Cepstral Filtering on a Columnar Image Architecture: A Fast Algorithm for Binocular Stereo Segmentation

YEHEZKEL YESHURUN AND ERIC L. SCHWARTZ

Abstract—**Many primate visual cortex architectures (including the human) have a prominent feature responsible for the mixing of left and right eye visual data: ocular dominance columns represent thin (about 5–10 minutes of arc) strips of alternating left and right eye input to the brain. In the present paper we show that such an architecture, when operated upon with a cepstral filter, provides a strong cue for binocular stereopsis. Specifically, the vector of binocular disparity may be easily identified in the output of the (columnar based) cepstral filter. This algorithm is illustrated with application to a random dot stereogram and to natural images. We suggest that this provides a fast algorithm for stereo segmentation, in a machine vision context. In a biological context, this may provide a computational rationale for the existence of columnar systems, both with regard to ocular mixing, and to other visual modalities which have a columnar architecture.**

Index Terms—**Brain, neural networks, segmentation, stereo, vision.**

I. Introduction

In this paper, an algorithm is presented which provides a fast, one step analysis of the binocular disparity of a pair of stereo images. This work is strongly motivated by architectural features of the visual cortex of monkeys and humans, and it has a close relationship to certain limitations and advantages which are shared with human stereo vision. Free use has been made of anatomical and psychophysical data, in the explanation and discussion of the algorithm. However, there is no attempt in the present work to construct a model of human stereopsis: we propose this work as a favorable algorithm for computational stereo applications.

Among the many cues which humans use for inferring the structure of a three dimensional scene, binocular stereopsis has been one of the most intensively investigated (both algorithmically and psychophysically). This cue is based on an (unknown) means of utilization of the slightly different views of a three dimensional scene, as projected onto the right and left retina. From a generic point of view, this problem reduces to a matching or correlation of two slightly different scenes, in order to find the (vector) displacement of small corresponding patches of projected image.

Although apparently simple, fast solutions to this problem have been elusive. Many proposed algorithms are based on a direct search for matching features in the left and right half-images. Although some recent approaches base this search on a relaxation, or variational approach, possibly utilizing multiscale image data structures (e.g., [1]), this approach has become less common: it has been pointed out [2] that inherently sequential approaches to early vision, such as relaxation (or cooperative) algorithms, are biologically implausible. The low pulse rate of cortical neurons, together with the very rapid response time of biological stereo segmentation argue against a sequential approach to stereo segmentation. Also, from a machine vision viewpoint, it would seem desirable to have a "one-shot" fast algorithm for stereo segmentation.

The present algorithm achieves this goal by using local context, in parallel, to arrive at a "one shot" measurement of the disparity vector.[1] A windowed cepstral filter, operating on an interlaced image format which is inspired by the structure of ocular dominance columns in primate visual cortex, provides this "one-shot" performance.

The basic idea of the algorithm is to apply a cepstral filter [3] to a stereo image which is formatted in a way suggested by the ocular dominance column pattern of primate visual cortex [4]. Specifically, this pattern presents the left and right images of a stereo pair in the form of thin "strips" of image, alternating between left and right half-images. Fig. 1 shows an example of a computer graphic reproduction of this pattern, reconstructed in our laboratory from the brain of a macaque monkey.

Cepstral filtering is a well known method of measuring auditory "echo": the power spectrum of the log of the power spectrum of an audio signal with an echo present has a strong and easily identified component which is a direct measure of the echo period [5], [3]. The binocular disparity measurement then reduces to that of application of a nonlinear local filter (cepstral filter), followed by peak detection. This filter is applied within windows which span an ocular dominance column pair. This approach

Manuscript received April 1, 1987; revised January 8, 1989. This work was supported by the Air Force Office of Scientific Research under Grant 85-0341 and by the Nathan S. Kline Research Institute.

Y. Yeshurun was with the Computational Neuroscience Laboratories, Department of Psychiatry, NYU Medical Center, 550 First Avenue, New York, NY 10016. He is now with the Department of Computer Science, Tel Aviv University, Tel Aviv, Israel.

E. L. Schwartz is with the Computational Neuroscience Laboratories, Department of Psychiatry, NYU Medical Center, 550 First Avenue, New York, NY 10016, and the Department of Computer Science, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012.

IEEE Log Number 8927503.

[1]Note that further processing of this sparse disparity map to "fill in" smooth surfaces may well require more complex processing, including cooperativity. We discuss only the first stage of obtaining a sparse disparity map in this paper.

0162-8828/89/0700-0759$01.00

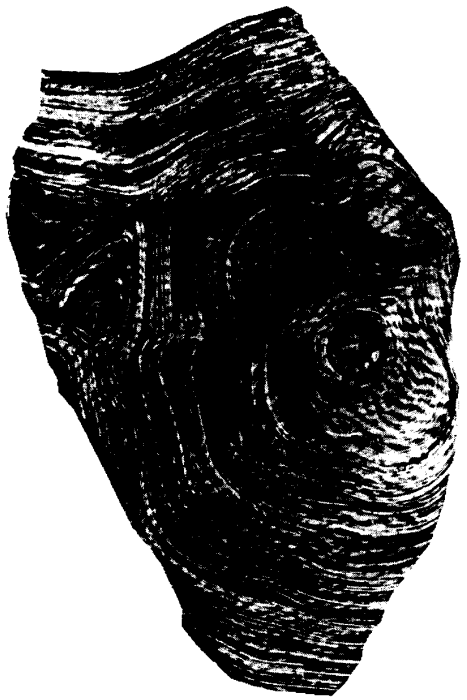

Fig. 1. The pattern of ocular dominance columns of a macaque monkey. The cortex has been digitally sectioned in the true tangential plane [26], numerically flattened with minimal error, and then the gray scale values of tissue stain have been texture mapped onto the flat cortical model. The data were obtained from a one eyed monkey whose brain was subsequently stained with a metabolic marker (for cytochrome oxidase). Darker regions indicate higher values of metabolism, i.e., they correspond to the present, active eye. The periodic pattern of the ocular dominance column system is clearly visible. In the context of the present paper, each of the dark stripes represents image input from the left eye, and each of the lighter stripes represents image input from the right eye. These strips are about 400 μ wide, and correspond to perhaps 5 minutes of arc in the visual field.

yields a strong stereo signal when presented with both natural images and random dot stereograms, and is resistant to image degradations such as blur, size difference, and intensity changes.

