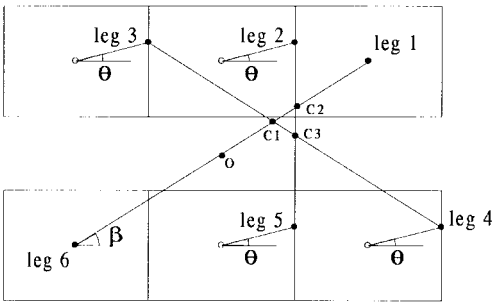


Fig. 16. Critical points for initially lifted pairs when $\alpha < \beta$ and $0 \leq \theta < \alpha$.

Remark: To use overlapping reachable cells in deriving the optimal fault tolerant locomotion in crab walking needs some precaution, since in extending the redefined reachable cells some unreachable areas are included in the cell. To explain more precisely, let us remind the shape of the overlapping reachable cell in Fig. 5(a), where unreachable regions are located in the upper right and left corners of the extended cell. Following the proposed sequences for crab walking, there might be cases where some lifted legs cannot place their feet on the front-end positions that are on the unreachable regions. For example, in the case of $\alpha < \beta$ and $\theta = \beta$, the front-end foothold position of leg 2 when it is initially lifted is the apex of the reachable cell in the upper right corner as shown in Fig. 15. But, if the hexapod has the overlapping reachable cells, the point is reduced to be in the unreachable region. Hence in using the overlapping reachable cells it is necessary to check if or not such a problem of kinematic limit occurs during the locomotion. If the hexapod has the problem inherently, the foothold positions should be changed to some feasible locations within the kinematic limit and consequently the proposed sequence should be adapted with changed foothold positions.

V. Conclusion

In this paper, we have shown that when the hexapod robot has the fault tolerant gait sequence in straight-line motion, each leg can have the overlapping redefined reachable cells of legs, improving the performance of the sequence with respect to the stride length. With overlapping reachable cells, the gait sequence for the locomotion in straight-line motion could be executed with the increased stride length of the center of gravity in one cycle, without causing any violation of the kinematic limit. In addition, we have presented that, as in straight-line motion, the optimal fault tolerant gait sequence of the hexapod for crab walking can be generated on perfectly even terrain. With the proposed sequence for crab walking, the hexapod could have fault tolerant capability and the maximum stride length in one cycle. It was shown that the order of lifting and placing of each leg in the proposed sequence is variant according to the relative values of the crab angle and the design parameters of the robot. The use of overlapping reachable cells in crab walking was also discussed.

References

[1] M. L. Visinsky, J. R. Cavallaro, and I. D. Walker, "A dynamic fault tolerance framework for remote robots," *IEEE Trans. Robot. Automat.*, vol. 11, pp. 477-490, Aug. 1995.

[2] R. B. McGhee and G. I. Iswandhi, "Adaptive locomotion of a multilegged robot over rough terrain," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, pp. 176–182, Apr. 1979.

[3] S. Hirose, "A study of design and control of a quadruped walking vehicle," *Int. J. Robot. Res.*, vol. 3, pp. 113–133, Summer 1984.

[4] T. T. Lee, C. M. Liao, and T. K. Chen, "On the stability properties of hexapod tripod gait," *IEEE J. Robot. Automat.*, vol. 4, pp. 427–434, Aug. 1988.

[5] S. M. Song and B. S. Choi, "The optimally stable ranges of $2n$-legged wave gaits," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, pp. 888–902, July/Aug. 1990.

[6] Y.-J. Lee, "A study on crab gait control and path planning for a quadruped robot on uneven terrain," Ph.D. dissertation, Dept. Electrical Eng., Korea Adv. Inst. Sci. Technol., Seoul, 1994 (in Korean).

[7] J.-M. Yang and J.-H. Kim, "Fault tolerant locomotion of the hexapod robot," *IEEE Trans. Syst., Man, Cybern. B*, vol. 28, pp. 109–116, Feb. 1998

[8] X. D. Qiu and S. M. Song, "A strategy of wave gait for a walking machine traversing a rough planar terrain," *ASME J. Mechan. Transmiss. Automat. Design*, vol. 111, no. 4, pp. 471–478, Dec. 1989.

