FCS

Grossberg, Cohen, and Mingolla invented one neural network architecture for form, color, and brightness perception. Fundamental to this architecture is a principle which they referred to as boundary-feature tradeoff [2,8,7,5,6]. The principle states that percepts are synthesized by the integration of two perceptual processes: boundary completion and featural filling-in. The boundary completion process synthesizes invisible outlines that corresponds to either "real" or "illusory" boundaries generated by the perceptual process. The featural filling-in process generates visible featural qualities by spreading values of brightness and colors. These two parallel processes extract different types of edge (contour) information from the image. This information is then integrated to form the final percept.

There are three systems in the model: the Boundary Contour System (BCS); the Feature Contour System (FCS); and the Object Recognition System (ORS). Two types of interactions are involved in modeling the form, color, and brightness perception: preattentive interactions from BCS to FCS and attentive interactions between ORS and BCS and between ORS and FCS. The processing begins with monocular preprocessed signals (MP) being sent independently to both the BCS and FCS. The BCS pre-attentively generates coherent boundary structures from these MP signals. These structures are passed to FCS and ORS for further processing. The ORS, in turn, rapidly sends attentively learned template signals to the BCS. These template signals can modify the pre-attentively completed boundary structures. The BCS passes these modification along to FCS. The signals from the BCS organize the FCS into perceptual regions wherein filling-in of visible brightnesses and colors can occur. This filling-in process is initiated by signals from the MP stage.

The BCS is triggered by the activation of oriented masks, or elongated receptive fields, at each position of perceptual space. The output signals from these oriented masks are sensitive to the orientation and to the amount of contrast, but not the direction of contrast, at an edge of visual scene. These outputs go through two stages of neighborhood competition processes to generate discontinuities in edge orientations, end points and corners. The global oriented cooperation and boundary completion stage then process the results from the previous stages to form both "real" and "illusory" contours.

The FCS follows a different set of rules for contrast detection than that of BCS. The process is insensitive to contrast orientation in a scenic edge, but it is sensitive to both the direction of contrast and the amount of contrast. The reasons being that feature filling-in involves spatial spreading of featural signals which requires no orientation information; but computation of relative brightness across a scenic boundary requires keeping track of which side of the boundary has a larger reflectance. Also, sensitivity to direction of contrast is needed to determine which side of a color boundary is of what color. It is important to point out that FCS's sensitivity to the amount of contrast removes the illumination gradients in a scene. This effects of "discounting the illuminant" is based on the assumption that difference in illumination are small across an edge and therefore, relative contrast across edges are good estimates of the object reflectances near the edge.

So far, both of the outputs of the BCS and FCS are edges. A further process is necessary to recover the featural properties of the scene. Grossberg, et.al., proposed a featural filling-in mechanism that takes signals from both systems. Grossberg envisions this mechanism to reside in a syncytium of cells. This is a regular array of interconnected nodes thus allowing for fast and simple propagation of electronic signals between neighboring cells.

When the feature contour signal activates a node, or compartment, that node rapidly sends signals activating its neighbors. This spreading mechanism occurs via the "electronic diffusion" of activity. The rate of diffusion across a compartment is controlled by a space constant that depends on the electrical properties of the compartment's interior and membrane. The boundary contour signals create barriers inhibiting the diffusive filling-in process from continuing past the boundary contours extracted by the BCS. The inhibition is implemented by allowing a boundary contour signal to decrease the diffusion space constant of its target compartment's membrane.

The passive propagation of of signals as modeled by the syncytium of cells results in integration process which is faster than Horn's iterative method. This is because Grossberg's BCS breaks up an image into many smaller regions which in effect converts the problem to integration processes of many smaller regions in parallel instead of one large region. However, the problem of normalization is left unsolved by Grossberg also. Although, his ORS provides a higher level process that could contain some absolute reference to fully recover reflectance.

In summary, Grossberg, et. al., suggested an extensive model that involves both low and high level visual processes to simulate brightness, color, and form perceptions. But he left out many details in the architecture. An important detail of implementation is the physical properties of the image sensor. His model only assumes a mathematical model of a sensor that maps input intensity to output response without putting any physical constraint on the sensor. In fact, Land and McCann, Horn, and Blake all ignored these constraints. One crucial physical constraint of image sensor is its dynamic range. No sensor can have unlimited dynamic range. Besides, there is a dynamic range and response sensitivity trade-off. The following architecture uses physiological results from vertebrate photoreceptors and from the neural structure of the retina to derive the physical constraints. These constraints are used as a guide to design a system that overcomes the dynamic range and response sensitivity tradeoff and uses only low level visual process to produce a similar Mach Band outputs of previous two models.