Because this algorithm is applied within fixed "windows," it cannot resolve changes in binocular disparity which occur within the window size. However, this is in agreement with the properties of primate stereo vision. Tyler has shown that humans cannot resolve changes in binocular disparity which vary over a scale finer than about 10 minutes of arc [6].[2] This is consistent with our algorithm, as the window size determined by the scale of the ocular dominance column pattern of monkey visual cortex lies within this region of angular size. Thus, the windowing method of our algorithm is consistent with the known spatial resolution of primate stereo vision. Our algorithm, like the human visual system, has extremely high stereo-acuity, but relatively low spatial acuity.

The present algorithm presents several unique properties:

1) It provides a computational justification for the existence of columnar interlacing, which is a common architectural feature of primate cortex.

2) It provides a "one shot," or purely parallel algorithm. There is no iterative component. Indeed, this algorithm could be implemented in a straightforward way by means of optical systems. It thus provides a candidate for "real time" performance in stereo segmentation.

3) It is in agreement with known psychophysical limitations and anatomical properties of stereo vision.

We will now briefly describe some architectural features of the primate visual system, and then the details of the algorithm.

II. Ocular Dominance Column Pattern

Many primates, including humans, possess "ocular dominance columns" in the primary visual cortex [4], [7]. The left and the right eye views of a scene interact for the first time at the level of striate cortex, and this interaction begins with the formatting of binocular data as thin (0.5 mm monkey; 1 mm human) strips of cortex which receive terminals from either the left or right eye. Fig. 1 shows a reconstruction of this pattern. Cortical magnification factor in the macaque monkey is estimated to be about 10–20 mm/deg [8], [9]. Since the width of a pair of columns is about 1 mm in macaque monkeys, the angular extent of a column pair is about 1/10–2/10 degree, or 6–12 minutes of arc.[3] Since visual acuity is about 1 minute of arc, a column pair extends over about 6–12 "resolution" units for acuity.

The function, if any, of the ocular dominance column pattern is currently unknown. One of the principal motivations of the present research has been to investigate algorithms which might be related to architectural features of striate cortex, such as this columnar pattern. For purposes of this discussion, the most notable feature of this pattern is that small patches of the left and right eye view of the scene are placed next to one another in layer IV of striate cortex. Fig. 2(a) schematically illustrates this situation. The task of stereo segmentation is to determine the vector displacement of these two image patches.

III. Cepstral Filtering

There are two aspects to the present algorithm. The first of these depends on the concept of "cepstral filter." The second depends on the use of a columnar image architecture. The Appendix provides a more detailed analysis of

[2]It is necessary here to distinguish between stereo-acuity, which represents the ability of humans to discriminate two nearby depth planes, and stereo positional accuracy. Stereo acuity is extremely high (about 2 arc seconds): humans can discriminate two stimuli which differ in depth by about 100 microns at three foot viewing distance. Stereo positional accuity, however, represents the ability of humans to perceive rapid changes in depth. This cannot be done when the rate of depth change is greater than 1 cycle/10 minutes of arc, which is about 10 times coarser than monocular spatial acuity.

[3]Note that we use the fact that the within-column magnification factor is really twice that of the global magnification factor usually measured, due to the double representation of columns within layer IV of cortex, as discussed by [4].

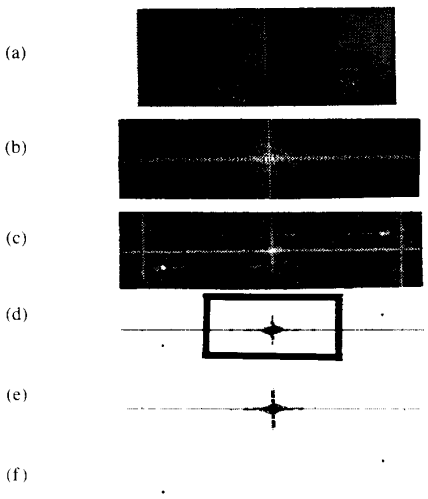

Fig. 2. (a) A pair of image patches. There is both horizontal and vertical disparity. (b) The power spectrum of (a). The origin of the frequency plane has been shifted to the center of the frame. (c) The cepstrum of (a). The disparity terms occur as bright spots in the cepstrum. These spots are easier to see in a thresholded version, shown in (d). The box marks a region of the cepstrum, described in the text, which does not need to be searched for a disparity signal. (e) The cepstrum of the left half image of (a). (f) The subtraction of the left cepstrum from the interlaced cepstrum. Only the two disparity "dots" remain.

some of the complexities which are associated with the "columnar architecture."

Consider an interlaced image $f(x, y)$ to be composed of a single columnar pair. Also, assume, for simplicity, that there is no binocular disparity, and that the data consist simply of an image patch $s(x, y)$ (the "right-eye" patch), and an identical patch "butted" against it (the left eye patch). Since there is assumed to be no disparity, the left and right eye images are identical, and the width of a single column is "D." We can mathematically represent this image pair as follows. (The * operator represents two-dimensional convolution.)

$$f(x, y) = s(x, y) * \{\delta(x, y) + \delta(x - D, y)\} \quad (1)$$

The Fourier transform of such an image pair is

$$F(u, v) = S(u, v) \cdot \{1 + e^{-i\pi(D-u)}\}. \quad (2)$$

By forming the logarithm of $F(u, v)$, the product structure becomes a sum:

$$\log F(u, v) = \log S(u, v) + \log (1 + e^{-i\pi D \cdot u}) \quad (3)$$

and the spectrum of (3) will have a prominent term located at the magnitude of the shift $(D, 0)$. In the Appendix, we derive the Fourier transform of the term $\log (1 + e^{-i\pi D \cdot u})$, and show that it consists of a principal term at the location $(D, 0)$, and a series of harmonics at integral multiples, with much smaller amplitude.

