

Early vision: Images, context and memory

Dov Sagi*

Department of Neurobiology, Brain Research

The Weizmann Institute of Science

Rehovot 76100, Israel

Visual perception involves image transformations, producing internal images with some useful features enhanced, possibly linked. Enhancement is a result of local transformations being context dependent, affected either by remote image regions or by visual memory. Psychophysical experiments show that early vision generates an image that, although dominated by the retinal output, contains a context dependent component. Context affects response of oriented filters through lateral interactions. Efficacy of lateral connections may change with experience (perceptual learning) and can be modulated by higher levels of processing (visual attention and imagery). Psychophysical results suggest that long range interactions are achieved by activation spread through multiple connections. Some properties of this model of early vision are examined and applied to a variety of problems, as lateral masking, texture discrimination and visual grouping.

1. Introduction

Our visual system transforms the apparently meaningless reality inflicted on our retinas into meaningful and behaviorally relevant objects. As the retinas are specialized in detecting picture points (pixels), we are faced with the problem of attributing these point data to objects. The embedment of point measurements within the proper context of a perceivable object, requires integration between remote image parts as well as with memory.

This problem is best exemplified by the hidden figure puzzle [1] shown in Figure 1. The figure seems to be composed of black and white patches devoid of any semantic significance, though a face is present. As this puzzle is difficult to solve by trial and error, it is suggested to look at Figure 5 briefly and then to return to the hidden figure, if only to be faced with the deep look of a face decorated with Christ styled hair and beard. Perception is changed dramatically, with some 'illusory' contours added (as at the top of the right shoulder), yet, these contours are supported by some minimal physical cues. It seems that the solution reached by our brains is based on integration

*Supported by the Israel Science Foundation administered by The Israel Academy of Sciences and Humanities - The Charles H. Revson Foundation.

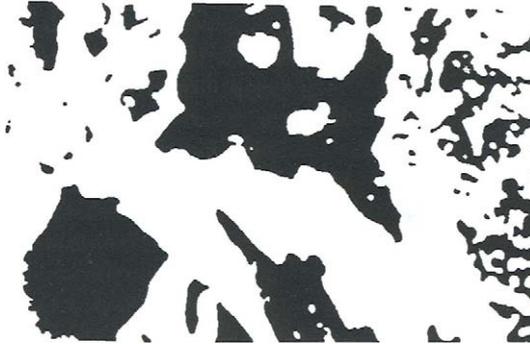


Figure 1.

A hidden figure: the puzzle. From Porter P. [1], copyright by the Board of Trustees of the University of Illinois.

of the physical context with some memory cues (previous experience). At what level of processing does this integration take place? Recent findings [2–6] suggest that both spatial and memory dependent context affect visual perception at a very early stage of processing. It is assumed here that a stage of processing is characterized by integration of information from different sources (lower, as well as higher stages) followed by a decision (threshold). Accordingly, the first stage of visual processing is described as spatial (linear) integration of the retinal input followed by a threshold. This stage, encapsulated within the preattentive bottom-up frame, is a principal component of theories of early vision [7–9], though formulation may differ between theories. Here the emphasis is being shifted to the context dependency of early vision, suggesting that already the first stage of processing is affected by ‘late’ processes and is involved in the creation of a unified perception (*Gestalt*). It is suggested that first stage filters are arranged topographically with excitatory connections between neighbors. Activation may spread in this network, and remote filters may interact, depending on efficacy of excitatory connections. Efficacy is determined by experience and is modulated by visual attention.

2. The primary visual network

2.1. Spatial filters

One of the basic findings in visual perception is that there exist many parallel pathways within the visual system, each of them is specialized to carry information about a different stimulus aspect. These findings were demonstrated in physiology with the introduction of the ‘receptive field’ concept [10,11] and in visual psychophysics with the introduction of parallel processing of color by Young [12] and Helmholtz [13]. Later work

by Hubel and Wiesel [14] and Campbell and Robson[15] opened a period of extensive research in determining the properties of visual channels analyzing form. Although the term ‘channel’ is not well defined, there is a general agreement on the existence of some mechanisms operating in parallel across the visual field and responding significantly to some stimuli but not others (selectivity). Some theories assume that these channel can be described as linear filters [15–18] while others elaborate nonlinear feature detectors [7–9,19].