2.4 UCLA-MPL Approach

2.4.1 Introduction

Using neurophysiological findings as a guiding principles, we define a class of parallel architectures that is characterized by large number of connections. The connectivity is flexible and can be sometimes irregular. This effort is contrasted to many "connectionist" architectures, unrelated to neurophysiology that have been made popular in the last three These connectionist architectures designed by various engineers and scientists infected by this "NEURONIA", usually have complete connectivity from every node to every node. They are limited to three layers and most often are intended to "learn" something.

The visual pathway is a layered and hierarchical structure with divergence at the earlier stages to convergence at latter stages. This structure is evident in the stack of five layers of cells in the retina: the photoreceptors responsible for image transduction and the bipolar, horizontal, amacrine, and ganglion cells.

Convergence is exhibited by the connections of many peripheral photoreceptors to a single bipolar cell. Another feature of natural vision system is lateral interaction. One example of lateral inhibition can be found in the lateral connections of amacrine cells to ganglion cells to provide contrast enhancement.

A distinct feature of human vision system is that a typical neuron has perhaps 2,000 to 16,000 inputs. The fan-in and fan-out factors are huge. The concept of visual receptive field is a convenient way of describing the pattern of connections of this complex network. A receptive field of a cell is a pattern of photoreceptors in the retina which results in a specific firing behavior of the cell. There is strong physiological evidence that different types of receptive fields function as features detectors [18].

Divergence is evident in visual cortex. Here, receptive fields of different sizes that detect features in different orientations are mapped into hyper columns. Each hyper columns has columns of feature detectors of different orientations covering the same receptive fields. Each column of the hyper columns has feature detectors of same orientation but different sizes that vary along the depth of the column [12,16].

Finally, the integral part of the visual system are the neurons. One can model neuron as a weighted sum and thresholded processor. This means the processor needs multiplication, summation and comparator functions.

Skrzypek and Shulman [21] proposed a computing architecture that tackles the lightness constancy problem plus the automatic setting of thresholds. This algorithm is inspired by the neurophysiological findings in the vertebrate retina. A module based on hexagonally organized, partially overlapping lateral inhibition operators of different sizes