Equations (1)–(3) describe a simple case of two identical images, placed side by side (simulating a small section of "columnar" image). It is easy to extend this situation to cases where a left or right shift of one of the image pairs has occurred, in order to simulate "binocular disparity." The same kind of analysis will apply, and it is possible to locate a strong peak in the cepstrum of the image patch whose location is equal to the basic columnar shift D, with an additional term in the disparity added to (or subtracted from) this shift. This situation is best illustrated with a simple image example.

Fig. 2(a) shows a small patch of image. There is a horizontal and vertical component of disparity. Fig. 2(b) shows the power spectrum of this image. Fig. 2(c) shows the cepstrum (power spectrum of log power spectrum). In these power spectra, zero frequency has been shifted to the center of the figure, so that there is bilateral symmetry about the origin. There are two bright dots, representing the disparity term in the cepstrum. These dots are made apparent by thresholding the cepstrum, as in Fig. 2(d). Note that the units of the cepstrum are the same as the original physical units of the image. This can be demonstrated by measuring the linear spacing of two features in the image plane (e.g., the eyes), and then measuring the distance from the origin to the peak in the cepstrum. Thus, spatial position in the cepstrum is a direct measure of the "disparity" of the left and right half images.

The image dependent terms in the cepstrum (i.e., the cepstrum of the half-images themselves) can be removed easily, since we have access to the two half-images. Fig. 2(e) shows the cepstrum of the left half image of Fig. 2(a). Subtracting this cepstrum from the total cepstrum, in Fig. 2(f), we isolate the disparity vector, shown as two dots whose vector displacement from the origin is the disparity vector.

In practice, we have not found it necessary to subtract the image dependent terms, as shown in Fig. 2(f). A simple peak detection algorithm has been capable of isolating and measuring the binocular disparity directly from the cepstrum of the columnar image pair, as shown in Fig. 2(c).

The key idea in this work is that the columnar interlacing shifts the disparity term, in the cepstral plane, by an amount "D," where "D" is the column size. Then, as long as peak detection is restricted to a region of the cepstral plane within $[D/2, 3D/2]$, it is quite easy to locate this term as a "bright dot" or peak, with no competing areas of high intensity (see the Appendix for a detailed discussion). Thus, in this case, it is a simple matter to measure the disparity. It is quite interesting to note that the present algorithm has its simplest case for disparities within the region of the width of a column, because human vision also finds its easiest stereo task when presented with stereo pairs whose disparity is within a psychophysically defined region called "Panum's area" [10], [11]. In humans, Panum's area corresponds to 5–10 minutes of arc, and it also corresponds to the estimated extent of human ocular dominance columns [12]. These issues are discussed more extensively in the Appendix.

IV. Summary of the Algorithm

Given a stereo pair, i.e., two images of the same scene, the algorithm is applied as follows.

1) The two images are interlaced to yield pairs of corresponding "patches." These patches can be of arbitrary shape and might even overlap, although only nonoverlapping rectangular areas are demonstrated here. One might also apply smoothing operators (e.g., Hanning windows) to the windowed images, although we have not found it necessary to do, so, as the stereo signal is extremely strong.

The size of the "window" is essentially determined by the size of the "ocular dominance columns" which cause the interlacing. Based on the human and monkey visual systems, this window size is estimated to be within the range of 5–10 minutes of arc.

2) Each window is processed (in parallel) by the cepstral operator.

3) The display vector is then found by a peak detection algorithm applied to the windowed cepstrum of the stereo pair. For disparities whose magnitude is $< D$, i.e., "Panum's area," the search is restricted to the region $[D/2, 3D/2]$ in the cepstral plane.

V. Performance of the Algorithm

The windowed cepstral filter is easy to implement (it depends on little more than access to an FFT algorithm). Its complexity is $O(N \cdot \log N)$, from the FFT stage. Full parallelism is easy to achieve, since the windowed filter can be run simultaneously (there is no interaction between neighboring image patches). Since this algorithm relies mainly on an ability to estimate power spectral densities, as well as a simple (logarithmic) nonlinearity, it would seem to be easy to implement in the context of optical computation.

Fig. 3 shows an example of a natural scene, in which a window of about 5–10 minutes of arc is indicated. In order to obtain sufficient resolution from this scene, we have photographically expanded this small segment of it, and performed the cepstral analysis on it. This corresponds to a pair of (foveal) ocular dominance column patches. Fig. 3(c) shows the cepstrum of the "columnar" pair of Fig. 3(b). The cepstral signal is evident as the two bright dots on the x-axis.

It is technically difficult to fully analyze a natural scene at this resolution, since we are essentially working on a scale in which human visual acuity is equivalent to the pixel size of the image. In other work, we show that this requires a (conventional, constant resolution) image of size 16 000 × 16 000 pixels (see Fig. 6). A more sensible path at this point would be to use (as does the human system) space variant image representations. However, we show in Fig. 4 a 3500 × 3500 random dot stereogram, fully segmented by the present algorithm. This stereo pair depicts a "pac man" that is detectable only by binocular cues. We have scaled this stereogram to match the following parameters of human vision. The size of each box of the stereogram subtends about 8 degrees of field, so that the percept within the box subtends about 5 degrees. The size of the windows in the box is 5 minutes of arc. Thus, the positional accuracy of this segmentation is no better