Here we assume that the basic mechanisms are linear filters, selective for orientation, location and spatial frequency. It is possible to describe these filters as some function of retinal coordinates, orientation and spatial frequency, $F(x, y \mid \omega, \theta, x_c, y_c)$, with parameters representing filter location (x_c, y_c) , its preferred orientation (θ) , and its preferred spatial frequency (ω) . Popular functions used for spatial filters are Difference Of Gaussians (DOG) [18] and Gabor functions [20]. Both filter types were used to model receptive fields of cortical simple cells [8,20].

Filter response is obtained by convolution with an input image $L(x,y)$, followed by a nonlinear transducer function ($\text{trf}[\]$):

$$R_{f1}(x_c, y_c \mid \omega, \theta) = \text{trf}\left[\sum_{x,y} F(x, y \mid \omega, \theta, x_c, y_c) \cdot L(x, y)\right] \quad (1)$$

This formulation is supported by psychophysical experiments indicating linear summation of subthreshold grating stimuli with parameters constrained to within a ‘filter bandwidth’ [16]. Contrast threshold for a sine wave grating is not affected by the presence of another grating that differs significantly by orientation or spatial frequency. Filters bandwidth can be estimated to be about 15° (half-width at half-height) of orientation [21,22], and between one to two octaves of spatial frequency [23–26]. Filter receptive field size varies according to its peak spatial frequency sensitivity, being about twice its optimal wavelength [27]. Some reasonable transducer functions are sigmoid functions [18] and threshold functions [28].

2.2. Excitatory interactions between filters

The filters described above integrate retinal input only. Next we consider excitatory inputs from neighboring units. Physiological studies indicate the existence of long range interactions in the primary visual cortex [29]. Psychophysical evidence for excitatory lateral interactions come from experiments demonstrating that target detection is enhanced by the presence of high contrast masks within some neighborhood [3]. In these experiments, targets and masks were 2D oriented Gabor signals (i.e cosine gratings with amplitude modulated by a Gaussian envelope with $\sigma = \lambda$, λ being the gratings wavelength; a parameter setting that serves as an efficient probe for spatial filters). In the psychophysical experiments, observers were asked to detect a low contrast Gabor signal (target) in the presence of two flanking high contrast but otherwise identical signals (masks). Figure 2 depicts observers thresholds as a function of target to mask distance, showing a threshold elevation (suppression) at small distances and threshold reduction (enhancement) with longer distances. The enhancement range was found to increase with

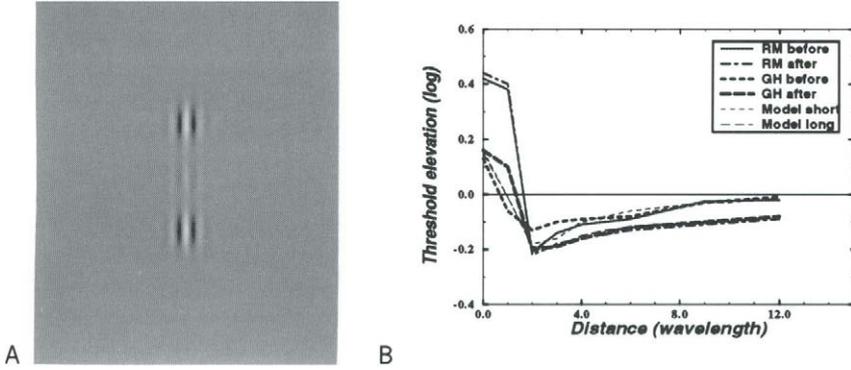


Figure 2.

(a) Stimulus used for exploring lateral interactions, with two high contrast masks (Gabor signals) flanking a low contrast target. (b) Data from the lateral masking experiments, before and after practice (redrawn from [4]). Observers detected a Gabor target flanked by two high contrast collinear Gabor signals at different distances. Contrast detection thresholds relative to absolute threshold (no mask) are plotted as a function of target to mask distance, using Gabor wavelength units. Data is shown for two observers (RM and GH) and a model (see text). Observers show an increased range of enhancement after extensive practice. The model emulates practice by increasing synaptic efficacy and thus increasing the range of signal propagation in a network with excitatory lateral connections. Note that suppression is not affected by practice in these experiments and simulations.

practice [4]. We take the suppression region as an indication for lateral inhibition and the enhancement region to support excitatory interactions. This effect is obtained only if target and masks have the same orientation and spatial-frequency, indicating connectivity between filters with similarly shaped receptive fields. Moreover, target and masks have to be collinear in order for this effect to be significant [30], implying a connectivity pattern with higher density in the direction defined by filter preferred orientation. Recently, Polat and Norcia [5] have shown that Visual Evoked Potentials (VEPs) elicited by a Gabor signal are enhanced by the presence of collinear Gabor signals and that the enhancement is reduced at high target contrasts ($> 20\%$).