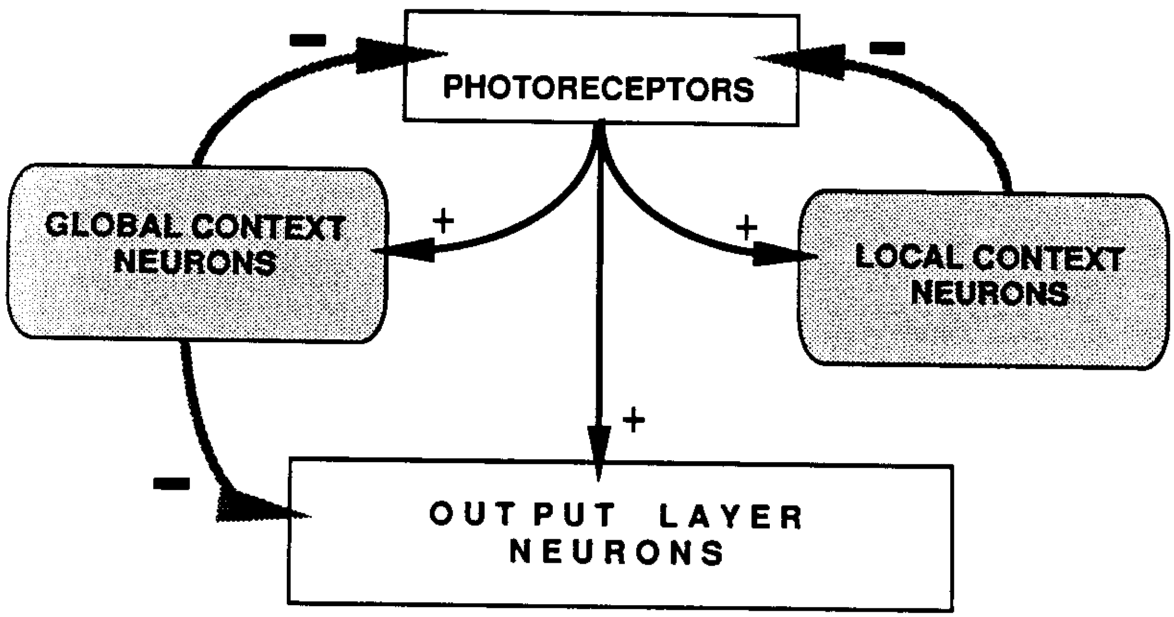


Figure 1: A schematic of a neural lightness algorithm that was implemented in SFINX and tested on realistic, gray level images.

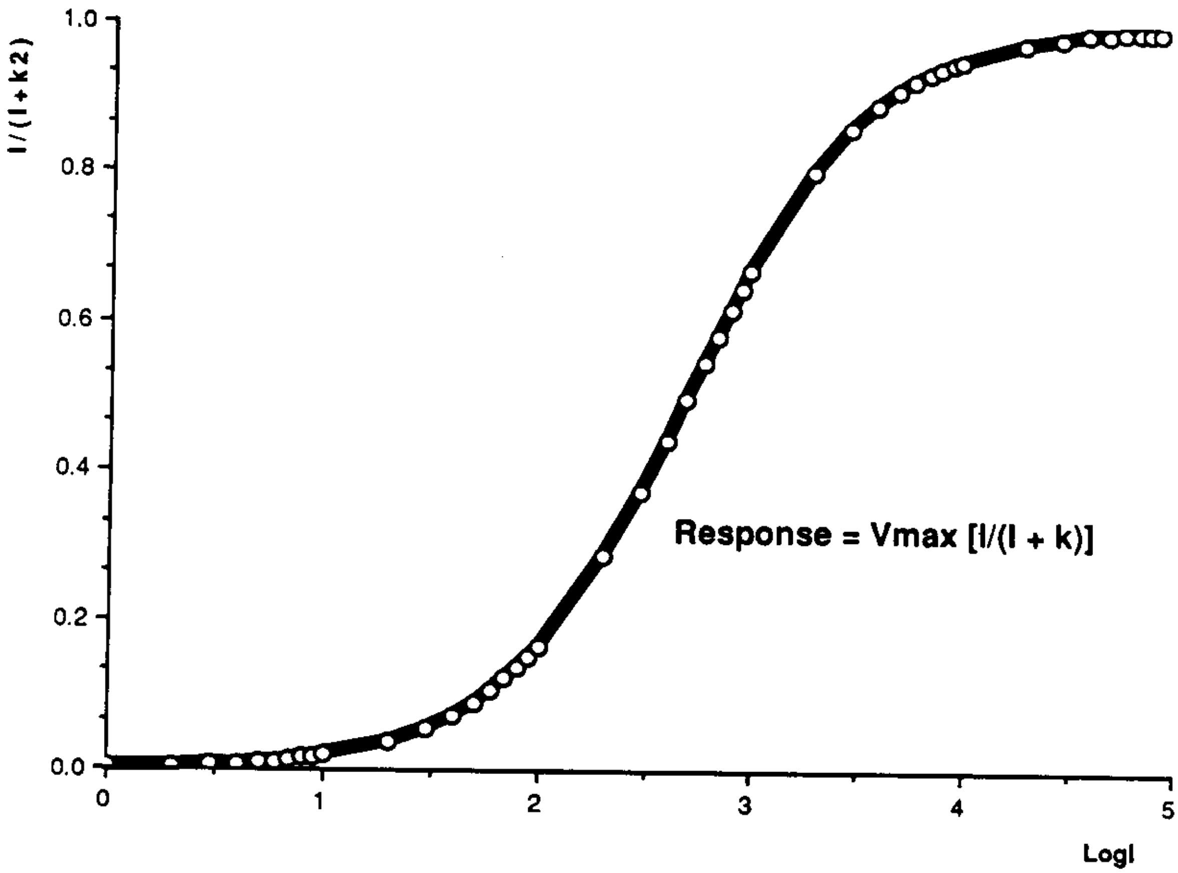


Figure 2: S-shaped transfer curve for a single "neuron"

consists of four functionally different layers of cells. This spatial preprocessing module for low level vision will maintain high sensitivity over the whole domain intensities without interfering with transmission of visual information embedded in spatial discontinuities of intensity.