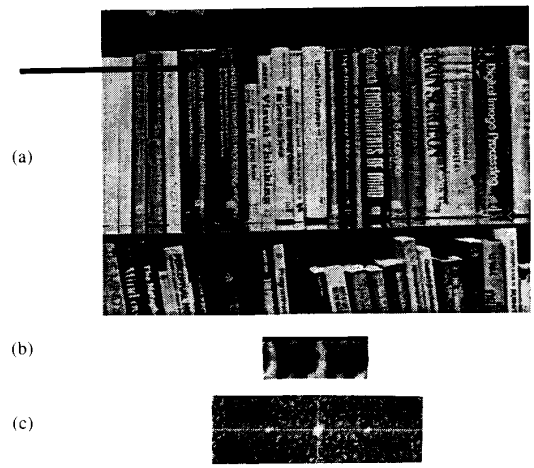

Fig. 3. (a) One frame of a stereo pair of a natural scene. (b) A patch of this stereo-pair, corresponding to five minutes of arc. The area of the patch is indicated by the arrow in the figure. (c) The cepstrum of the interlaced patches of (b).

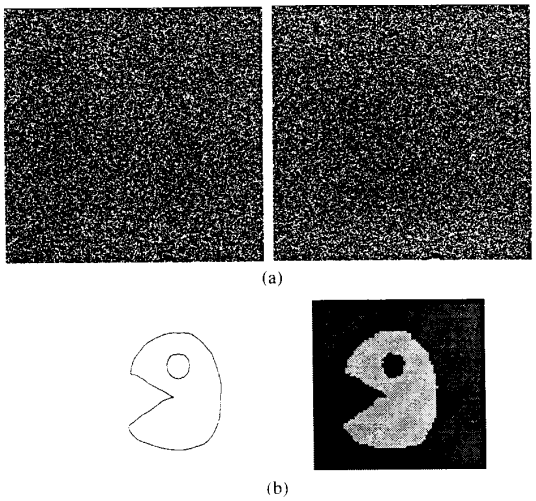

Fig. 4. (a) A random dot stereogram, in which a "pac-man" figure exists as a pure stereo signal. (b) The segmented figure, after windowed cepstral filtering and peak detection. The gray scale of the segmented figure is proportional to stereo disparity. The original stereo frames were constructed at 3500 × 3500 pixels, and correspond, for typical reading distances, to about eight degrees of field. In other words, if the size of the box is 4 cm, and is held at 30 cm from the eye, it will subtend 8 degrees. The window size used was 5 minutes of arc (32 pixels).

than 5 minutes of arc. This lack of fine detail is evident in slight aliasing of the boundaries of the stereo percept. Based on Tyler's measurements, this aliasing, or an equivalent lack of positional detail, should be present in human segmentation of the stereo pair shown in Fig. 4. Humans do not see aliased edges in random dot stereograms (based on our subjective experience), but merely cannot resolve fine positional detail, below about 10 minutes of arc, in pure stereo images. Stereo fusion, which we have not addressed, would need to be considered to complete the segmentation algorithm presented in this paper, and it would be of interest to compare the subjective

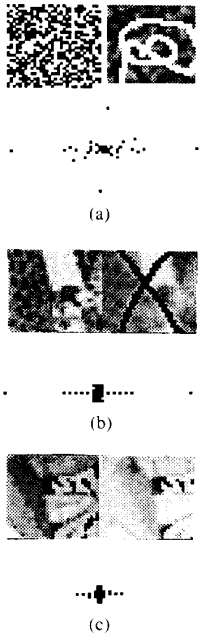

Fig. 5. (a) A random dot stereogram. The right image has been blurred (Gaussian convolution), and it has been "scribbled." (b) A small segment of a natural scene, in which blur and "scribbling" has also been applied. (c) A small segment of a natural scene, in which the left frame has been histogram equalized, changing its intensity values. The disparity peaks are clearly evident, outside of the region between $\{D/2\}$, $\{3D/2\}$, in each thresholded cepstrum.

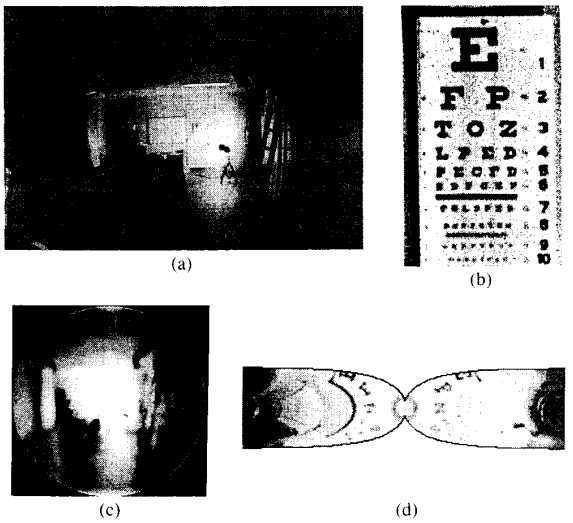

Fig. 6. (a) A wide angle fish eye view of a scene in the hall of our laboratory. A ladder is to the right, an eye chart is in the very center of the frame (almost invisible). The original version of this scene was digitized to an effective resolution of 16000 × 16000 pixels by a polar coordinate mosaic techniqiue. A "blow-up" of the central region of this original frame is shown in (b). This is an eye-chart, and the distance to the chart was 20 feet. In the original, line 7 of the chart could be easily read, indicating an effective "acuity" of 20/30, or about 1.5 minutes of arc. The purpose of this work was to simulate a wide angle scene (about 100 degrees), roughly comparable to human vision, at human visual acuity. (c) This scene blurred by a space variant filter which is modeled after human visual acuity. (d) The image of (a), modeled in terms of a complex logarithmic model [12] of human visual cortex. The eye-chart occupies almost half of the surface of visual cortex, although it occupies a tiny fraction of the original scene. The ladder, and the windows of the original are compressed to almost the same size as the centrally fixated letters of the eye-chart. This illustrates the tremendous space variant compression of human vision. Variations in linear size of about 100 : 1 (10^4 in solid angle) occur from the center to the periphery of the human visual system.

qualities of such a fusional algorithm with human performance.