Following Polat and Sagi [4] we assume here that these excitatory connections are limited to between close neighbors, but signal transmission through multiple connections is possible. Due to transmission constrains (synaptic) signal propagation through lateral connections involves attenuation (by a factor $0 \leq \kappa < 1$) and requires additional time steps (with a temporal attenuation by a factor $0 < \nu < 1$), resulting the equation:

$$\frac{\partial}{\partial t} R_{\theta}(r_{\theta}, t) = -\nu R_{\theta}(r_{\theta}, t) + \kappa [R_{\theta}(r_{\theta} - \delta_{r_{\theta}}, t) + R_{\theta}(r_{\theta} + \delta_{r_{\theta}}, t)] + \frac{\partial}{\partial t} \Phi(r_{\theta}, t) \quad (2)$$

with r_{θ} as the axis of propagation in direction θ . R_{θ} is the response of a spatial filter with an orientation θ , and some spatial frequency parameter ω , and $\Phi(r_{\theta}, t)$ is the sensory input to this filter. (Only filters with the same spatial frequency parameter are assumed to be laterally connected.)

And if sampling density is high enough:

$$\frac{\partial}{\partial t}R = (2\kappa - \nu)R + \kappa\delta_{r\theta}^2 \frac{\partial^2}{\partial r^2}R + \frac{\partial}{\partial t}\Phi. \quad (3)$$

It is further assumed that excitatory weights are modulated by visual attention (see Section 4). This assumption can allow for a selective increase (or decrease) in connectivity at selected image regions within a time frame defined by network integration time. On a slower time scale, excitatory weights are subjected to perceptual learning [4]. Network response level is also controlled by a normalization process, that is by divisive inhibition.

2.3. Inhibitory interactions

Response normalization is supported by data from masking experiments, where contrast thresholds increase with increasing background contrast. Psychophysical [28,31] and physiological [32] models of filter contrast response assume divisive inhibition, where filter response is divided by a ‘local energy’ measure reflecting the total activation at some small neighborhood around the normalized filter. These inhibitory filters seem to have an isotropic spatial weighting function[30] and probably operate on a slower time scale [33]. Thus, the normalization factor is considered here to be filter response averaged across time. The orientation selectivity of the detection suppression (see Figure 2) observed in lateral masking experiments [30] indicates short range (of about twice the filter size) orientation selective lateral inhibition. Thus the response of a given filter is inhibited by activity of filters with different orientations at the same location, and by spatially adjacent filters with the same orientation. Such a connectivity pattern may serve, in addition to local response normalization, detection of texture boundaries by reducing activity at regions with uniform orientation.

3. Application to some perceptual phenomena

3.1. Lateral masking: a probe on spatial interactions

The range of spatial interactions (i.e the space constant) can be estimated by measuring detection thresholds for a target with high contrast signals (masks) positioned at some distance away from it [3,30]. Data from such an experiment is depicted in Figure 2. These experiments show detection suppression when masks are presented in close neighborhood to the target (less than twice the target wavelength, as predicted by the divisive inhibition). At larger distances an enhancement is observed, reflecting an excitatory input from a distance up to ten times signal wavelength, much beyond the first stage receptive field size. In order to derive model predictions for this case, we assume a narrow input at the origin with a space/time spread of one unit, $\Phi(r_\theta, t) = e^{-(t+r_\theta)}$ (in Equation 3). Taking a non-oscillating solution of Equation 3 one finds (for some v, ξ with $v = \nu - \kappa(2 + \delta_{r\theta}^2 \xi^2) > 0$),

$$R = R_f[e^{-(vt+\xi r)} + C e^{-(t+r_\theta)}]. \quad (4)$$

Model predictions for this case, depicted in Figure 2, were calculated assuming temporal integration and using a Gaussian weighting function for the inhibitory filter with

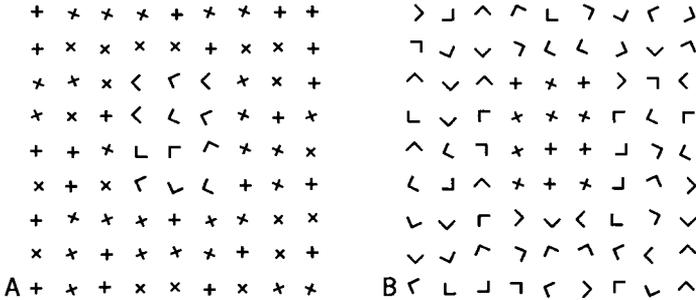


Figure 3.