In terms of signal hierarchy, counting number of neurons from the input there are three layers of processing (Fig. 1). The first layer is composed of photoreceptors with logarithmic transfer function spanning finite domain of intensity. The transfer function has decreasing sensitivity at the lower end and saturation at the higher end. This S-shaped I-R curve spans at most three logarithmic units of intensities. To cover the whole domain of 10 log units, a mechanism is proposed that automatically shifts the response curves of the photoreceptors to be in register with the prevailing ambient light level (Fig. 2). Thus the transfer curve is a steep (high-gain) intensity-response characteristics with the automatic setting of threshold mechanism that is optimally selected by the positioning of the photoreceptors operating point.

To achieve this automatic setting of threshold, two types of context operators in the second layer are introduced. One type are global average operators. The other type are local average operators. A combination of the global and local responses is feedback to each individual photoreceptor to reset its own operating point.

Finally, the output layer uses two levels of resolution to accurately reconstruct the original lightness signals and increase the chance of detecting actual spatial discontinuities in intensity. Here, we use the center-surround form of operators with two types of polarities. One is center minus surround and the other surround minus center. The result computed by the output layer is then roughly the average second difference of the input intensities.

In summary, with the assumptions that the photoreceptors converging onto a given surround, output layer operators are linearly combined and that inhibition is a simple linear operation our computational structure which performs the lightness constancy function resembles something as simple as a difference of Gaussians [20].

2.4.2 Architecture Description

The structure was designed and tested using UCLA SFINX simulator [17]. Each layer of Skrzypek's architecture has a corresponding buffer array. There are five buffer arrays. The architecture is illustrated in Fig.3. The buffer arrays are represented by the parallelograms. The links represent data flows. The texts near the links are associated node_functions that operate on the data links. Keep in mind that each node_function operates on each element of a buffer array.

The input buffer at the top of figure 3 contains the actual intensity from a scene to the image sensors. This input buffer is fed through the buffer function, bf_sensor, which has another input from the feedback buffer. The bf_sensor function performs the log transform and has a sigmoidal response curve. Its function is to simulate the cone receptor's behavior, which are endowed with the automatic shifting mechanism. The output of the bf_sensor function is stored in the sensor buffer. The bf_local function performs a convolution with a local mask. Typically, a local mask is of sizes 3x3 through 7x7. The output of the convolution is then stored in the local buffer. This local convolution is equivalent to [20] local average operator. Similarly, the bf_global is a convolution with a global mask. Again, this function performs [20] the computaion of global average operator. Global mask is about four-ninth of the image area. The global buffer contains the outputs of the bf_global function. The bf_lincom function takes inputs from both the local and global buffers to compute the new biases for each of the sensors. These biases then adjust the operating curves of each sensor's response function, bf_sensor, to produce new sensor output values.

The sensor response curve is modeled by the following equation.