A. Preprocessing of Stereo Images

We emphasize that the input to this algorithm is not necessarily a "gray scale" image, as we have used in the figures. Naturally, any preprocessing (e.g., high pass filtering, edge enhancement, etc.) is compatible with this approach. However, it is important to emphasize that we obtain good performance without image preprocessing. Simple gray scale data, as in the figures, is sufficient for good performance. This is in marked contrast to other stereo algorithms, which require edge enhancement or feature detection in order to perform at all.

Small regions of the image which have no spatial detail will of course fail to yield a disparity signal. This is true of any matching algorithm.

B. Robustness of Algorithm

We have found the columnar cepstral algorithm to resist a wide range of image degradations. Fig. 5(a) shows a pair of image patches (random dot stereogram) in which a Guassian blur was applied to the right frame, and then some random "scribbling" was added. The threshold cepstrum is shown below. Fig. 5(b) shows a pair of natural images, with the same Gaussian blur and scribbling applied. The thresholded cepstrum is also shown beneath it. Fig. 5(c) is a natural image pair, in which the left frame was histogram equalized,[4] and the right frame was not, with the thresholded cepstrum shown below. Size changes of up to 15 percent and rotations of ten degrees of one of the stereo frames can be routinely accepted by this algorithm. Considerable intensity changes can be applied to one of the stereo frames without disrupting the algorithm. In Fig. 5, we show an example in which one of the stereo pairs is "histogram equalized" and the other is left in its original (low contrast) state. The algorithm of the present paper was not disturbed by this intensity difference, nor by simple additive intensity increments of 50 percent to one image of a stereo pair. In fact, positive and negative stereo pairs can be processed with no difficulty, as is evident from the mathematical structure of the cepstral filter. Humans can fuse stereo pairs which differ considerably in intensity, and can fuse positive–negative pairs of line stereograms, but cannot fuse positive–negative random dot stereograms [14]. In other work [15], we show

[4]Histogram equalization is a method of expanding the contrast of an image. The histogram of gray levels of the original image is obtained, and a remapping of gray-scale levels is performed so that the resultant histogram of gray-scale values approximates some desired (e.g., uniform) distribution [13].

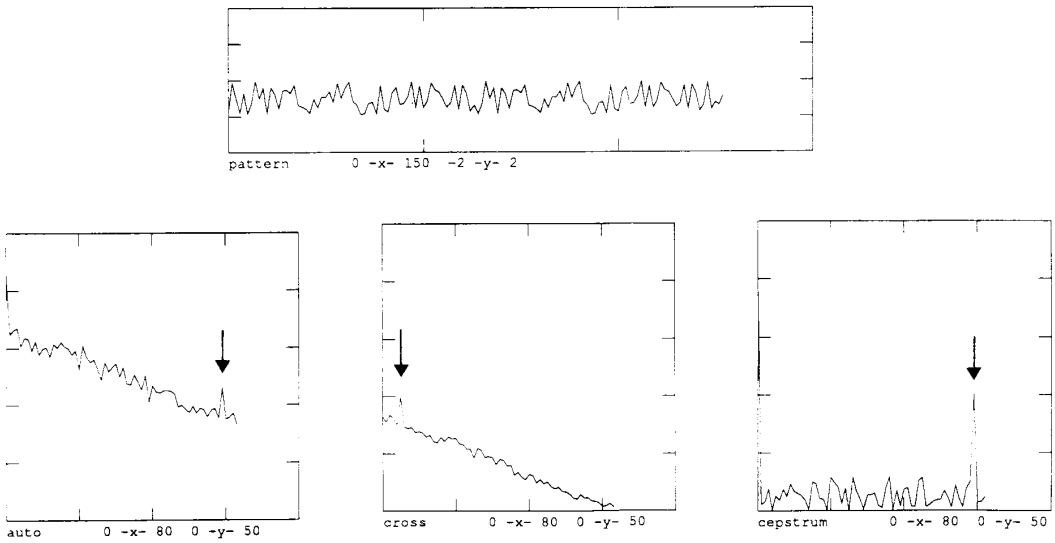

Fig. 7. The top shows a random (white noise) signal (1-D) which is repeated with a slight shift, to simulate a "columnar" stereo signal in one dimension. Below the auto-correlation and cross-correlation of this signal is shown. The "disparity" is indicated by the small peak in the correlation spectra (arrow). On the bottom, right the cepstrum of the same signal is shown. The peak indicating the disparity is has a much higher signal-to-noise ratio than that of the correlation examples.

that these details are compatible with a cepstral filter. Similarly, humans can fuse stereo pairs which have size differences of up to 15 percent and rotational differences of up to 15 degrees. In these respects, the cepstral algorithm has similar robustness to human stereo vision.

C. Relation to Other Correlational Methods

The ceptral filter is closely related to auto-correlation. The question naturally arises whether auto-correlation would perform as well as the cepstral filter in this application. Similarly, does simple cross-correlation of two stereo scenes perform well?

Our experience with auto-correlation applied to a columnar architecture, using the same data with cepstral filtering, is that the cepstral filter has superior signal-to-noise properties.

Fig. 7 shows an example of the same 1-D data set, formatted as a single left-right "column," and processed by auto-correlation, cross-correlation, and cepstrum. The cepstral transform clearly has superior signal-to-noise ratio.

Since the computational cost of the cepstrum is virtually identical (both in complexity and in reality!) to auto-correlation, and its performance on columnar images is clearly superior, it is the method of choice in the present context.