Effortless texture segmentation in (a) is somewhat easier than in (b), though boundaries are the same. Orientation selective spatial filters predict larger response variability for textures composed of L's, as L's are more elongated than X's (when convolved with larger receptive fields). As a consequence, filter based models generate more boundaries in (b), making foreground localization difficult.

a standard deviation equal to the first stage filter wavelength. Excitation was calculated as a spatial convolution of filter responses with a sum of exponential weighting function, as derived in Equation 4. The best data fitting parameters for Equation 4 yield an input dependent weighting function which is 16 times stronger than the weighting function accounting for lateral range interactions (i.e $C = 16$). The same ratio was also used by Zenger and Sagi [28]. The space constant of the long range excitation depends on experience. The experimental data depicted in Figure 2 show an increase range of interactions with practice (perceptual learning), implying an increase of the corresponding space constant from 6 to 12 filter wavelengths ($\xi = 0.16$ to $\xi = 0.08$).

3.2. Texture segmentation

In texture segmentation tasks one has to detect or locate a texture region (foreground) embedded within another texture (background). A typical example is shown in Figure 3. As texture segmentation is sometimes 'effortless' [7] or attentiveless [47], it is believed to be carried out, and thus to be limited, by early visual mechanisms. Human performance on texture segmentation tasks can be accounted for by a two stage filtering model [41–44] with a similar design to the one described above. The key component in texture segmentation is the first stage filter, being local, orientation selective and spatial-frequency selective. Detection of texture boundaries can be modeled as detection of local activity differences (edge detection) within a filtered image. As a typical texture creates local activity differences across all image regions (e.g. the textures depicted in Figure 3 generate highly variable activity patterns when convolved with oriented filters, due to the random local orientation), the problem is one of detecting an activity edge in a noisy image [44]. Here, spurious activity variations are smoothed by excitatory lateral interactions,

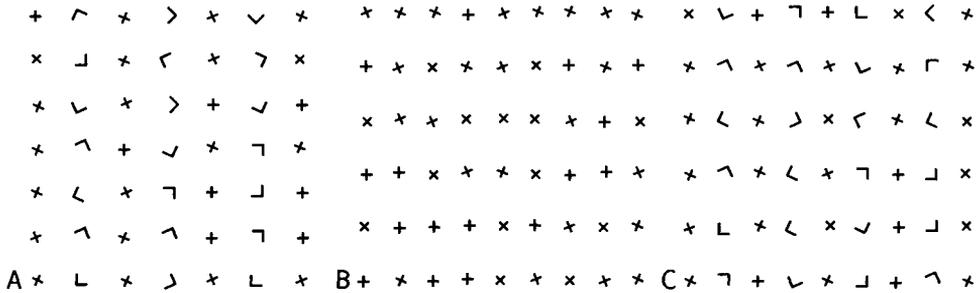


Figure 4.

Visual grouping by similarity (a), proximity (b) and both (c).

and activity at regions of uniform response is reduced by (divisive) inhibition. A similar model was shown to successfully predict human performance on texture segmentation tasks [44].

To account for preattentive texture segmentation, where all image regions are processed in parallel and texture boundaries are to be detected, we assume excitatory weights to be low and constant across the processed image, allowing for good localization performance. However, performance is experience dependent. Observers practicing texture segmentation improve with practice [2]. Improvement occurs through a period of a few days and lasts for years [45]. Learning is specific for texture orientation, location and eye (practicing with one eye does not produce improvement when testing with the other eye), implying plasticity at an early stage of visual processing. These results imply that the network weights used here should be modifiable with experience. It is possible that increase in efficacy of excitatory weights produces a more efficient smoothing of spurious local variations in filters activity. It is also possible that an increase of inhibitory gain further reduces activity at regions with low response variability, thus enhancing texture boundaries.