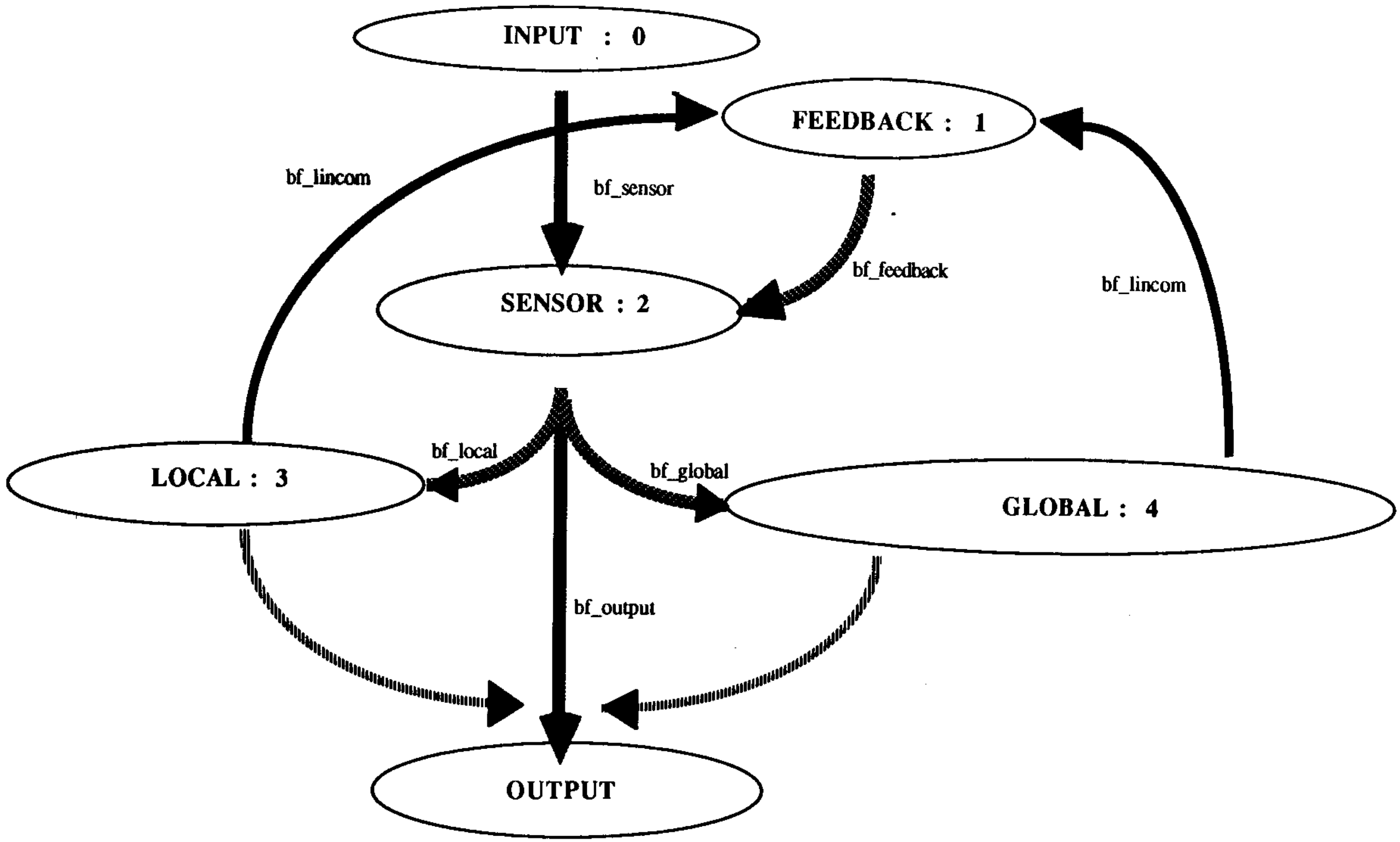


Fig.3. Architecture for Lightness Constancy

$$R = R_{max}(\frac{I}{I+k}) \tag{12}$$

where R is the response, I is the input intensity, k is the input intensity that yields a half-maximum response and R_{max} is the saturation response. In this equation, k is the adjustable bias that is capable of shifting the response up and down along the intensity axis. During each feedback cycle, a new k is computed from the current k and a linear combination of the global and local averages using the next equation.

$$k' = k(lc/(R_{max} - lc) \tag{13}$$

where k' is the new k and lc is computed as follow.

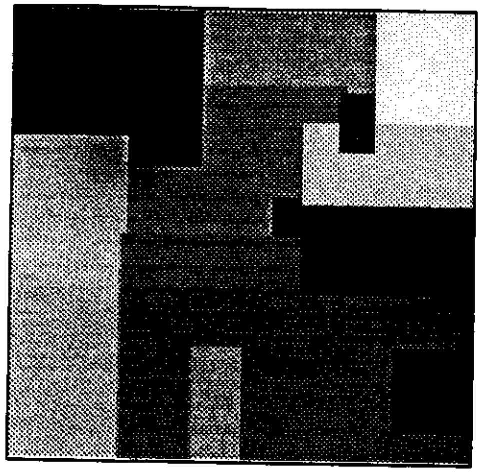
$$lc = a1 \times b + a2 \times c \tag{14}$$

where $a1$ and $a2$ are constants and b is the global average and c the local average. After several cycles, the output of the sensor array is stablized when changes in successive $k's$ is small. [20] suggested a way to use the contents of global averages, local averages and the sensor outputs to compute an averaged second difference of the input intensities which emphasizes edges.