VI. Discussion

Recent approaches to stereo segmentation (see [10] for review) typically have a sequential, iterative, or relaxation component. Thus, some search procedure (e.g., relaxation, or search over multiple scales) is used to resolve ambiguities in local image matches. As Marr [2] has pointed out, these approaches do not seem biologically plausible: the ability of biological systems to function in extremely short time intervals seems to argue against elaborate variational, cooperative, or relaxation processes. In fact, any iterative step at all seems questionable when considering the performance of preattentive segmentation in humans: a time period of about 200 ms is enough for a complete preattentive segmentation of a complex scene. Yet, 200 ms is about the same amount of time it takes a signal to propagate through all layers of visual cortex, from the retina. The implication is that preattentive segmentation in humans is a "one-shot" procedure, since it seems to occupy little more than a single pass through the cortical "machine."

The cepstral algorithm described in this paper is purely parallel, so it is not subject to Marr's critique of iterative algorithms. Without columnar image format, this algorithm would be problematic. If the two images were simply superimposed (e.g., by addition), the performance would be severely degraded, since there would be no clear "echo" signal. Moreover, the very small disparity terms in the cepstrum would be masked by neighboring frequency components in the images.

A. Limitations of this Algorithm

The principal limitation of this algorithm is that, insofar as it is a "textural" approach, it can only operate in windows of size greater than the resolution limit of the image.

Ocular dominance columns are perhaps five times larger than the basic acuity limit of the visual system. This is a

comfortable range for our window operator, and is in agreement with psychophysical measurements of the spatial limits of disparity sensitivity in humans.

Other algorithmic approaches, based on local feature matches for example, might be capable of extending this limit to within the range of acuity. However, they do so at considerable computational expense, and would, in the process, exceed the actual abilities of the human visual system. (We are implicitly considering the performance of human vision as optimal in this context, rather than consider an unconstrained definition of the term "vision.")

B. Biological Implementation and Relation to "Spatial Frequency Channels"

The method of implementation which we have used to simulate the cepstral filter requires little more than access to the spatial frequency content (power spectrum) of an interlaced stereo pair. For many years, there has been considerable interest in the "spatial frequency" tuning properties of the human visual system. It is interesting to note that an appropriate set of bandpass frequency filters are adequate to implement the cepstral filter of this paper. In pilot work, we have succeeded in simulating an estimation of the power spectrum of a columnar image using medium bandwidth (1.5 octave) filters to provide cepstral estimates which are comparable to those shown in the present paper, which were obtained from a digital FFT.

It is important to emphasize that access to a Fourier transform is not necessary for the present algorithm. Thus, phase information is not required. Estimates of power spectral density which could be provided by simple space domain filtering are sufficient.

We have not yet studied the implications of cepstral filtering in a biological context. Nevertheless, it is interesting to point out that a recursive application of spatial filtering, as in the cepstral filter, can have powerful image processing capabilities if the underlying architecture (e.g., columns) is correct. Finally, a prediction associated with the present algorithm is worth mentioning. Although binocularly tuned neurons are known to exist in primate visual cortex [16], it is not known whether there is any systematic spatial organization (e.g., disparity columns). An arrangement of disparity tuning in a direction perpendicular to the boundaries of ocular dominance columns would be a strong support for the cepstral filtering mechanism proposed in this paper.

C. Generalization to Other Columnar Systems

The columnar architecture which is at the basis of the present algorithm is common in the primate neo-cortex. Orientation of edges[5] and direction of motion are two other modalities which are known to be formatted in columnar terms. Color and spatial frequency have been claimed to have columnar architecture, and regions of frontal cortex with unknown function are also known to have a columnar architecture [17]. In earlier work, the frequency modulation aspect of ocular dominance columns was pointed out [18], and it was suggested that columnar systems occur in cortex whenever two slightly different modalities need to be compared, and their differences extracted. The cepstral operator is sufficiently simple, and the columnar architecture sufficiently widespread, to provide a hope that perhaps a generic operation of the visual system is provided by the analysis of this paper. If so, then support would be provided for the notion that the visual system uses the elaborate functional architecture which it has constructed for computational purposes.

D. Relation to Other Psychophysical Phenomena and Other Models of Stereo

Many other biological and machine approaches to steteo have been proposed (see [10], [19], [20] for review, and [21], [1], [22], for some examples of well known models). However, none of this work makes use of the formatting of stereo data in visual cortex, in the form of ocular dominance columns, as is the basis for the present algorithm.

There are also many complex psychophysical details related to stereo vision. For example, the classical concept of Panum's area used in the present paper has been revised due to work of Burt and Julesz [23]. It must be emphasized that our present work is not meant to provide a biological theory of stereo. We merely seek to point out that some of the first-order spatial constants of stereo vision match in a rough way with the size of the ocular dominance column system, and may indicate that a columnar-windowed algorithm, of the kind outlined in this paper, may occur at one or more levels of the visual system. Note that the ocular dominance columns of V-1 provide one scale, as discussed in the text, but that ocular dominance may occur also in V-2 and perhaps elsewhere. This may explain some of the complexity in the details of the concept of "Panum's area."

E. Space Variant Vision

In order to make effective use of the algorithm of this paper, it is necessary to use windows which approximate the size of a pair of human ocular dominance columns. This is estimated to be about 6–12 minutes of arc (foveally). It is reasonable to require perhaps 20 pixels in such a window. Thus, a 512 × 512 pixel image would span only a few degrees. Clearly, a much higher resolution sensor than 512 × 512 is desirable if any appreciable angular extent is to be covered.[6] An interesting alternative would be to use a space variant sensor (i.e., a foveal sen-

[5]The orientation column system of primate cortex would provide, when operated upon by the cepstral algorithm, a measure of boundary curvature: difference of orientation is curvature. Thus, stereo and boundary curvature extraction could be provided by the same underlying mechanism, operating on the respective columnar systems of ocular dominance and orientation selectivity.

[6]Note that the random dot stereogram of Fig. 4 was of size 3500 × 3500, in order to cover 8 degrees of simulated field at 5 minute/window resolution!

sor), patterned after the human visual system [12]. This would allow a wide angle of coverage, but still provide a reasonable allocation of sensor resources, mimicking the approach of biological visual systems. In other work, we have begun to study the image processing aspects of such space variant systems (see Fig. 6 for a computer simulation of a natural scene, imaged by a space variant system such as primate visual cortex) [24], [25]. It is interesting to note the possible synergy between the biologically motivated windowed cepstral filter, and the need to consider space variant (e.g., logarithmic) visual systems.