3.3. Perceptual grouping

Visual stimuli, as Figure 4, give rise to spontaneous perception of some global organization. Thus it seems that our visual system implements specific rules for the creation of structure in seemingly ambiguous situations. While texture segmentation is marking object boundaries, grouping is a process involved in linking image regions into unified objects. The Gestalt laws of perceptual organization [48] assume that image parts group together when being similar in shape or when being in close proximity, or when creating a

'good form'. The Gestalt laws, though applicable in many situations, lack a quantitative formulation. It is not clear how to define shape and similarity or what entities to use for proximity measurements. Furthermore, often images contain multiple organization cues (e.g. proximity and similarity in Figure 4c) thus calling for some quantitative method to combine the different rules into a unified framework. The primary visual network described above provides a natural framework for handling internal image relations, as required in perceptual organization. Spatial filters can be used to construct a similarity metric (as in texture segmentation), and anisotropic excitatory lateral connections can be used to implement spatial relations.

A recent theory of grouping accounts for human performance on proximity and similarity based grouping tasks by assuming an autocorrelation type of process [49]. Though the autocorrelation computations can be applied to filter responses, a simpler model is possible. In this model [49], the input image is represented as an intensity map, $I(x, y)$, thus skipping the filtering stage. The basic operation performed is an estimate of total weighted directional correlation at each image point:

$$R_{\theta}(x, y) = \sum_{\xi} I(x, y) \cdot I(x + \xi \cos \theta, y + \xi \sin \theta) \cdot e^{(-\xi/\xi_0)} \quad (5)$$

Here, $R_{\theta}(x, y)$ represents the total weighted correlation of the image point $I(x, y)$ with all image points along a line with orientation θ passing through the point (x, y) . Long range correlations are given a smaller weight (exponentially decaying), thus introducing proximity effects. This equation was successfully applied to psychophysical experiments using stimuli (Figure 4) where global orientation was defined either by elements proximity and/or shape similarity and/or luminance similarity. As the psychophysical task in the experiments was a two alternative forced choice between vertical and horizontal global organizations, a model decision parameters was constructed (Λ),

$$\Lambda = \frac{\sum_{x,y} R_{\theta=vertical}(x, y)}{\sum_{x,y} R_{\theta=horizontal}(x, y)} \quad (6)$$

so that $\Lambda > 1$ indicates vertical organization, $\Lambda < 1$ indicates horizontal organization, and an unbiased guess is forced with $\Lambda = 1$.

Performance on the grouping tasks was found to be time dependent [49]. In the experiments, stimuli (like in Figure 4) were briefly presented (20 msec) followed by noisy unorganized pattern to mask the grouping perception. For smaller stimulus to mask temporal separations ($t < 100$ msec), the perceived stimulus organization was dominated by proximity relations, while additional processing time allowed for similarity based grouping to take over. Equation 4 can account for this behavior if the space constant ξ_0 (the model free parameter) is allowed to increase with time. Thus, the model suggests an increasing integration range with time.

Our network can account for human behavior by assuming that the weighted spatial integration is performed by the excitatory lateral connections. The exponential distance

weight applied to the correlation equation is predicted by signal propagation through lateral connections (section 2.2). Dynamic control of connectivity can result in an increase of excitatory spread within network integration time (< 200 msc), predicting increasing space constants with time and thus accounting for the psychophysical findings [49]. This fast dynamic of excitatory transmission is assumed here to be controlled by visual attention. It is also possible that fast modulations of synaptic efficacies are generated by response correlations [50,51](in filters or some other responding units), however, such a mechanism would not allow transmission of activity through unstimulated image regions. Perceptual organization involves, not only enhancement of activity correlations, but also filling in gaps where information is missing (as in Figure 1). Psychophysical experiments indicate that perceptual grouping depends on the availability of visual attention [39,40], supporting the assumption of attentional control over excitatory efficacy.

4. Visual attention

Visual attention is assumed to link image features into objects [7,9], to integrate filter responses [34,35], or to select a region in the visual field for enhanced processing [36]. As image features (e.g. orientation, size, color) are not represented uniquely within the filter representation used here (assuming filters perform weighted linear broadband integration within a multidimensional space), attention is assumed here to derive image features from filter representations by integrating filter responses. On this account, attention drives visual process that reconfigure filters by modulating their (excitatory) interactions to create ensembles that are better tuned to specific image features. Though these processes can be implemented in a feed-forward fashion by creating higher level filters using dynamic receptive fields [37], it is attractive to assume a single layer feedback network capable of handling global process (as needed for shape analysis).