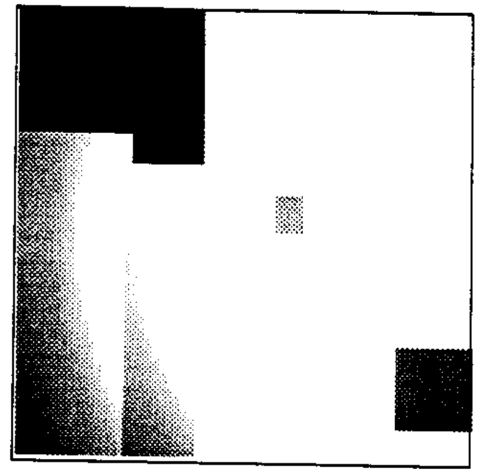
2.4.3 Simulations in SFINX

Additional buffer arrays are included to simulate the effect of different illumination conditions. This is accomplished by the bf_gauss function. This function takes available illumination functions and convolves it with a gaussian point-spread function. Thus, a parameter that determines the maximum intensity of the illumination on the image plane and a sigma parameter that specifies the rate of decay of the gaussian curve are incorporated in the function. In effect, this function multiplies the input image with a specified point-spread illumination to produce an output image. If the input image is taken under evenly illuminated condition, the image is linearly proportional to the reflectance of the scene. The bf_gauss function, then, produce an output that includes an controlled illumination component to the input image.

The ability to control illumination conditions is important to simulate the enormous dynamic range of natural illumination conditions that can be accommodated by our architecture.



(a) Original Mondrian



(b) Simulated Non-uniform Illumination

Figure 4: Gray level, mondrian like input image (a) was illuminated by a strong spot light source (b) to simulate various non-uniform illumination conditions.

Artificial Mondrian test images of 64 pixels by 60 pixels were used to evaluate our architecture (Fig. 4). The evaluation were done in two parts. The first part is to experiment with the different linear combination of local and global averages. The second part is to evaluate the constancy of outputs across various illumination conditions.

In order to isolate the effect of averaging mask size to the output image, feedback of 3x3, 7x7, 15x15, and 42x42 convolution averages were applied to the sensors separately. Mach bands effects were clearly observed at the edges of grey patches. The rate of decay of the response is determined by mask size. Smaller mask size produces faster decay and larger mask size gives slower decay. In all cases, large uniform reflectance areas have middle grey response. This is expected because the feedback mechanism is functionally similar to a difference of gaussian, which is a form of differentiation. And differentiation produces no response when no changes are in the input.

After analyzing the independent effects of mask size, we next combined linearly two different sizes of averaging masks to record the difference. No noticeable difference was seen that would be different from the outcome of the 15x15 mask size output.

To conduct the second part of the experiment, two images with different illumination conditions were used. The illumination conditions are varied in two aspects. The first aspect explores the variation in the position of the light source. As shown in figure 4, the left one has the light source located at the center of the image and the right one has the source located at the middle right of the image.