VII. Conclusion

Computation consists of algorithms applied to data structures. By means of formatting visual data, in the brain, in terms of adjacent "columns," a disparity signal is modulated onto a particular component of the cepstrum of the cortical image. Thus, good data structure (columnar interlacing) allows a simple algorithm (cepstral filter) to provide a robust solution to a computationally intensive application.

A possible lesson from biological vision may be the importance of using simple algorithmic procedures, applied to novel spatial architectures. At the same time, there is a tradeoff of positional resolution for discrimination: humans can discriminate stereo differences with extraordinary accuracy (stereo-acuity), but cannot resolve high spatial frequencies in the stereo modality.

These design choices, which are not intuitively obvious, are the product of an extremely long "burn in": evolution. Perhaps robotics vision applications can profit from this process.

Appendix

A. Multiple Disparities Due to Large Windows

The log Fourier transform of an interlaced image pair is given by

$$\log F(u, v) = \log S(u, v) + \log (1 + e^{-i\pi D \cdot u}). \quad (1)$$

The power spectrum of this image is the cepstrum. We show that the second term of (1) above consists of a series of delta functions whose spatial location in the cepstrum is a measure of the shift D.

Using the expansion $\log (1 + z) = \Sigma_1^\infty (-1)^{n+1}(z^n/n)$, valid for $z\bar{z} \leq 1$, and $z \neq -1$, we can evaluate the Fourier transform

$$\int_{-\infty}^{\infty} e^{i\pi X u} \log (1 + e^{-i\pi D \cdot u})\, du = \sum_1^\infty (-1)^{n+1} \frac{\delta(X - nD)}{n}. \quad (2)$$

Thus, we get delta functions at locations which are integral multiples of the shift term D. The weight of these delta functions decreases rapidly, and, in particular, there is only a single delta function in the interval $[D/2, 3D/2]$.

B. Multiple Disparities Due to the Columnar Architecture

The situation is best illustrated graphically. Consider the table below:

L0	R1	L1	R2	L2	d1	d2	Actual Shift
xxxxxx	123456	123456	789ABC	789ABC	6	—	0
xxxxx1	123456	234567	789ABC	89ABCx	5	1	−1
xxxx12	123456	345678	789ABC	9ABCxx	4	2	−2
xxx123	123456	456789	789ABC	ABCxxx	3	—	−3
xx1234	123456	56789A	789ABC	BCxxxx	2	4	−4
x12345	123456	6789AB	789ABC	Cxxxxx	1	5	−5
xxxxxx	123456	x12345	789ABC	6789AB	7	13	1
xxxxxx	123456	xx1234	789ABC	56789A	8	14	2
xxxxxx	123456	xxx123	789ABC	456789	9	15	3
xxxxxx	123456	xxxx12	789ABC	345678	10	16	4
xxxxxx	123456	xxxxx1	789ABC	234567	11	17	5
xxxxxx	123456	xxxxxx	789ABC	123456	—	18	6

We have constructed a schematic model of "columns" 1 unit high and 6 units wide, depicted by L0 to L2. An image is simulated in this table as a sequence of the characters "123456789ABC", and "x" represents arbitrary data. The table summarizes the possible disparity matches which can exist as a function of the "actual shift" of the underlying images. The complication in this simulation is due to a "wrap-around" condition which is introduced by the existence of periodic columns of interlaced image, and by considering windows of more then two adjacent columns. Thus, the complexity of this situation is due to the actual architectural complexity of the primate visual cortex.

The possible disparity values which can exist are indicated in the table as d1 and d2. If the algorithm compares only adjacent colums (e.g., R1 and L1), then a single peak will be detected by the algorithm. However, its intensity is related to the degree of correspondence between these columns: for actual shift of −1, for example, the two columns share the part of the image represented by "23456", while for an actual shift of −4, this part is "56" only, which will yield a weaker cepstral peak. Thus, comparing two adjacent columns, disparities up to the size of the column can be detected, but the corresponding peak is attenuated as the shift grows, until it vanishes when it reaches the size of the column.

Another point which should be addressed is the possibility that the "disparity signal" will be masked by the

signal of the image [see (1)] for large negative shifts (e.g., -4 and -5 in the table). This could be avoided by either considering disparities in the range $[D/2, 3D/2]$ (i.e., actual shift of -3 to 3 in the table), or, as mentioned in the description of the cepstral filter, by subtracting the cepstrum of the image from the columnar cepstrum before the peak detection.

Thus, we feel that the execution of this algorithm is simplest for disparities whose range allows the shifted image terms to remain within neighboring windows. This will occur for disparities in the range of a single column width, which corresponds to about 6–12 minutes of arc.

Panum's area [11] refers to a range of stereo disparity over which humans can easily "fuse" stereo frames. It is possible for humans to process stereo at larger disparities, but special conditions are required: first, fusion is effected within Panum's area, and then the stereo frames are "pulled" slowly to larger disparities.

It is thus interesting to note that our column based algorithm has similar performance characteristics to human stereo. It is limited in positional resolution, as outlined in the text, but is also limited in the range over which it can most simply process stereo images. Both of these limitations are determined by the width of ocular dominance columns. Thus, there is a basic scale factor in human stereo, of magnitude 5–10 minutes/arc, which determines the limit in both positional resolution for stereo, and the limiting region for easy stereo processing. Our algorithm has very similar limitations.

Acknowledgment

We thank R. Hummel and D. Lowe for reading this work, and for making suggestions towards its improvement, and N. Brennan for editorial assistance.