Selective attention seems to be controlled by grouping process [35,38], as the number of image items that can be processed in parallel within the 'spot-light' of attention depends on stimulus parameters affecting perceptual grouping. Two image items having the same spatial frequency, or location (e.g Gabor signals [35]) require the same processing time as a single item to be identified, implying some similarity and proximity based grouping processes preceding attention allocation. Excitatory interactions can underlie these grouping process, with a similarity metric defined by filter parameters (orientation, spatial frequency and location). It is interesting to note that grouping processes (see section 3.3) are critically dependent on the availability of attention [39,40], in agreement with the assumption of attentive control over the efficacy of excitatory interactions. Thus attention and grouping seem to be mutually dependent. Increasing efficacy of excitatory interactions enables long range correlations to affect grouping and thus an optimization of attentive resources allocation.

Duncan and Humphreys [38] suggested that hierarchical segmentation and grouping is a basic capacity-free stage in visual processing. This visual selection theory has three components: a parallel stage of perceptual description, a selection process and the entry

of selected information into visual-short-term-memory (VSTM) which allows control of access to awareness. According to this theory, the perceptual description is made by a process of hierarchical segmentation of the image into linked groups and subgroups (structural units). Each structural unit is described by its elementary sensory properties (relative position, color, size, motion etc) and categorical properties (based on meaning). The above described process of segmentation and description is said to be parallel and resource free, selection starts when all parts of the display compete for access to the VSTM. Linked structural units tend to gain or lose resources (activity strength) together. Increased assignment of resources to any structural unit increases its speed and probability for access to VSTM. Once a structural-unit emerges the 'winner' in the visual selection process it accesses VSTM. Duncan and Humphreys [38] propose that "Structural units act as wholes, competing for and gaining access to VSTM with all their associated descriptions". It is possible that structural units are created by linking (grouping) processes within the early filter representation. If this is the case, attention allocation should be constrained by early vision architecture, that is, by excitatory connectivity. Filters that are connected by excitatory interactions can be linked to create a 'structural unit'. The linking process would involve 'attentive' fast synaptic modifications to strengthen the desired links, or to weaken the links between other filters. However, on this account, the linking processes is not 'resource free' as the system is limited by connectivity, and to some extent by the flexibility of the attentive synaptic gating mechanism.

5. From attention to memory

Lateral interactions are probably modulated by attention on a fast time scale (msecs) and can go under longer term (years) modifications as in perceptual learning. These two processes operate on different time scales but are probably linked through some intermediate memory structures operating on time scales of several minutes to a few hours. Experiments involving lateral masking learning show an increase of interaction range with practice [4]. In order for practice to be successful, observers had to practice all target to mask distances within each practice session. Learning was not observed when observers were practicing on long distances alone, without performing on intermediate distances. Increasing the difference between the distance samples used in the experiments also prevented learning [4]. It seems that different distances have to be experienced within a single session, probably with neighboring filters activated within a time window of a few minutes, enabling generation of associations between stimuli that are close in space and time.

Further evidence for a low level memory operating on a time scale of a few minutes come from visual imagery experiments. In these recent experiments [55] observers were asked to imagine the high contrast masks while detecting the low contrast target in the lateral masking experiment (see Figure 2). Results from these imagery experiments show enhanced detection following the perception effects, though at about half the magnitude. However, enhancement by imagery can be obtained only within a time window of a few

minutes after *performing* the ‘real’ perception task, and with the same targeted eye. Passive inspection of the mask prior to the imagery action, or running only as few as ten trials before the imagery task are not sufficient to obtain the imagery based enhancement. These results support the existence of a low level (monocular) memory, probably iconic in nature, that stores filters activity for a few minutes. This storage is activated by higher level processes (e.g. attention, as task is necessary) and is accessible by higher level processes (as imagery). Thus this memory seems to match the one that is used for establishing spatio-temporal associations. [Though spatio-temporal associations are established during a time period of a few minutes, a few hours may be required for their consolidation [45,54].]

6. Conclusion

Non-isotropic excitatory interactions between oriented spatial filters provide activity enhancement from neighboring filters, thus enabling some useful global image characteristics to spread within the network of early vision and to be detected (e.g. enhancement of long lines [3,53], closure [48]), enhancement of closed contours [52]). Furthermore, it is assumed here that excitatory weights are modulated by a top-down process (i.e. visual attention), introducing memory dependent context. It is suggested that excitatory transmission efficacy is set to some low default value under nonattentive conditions and it can increase with increasing levels of attention. Though a global control of excitatory efficacy is assumed here, it is also possible that some specific weights, or pattern of weights, can be modulated by selective attention providing a fast space invariant object filter.