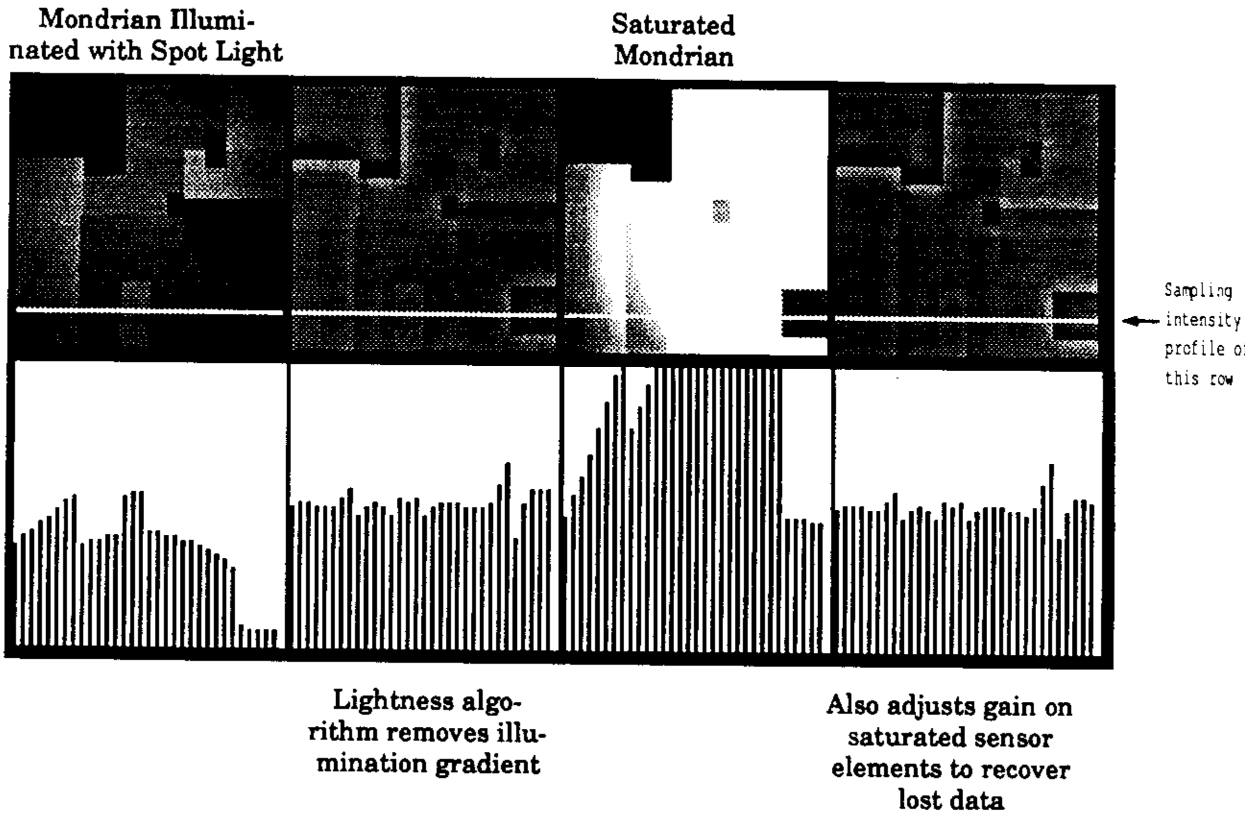


Figure 5: The results of the Lightness algorithm show that it can compensate for arbitrary illumination gradients, thereby preserving useful topological information.

The second aspect takes the illumination intensity of the light source as a variable. The right image's illuminant is 1000 times that of the left image. In fact, the right image represents the output of sensors with fixed biasing. Almost the entire right half of the image is saturated due to the wide dynamic range of the illuminant. Despite this wide dynamic range, the feedback mechanism of our architecture are able to adjust their own biasing to adapt to the illumination changes and produce constant outputs as demonstrated in Figure 5.

A comparison of the two input images and their respective output profiles is also intensity. shown. The intensity profiles are taken at 50th row of each images. Again there is no significant difference in the output profiles despite drastic differences in the input.

2.4.4 Discussions

From our simulation results, we have demonstrated our architecture's capability to produce constant outputs with illumination variations in light source location and intensity. As to the temporal aspect of this architecture, all outputs were produced after 3 to 4 cycles of feedback with the initial biasing of the sensors set at the middle grey level. We suspect random biasing for the initial condition will delay the convergence of the outputs, but not significantly.

However, several critical questions arised from our simulations. One is how to combine the multiple resolutions of averaging for the feedback. Another related problem is how many of these different sizes are needed and at what distance apart. [20] also pointed out these problems and suggested ways to combine the multi-resolution feedbacks.

Although our outputs are constant, but some sort of global integration process is needed to recover the desired lightness. An open question will be how to implement a bf_output function in our architecture that could utilize the outputs of the averaging operators and the sensor outputs to reconstruct a reflectance map of the artificial mondrian. Will additional information be needed to solve this integration process?

Horn and Blake used iterative integration method to reconstruct lightness up to a constant [11,1]. Grossberg proposed a much more elaborated method using his boundary contour system and feature contour system interacting at syncytium to reconstruct lightness. Horn and Blake's methods took thousands of iterations which is too slow and Grossberg's method is faster but suggest that reconstruction of lightness is much higher visual processing that is done not at the retina, but somewhere higher up in the visual cortex.