References

[1] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proc. Roy. Soc. London*, vol. B204, pp. 301–328, 1979.
[2] D. Marr, *Vision*. New York: Freeman, 1982.
[3] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
[4] S. LeVay, D. H. Hubel, and T. N. Wiesel, "The pattern of ocular-dominance columns in macaque visual cortex revealed by a reduced silver stain," *J. Comput. Neurol.*, vol. 159, pp. 559–576, 1975.
[5] B. P. Bogert, W. J. R. Healy, and J. W. Tukey, "The frequency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking," in *Proc. Symp. Time Series Analysis*. New York: Wiley, 1963, pp. 209–243.
[6] C. W. Tyler, "Spatial organization of binocular disparity sensitivity," *Vision Res.*, vol. 15, pp. 583–590, 1975.
[7] J. C. Horton and E. T. Hedley-White, *Phil. Trans. Roy. Soc. London*, vol. B304, pp. 255–272, 1984.
[8] B. M. Dow, A. Z. Snyder, R. G. Vautin, and R. Bauer, "Magnification factor and receptive field size in foveal striate cortex of monkey," *Exp. Brain Res.*, vol. 44, pp. 213–228, 1981.
[9] D. C. Van Essen, W. T. Newsome, and J. H. R. Maunsell, "The visual representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability," *Vision Res.*, vol. 24, pp. 429–448, 1984.
[10] B. Julesz, "Stereoscopic vision," *Vision Res.*, vol. 26, pp. 1601–1612, 1986.
[11] C. W. Tyler, "Sensory processing of binocular disparity," pp. 199–295 in *Vergence Eye Movements: Basic and Clinical Aspects*, C. Schor and K. Ciuffreda, Eds. Boston, MA: Butterworths, 1983.
[12] E. L. Schwartz, "Computational anatomy and functional architecture of striate cortex: A spatial-mapping approach to perceptual coding," *Vision Res.*, vol. 20, pp. 645–670, 1980.
[13] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1977.
[14] B. Julesz, *Foundations of Cyclopean Perception*. Chicago, IL: University of Chicago Press, 1971.
[15] Y. Yeshurun and E. L. Schwartz, "Ocular dominance columns in macaque V1: A two dimensional stereomap," in *Computational Neuroscience*, E. Schwartz, Ed. Cambridge, MA: MIT Press, 1989.
[16] G. Poggio and B. Fischer, "Binocular interaction and depth sensitivity in striate cortex and prestriate cortex of behaving rhesus monkey," *J. Neurophys.*, vol. 40, pp. 1392–1405, 1977.
[17] P. Goldman and W. Nauta, "Columnar distributions of cortico-cortical fibers in frontal, associational, limbic and motor cortex of developing rhesus monkey," *Brain Res.*, vol. 122, pp. 393–413, 1977.
[18] E. L. Schwartz, "Columnar architecture and computational anatomy in primate visual cortex: Segmentation and feature extraction via spatial-frequency-coded difference mapping," *Biol. Cybern.*, vol. 42, pp. 157–168, 1980.
[19] F. A. Almgren and A. Arditi, "Binocular vision," in *Handbook of Perception and Human Performance*, K. Boff and L. Kaufman, Eds. New York: Benjamin, 1986.
[20] S. T. Barnard and M. A. Fischler, "Computational stereo," *Comput. Surveys*, vol. 14, 1982.
[21] J. E. W. Mayhew and J. P. Frisby, "Psychological and computational studies towards a theory of human stereopsis," *Artificial Intell.*, vol. 17, pp. 349–385, 1981.
[22] W. E. L. Grimson, in *From Images to Surface: A Computational Study of the Human Visual System*. Cambridge, MA: MIT Press, 1981.
[23] P. Burt and B. Julesz, "A disparity gradient limit for binocular fusion," *Science*, vol. 208, pp. 615–617, 1980.
[24] Y. Yeshurun and E. L. Schwartz, "Space-variant image-processing IV: Contour-based blending of multi-fixation log views of a scene," Computational Neurosci. Lab., NYU Med. Center, Tech. Rep. CNS-TR-11-86, 1986.
[25] E. Wolfson, Y. Yeshurun, and E. L. Schwartz, "Space-variant image-processing II: Image-blending of multi-fixation logarithmic views," Computational Neurosci., NYU Med. Center, Tech. Rep. CNS-TR-10-86, 1986.
[26] W. Light and E. L. Schwartz, "A digital tangential microtome built from a voxel-based surface tracker: The brain peeler," Computational Neurosci. Lab, NYU Med. Center, Tech. Rep. CNS-TR-5-86, 1986.
[27] E. L. Schwartz and B. Merker, "Computer-aided neuroanatomy: Differential geometry of cortical surfaces and an optimal flattening algorithm," *IEEE Comput. Graphics Applicat.*, vol. 6, pp. 36–44, 1986.

Yehezkel Yeshurun received the Ph.D. degree in mathematics from Tel Aviv University, Tel Aviv, Israel, in 1985.

From 1978 to 1985 he was the head of the systems programming group of the TAU computation center. During 1986–1987, he pursued postdoctoral studies at the Computational Neuroscience Labs of New York University Medical Center, New York, NY. He is currently with the Department of Computer Science of Tel Aviv University. His research interests are in computational vision, robotics, and neural networks.

Eric L. Schwartz (M'83) received the A.B. degree in physics and chemistry in 1967 and the M.A., M.Phil., and Ph.D. degrees in high energy physics from Columbia University, New York, NY.

He is currently Associate Professor at the Computational Neuroscience Laboratories of New York University Medical Center, is a Research Scientist at Nathan Kline Research Institute, and is an Adjunct Associate Professor of Computer Science at the Courant Institute of Mathematical Sciences of New York University. His current research interests are the anatomy and physiology of primate visual systems, computational vision and robotics, and computer and mathematical simulation of the nervous system. A major theme of this research is the application of mathematical and computer models to describe the various architectures that have been experimentally observed in primate visual systems, and to infer the possible computational significance of these architectures for both machine and biological visual computation.