Spatial selectivity is inherent in the concept of visual attention as visual detection is enhanced within attended regions of the visual field [36,46]. Attention was also suggested to be involved in binding image features to objects [7,9], thus serving as the bridge between featural image representation and the perceived Gestalt. Within the context of the present framework, binding can be achieved by selective modulation of weights, probably determined also by memory stored patterns, with objects parts linked by increased connectivity between their corresponding filters. Note that we do not allow for memory dependent context to be effective without any input dependent context available, as memory affects only connectivity. (Transmission efficacy may also be controlled by weak top-down signals providing subthreshold input to filters and thus improving their responsiveness to other inputs [56], sensory as well as lateral.) Visual inputs should assume some spatial correlations in order for connectivity modulations to be efficient. However, a rich enough (or noisy) input can generate patterns of spurious correlations with sufficiently organized filter activity to enable modulation by memory. The effect of these spurious meaningful images on perception will depend then on the weighting factors, which may vary with time, or between brains.

REFERENCES

1. Porter, P. (1954). Another picture-puzzle. *American Journal of Psychology* 67, 550–551.
2. Karni, A. and D. Sagi (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Science, USA* 88, 4966–4970.
3. Polat, U. and D. Sagi (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research* 33, 993–999.
4. Polat, U. and D. Sagi (1994). Spatial interactions in human vision: from near to far via experience dependent cascades of connections. *Proceedings of the National Academy of Science, USA* 91, 1206–1209.
5. Polat, U. and A. M. Norcia (1996). Neurophysiological evidence for contrast dependent long range facilitation in the human visual cortex. *Vision Research*, in press.
6. Sagi, D. and D. Tanne (1994). Perceptual learning: learning to see. *Current Opinion in Neurobiology* 4, 195–199.
7. Julesz B. (1981). Textons, the elements of texture perception and their interactions. *Nature* 290, 91–97.
8. Marr, D. (1982). *Vision*. Freeman, New-York.
9. Treisman, A. (1985). Preattentive processing in vision. *Computer Vision Graphics and Image Processing* 31, 156–177.
10. Hartline H.K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology* 121, 400–415.
11. Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *Journal of Physiology, London* 119, 69–88.
12. Young, T. (1802). On the theory of light and colors. *Philosophical Transactions of the Royal Society* 92, 12.
13. von Helmholtz, H. (1909). *Physiological Optics*. Optical Society of America, 1925. Republished by Dover, New York.
14. Hubel, D. and T. Wiesel (1962). Receptive fields, binocular interaction and functional architecture in the cat's striate cortex. *Journal of Physiology, London* 166, 106–154.
15. Campbell, F. and J. Robson (1968). Application of fourier analysis to the visibility of gratings. *Journal of Physiology, London* 197, 551–566.
16. Graham N. (1977). Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels. *Vision Research* 17, 637–652.
17. Watson A.B. (1983). Detection and recognition of simple spatial forms. In *Physical and Biological Processing of Images*. Eds. Braddick O.J. & Sleight A.C., Berlin: Springer-Verlag.
18. Wilson H.R. (1983). Psychophysical evidence for spatial channels. In *Physical and Biological Processing of Images*. Eds. Braddick O.J. & Sleight A.C., Berlin: Springer-Verlag.
19. Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual

- psychology? *Perception* 1, 371-394.
20. Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America* 70, 1297-1300.
 21. Campbell, F. and J. Kulikowski (1966). Orientation selectivity of the human visual system. *Journal of Physiology, London* 187, 437-445.
 22. Phillips, G. C. and H. R. Wilson (1984). Orientation bandwidths of spatial mechanisms measured by masking. *Journal of the Optical Society of America A* 1, 226-232.
 23. Blakemore, C. and F. Campbell (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology, London* 203, 237-260.
 24. Stromeyer C.F. and Julesz B. (1972). Spatial-frequency masking in vision: Critical bands and spread of masking. *Journal of the Optical Society of America*, 62, 1221-1232.
 25. Legge G. E. (1979). Spatial frequency masking in human vision: binocular interactions. *Journal of the Optical Society of America*, 69, 838-847.
 26. Sagi, D. and S. Hochstein (1983). Discriminability of suprathreshold compound spatial frequency gratings. *Vision Research*, 23, 1595-1606.
 27. Watson A.B., Barlow H.B. and Robson J.G. (1983) What does the eye see best? *Nature*, 302, 419-422.
 28. Zenger, B and D. Sagi (1996). Isolating excitatory and inhibitory non-linear spatial interactions involved in contrast detection. *Vision Research*, in press.
 29. Gilbert, C.D. (1993) Circuitry, architecture and functional dynamics of visual cortex. *Cerebral Cortex* 3, 373-386.
 30. Polat, U. and D. Sagi (1994). The architecture of perceptual spatial interactions. *Vision Research* 34, 73-78.
 31. Foley, J. M. (1994). Human luminance pattern-vision mechanisms: masking experiments require a new model. *Journal of the Optical Society of America A* 11, 1710-1719.
 32. Heeger, D. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience* 9, 181-197.
 33. Georgeson, M. A. and J. M. Georgeson (1987). Facilitation and masking of briefly presented gratings: Time-course and contrast dependence. *Vision Research* 27, 369-379.
 34. Sagi, D. (1990). Detection of an orientation singularity in gabor textures: Effect of signal density and spatial frequency. *Vision Research* 30, 1377-1388.
 35. Adini, Y. and D. Sagi (1992). Parallel processes within the "spot-light" of attention. *Spatial Vision* 6, 61-77.
 36. Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology* 32, 3-25.
 37. Moran, J., and R. Desimone (1985) Selective attention gates visual processing in the extrastriate cortex. *Science* 229, 782-784.
 38. Duncan, J. and G. W. Humphreys (1989). Visual search and stimulus similarity. *Psychological Review* 96, 433-458.

39. Ben-Av, M. B., D. Sagi and J. Braun (1992). Visual attention and perceptual grouping. *Perception & Psychophysics* 52, 277-294.
40. Mack, A., B. Tang, R. Tuma, S. Kahn, and I. Rock (1992). Perceptual organization and attention. *Cognitive Psychology* 24, 475-501.
41. Fogel, I. and D. Sagi (1989). Gabor filters as texture discriminator. *Biological Cybernetics* 61, 103-113.
42. Landy, S. L. and J. R. Bergen (1991). Texture segregation and orientation gradient. *Vision Research* 31, 679-691.
43. Malik, J. and P. Perona (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A* 7(5), 923-932.
44. Rubenstein B. S. and Sagi D. (1990) Spatial variability as a limiting factor in texture discrimination tasks: Implications for performance asymmetries. *Journal of the Optical Society of America A*, 7, 1632-1643.
45. Karni, A. and D. Sagi (1993). The time course of learning a visual skill. *Nature* 365, 250-252.
46. Sagi, D. and B. Julesz (1986). Enhanced detection in the aperture of focal attention. *Nature* 321, 693-695.
47. Braun, J. and D. Sagi (1990). Vision outside the focus of attention. *Perception & Psychophysics* 48, 45-58.
48. Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt Brace.
49. Ben-Av, M. B. and D. Sagi (1995). Perceptual grouping by similarity and proximity: Experimental results can be predicted by intensity autocorrelations. *Vision Research*, 35, 853-866.
50. Malsburg, C. von der (1985). Nervous structures with dynamical links. *Physical Chemistry* 95, 703-710.
51. Braun, J., E. Niebur, H.G. Schusler and C. Koch (1995). Perceptual contour completion: a model based on local, anisotropic, fast-adapting interactions between oriented filters. *Society of Neurosciences Abstracts* 20, 1665.
52. Kovács, I. and B. Julesz (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Science USA* 90, 7495-7497.
53. Field, D. J., A. Hayes, and R. F. Hess (1993). Good continuation and the association field: evidence for local feature integration in the visual system. *Vision Research* 33, 173-193.
54. Polat, U., and D. Sagi (1995) Plasticity of spatial interactions in early vision. In B. Julesz & I. Kovács (Eds.), *Maturational Windows and Adult Cortical Plasticity*, SFI Studies in the Sciences of Complexity, Vol XXIV, Addison-Wesley, Reading MA.
55. Ishai, A., and D. Sagi (1995). Common mechanisms of visual imagery and perception. *Science* 268, 1772-1774.
56. Aertsen A.M.H.J., G. Gerstein, and P. Johannesma (1986). From neuron to assembly: Neuronal organization and stimulus representation. In G. Palm & A. Aertsen (Eds.), *Brain Theory*, Springer-Verlag, Berlin.



Figure 5.
A hidden figure: solution.