3 Summary

The computation required for many early visual functions could be reduced by embedding these functions in the complex network of neurons. Computation at this level should take the advantege of massive parallelism built into image data. The neural networks are characterized by three-dimensional highly connected structures that are well suited to data parallelism. This could satisfy the requirement of recognizing objects in real time by computing object's attributes from local information.

The conventional von Neumann architectures are insufficient for the amount of I/O bandwidth required. The real-time requirement and technical feasibility proliferated the growth of different parallel architectures. Currently, electronic computing dominates the implementation of the machine vision architectures. However, the two-dimensional layout of VLSI chips puts a crucial limitation on the number of interconnections among processors. This limitation prevented the more interesting connectionist architectures from being developed in hardware. Optical computing overcomes the connectivity limit. Since optical pathways do not interfere with one another even when they are crossed, optical interconnections provide wider freedom in signal communications. Unfortunately, optical computing is not practically feasible at this time.

References

[1] A. Blake. Boundary Conditions for Lightness Computation in Mondrian World. *Computer Vision, Graphics, and Image Processing*, 32:314–327, 1985.

[2] M.A. Cohen and S. Grossberg. Neural dynamics of brightness perception: features, boundaries, diffusion, and resonance. *Perception and Psychophysics*, 36(5):428–456, 1984.

[3] T.N. Cornsweet and J.I. Jr. Yellot. Intensity-dependent spatial summation. *Journal of Optical Society of America A*, 2(10):1769–1786, October 1985.

[4] J.J. Gibson. Perception as a function of stimulation. In S. Koch, editor, *Psychology: A Study of a Science*, McGraw-Hill, New York, 1959.

[5] S. Grossberg. Cortical dynamics of three-dimensional form, color, and brightness perception: i. monocular theory. *Perception and Psychophysics*, 41(2):87–116, 1987.

[6] S. Grossberg. Cortical dynamics of three-dimensional form, color, and brightness perception: ii. binocular theory. *Perception and Psychophysics*, 41(2):117–158, 1987.

[7] S. Grossberg and E. Mingolla. Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 92(2):173–211, 1985.

[8] S. Grossberg and E. Mingolla. Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Perception and Psychophysics*, 38(2):141–171, 1985.

[9] H.V. Helmholtz. *Physiological Optics.* Volume 2, Optical Society of America, New York, 3rd edition, 1924. (English Translation, J.P.C. Southall).

[10] E. Hering. *Outlines of a Theory of the Light Sense.* Harvard Press, Cambridge, Mass., 1964. (English Translation, L.M. Hurvich and D. Jameson).

[11] K.P.H. Horn. Determining Lightness from an Image. *Computer Vision, Graphics and Image Processing*, 3:277–299, 1974.

[12] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's cortex. *Journal of Physiology (London)*, 160:106–154, 1962.

[13] W. Kohler. *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology.* Liveright, New York, 1947.

[14] E.H. Land and J.J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, January 1971.

[15] R.A. Normann and F.S. Werblin. Control of retinal sensitivity. i. light and dark adaptation of vertebrate rods and cones. *Journal of General Physiology*, 63:37–61, 1974.

[16] R. Ornstein and R.F. Thompson. *The Amazing Brain.* Houghton Mifflin, Boston, 1984.

[17] E. Paik, D. Gungner, and J. Skrzypek. UCLA SFINX - A neural network simulation environment. In *Proceedings of the IEEE First Annual International Conference on Neural Networks*, 1987. San Diego, California, June 21-24.

[18] R.W. Rodieck. Receptive fields in the cat retina: a new type. *Science*, 157(3784):90–92, 1967.

[19] J. Skrzypek. In preparation. 1989.

[20] J. Skrzypek. A Unified Computational Architecture for Preprocessing Visual Information in Space and Time. In *Proceedings of SPIE Conference on Computer Vision for Robotics*, pages 258–263, 1985.

[21] J. Skrzypek and D. Shulman. *Computing Architecture for Spatial Preprocessing of Visual Information.* Technical Report 001, Northeastern University Vision and Robotics Research Lab., May 1985.

[22] J. Skrzypek and F.S. Werblin. Lateral Interactions in Absence of Feedback to Cones. *Journal of Neurophysiology*, 49(3), 1983.

[23] H. Wallach. Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology*, 38:310–324, 1948.

[24] F.S. Werblin. The Control of Sensitivity in the Retina. *Scientific American*, 228(1):71–79, January 1973.