Retinally Reconstructed Images: Digital Images Having a Resolution Match with the Human Eye

Turker Kuyel, Wilson Geisler, and Joydeep Ghosh

***Abstract*— Current digital image/video storage, transmission and display technologies use uniformly sampled images. On the other hand, the human retina has a nonuniform sampling density that decreases dramatically as the solid angle from the visual fixation axis increases. Therefore, there is sampling mismatch between the uniformly sampled digital images and the retina. This paper introduces retinally reconstructed images (RRI's), a novel representation of digital images, that enables a resolution match with the human retina. To create an RRI, the size of the input image, the viewing distance and the fixation point should be known. In the RRI coding phase, we compute the "retinal codes," which consist of the retinal sampling locations onto which the input image projects, together with the retinal outputs at these locations. In the decoding phase, we use the backprojection of the retinal codes onto the input image grid as B-Spline control coefficients, in order to construct a three-dimensional (3-D) B-spline surface with nonuniform resolution properties. An RRI is then created by mapping the B-spline surface onto a uniform grid, using triangulation. Transmitting or storing the "retinal codes" instead of the full resolution images enables up to two orders of magnitude data compression, depending on the resolution of the input image, the size of the input image and the viewing distance. The data reduction capability of retinal codes and RRI is promising for digital video storage and transmission applications. However, the computational burden can be substantial in the decoding phase.**

***Index Terms*—Compression, fovea, image coding, reconstruction.**

I. Introduction

The properties of the human visual system have enabled technologies for many applications. The 50 Hz temporal resolution

Manuscript received August 27, 1996; revised August 13, 1998. This work was supported by AFOSR Contract F49620-93-1-0307 and ARO Contract DAAH04-94-G0417. This work was presented in part at the the SPIE Conference in Human Vision and Electronic Imaging III, San Jose, CA, 1998.

T. Kuyel is with Texas Instruments Inc., Dallas, TX 75206 USA (e-mail: kuyel@ti.com).

W. Geisler is with the Department of Psychology, University of Texas, Austin, TX 78712 USA.

J. Ghosh is with the Department of Electrical and Computer Engineering, University of Texas, Austin, TX 78712 USA.

Publisher Item Identifier S 1083-4427(99)01458-7.

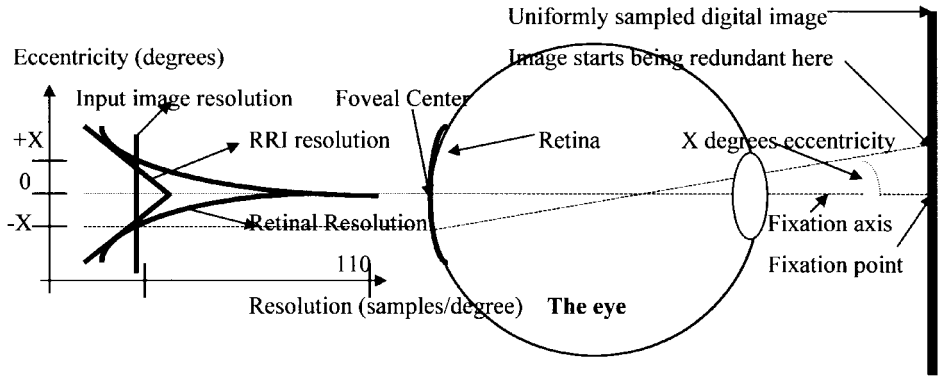


Fig. 1. The demonstration of the varying resolution across the retina and the possible redundancy of a uniformly sampled picture. The effects of the lens is ignored. A lot of anatomical facts are also ignored for simplicity.

of the human visual system has allowed the development of motion pictures and television, the trichromacy of the human visual system has allowed the development of color TV, and the spatial-frequency dependence of human contrast sensitivity has allowed spatial-frequency dependent video compression, as in the MPEG-1 and MPEG-2 standards. In this paper, we propose a way of exploiting the foveated nature of the human visual system for data compression.

Fig. 1 sketches the key idea behind the foveated retinally reconstructed images (RRI's) described in this paper. The spatial resolution of the human visual system near the point of fixation is approximately 60 cycles/$^\circ$, which is higher than typical image resolutions at typical viewing distances. However, beyond a certain eccentricity (angular deviation from the fixation axis), the resolution of the digital image exceeds the retinal resolution, and thus starts becoming redundant for the human observer.

An RRI is a uniformly sampled digital image which uses the fixation point and viewing distance information to create a better resolution match with the retina. Data compression obtained by using RRI's, instead of full resolution images, is lossy. However, this lossy compression becomes "perceptually lossless" if the RRI resolution exceeds sufficiently the resolutions of the human visual system at all eccentricities. Furthermore, allowing some "perceptual loss" in the image away from the point of fixation may not significantly affect performance in many visual tasks, because human observers often do not utilize the visual information in the peripheral regions of the visual field nearly as efficiently as the information near the point of fixation.

By using RRI's and fixation information, up to two orders of magnitude of "perceptually lossless" data compression can be obtained, depending on the size of the pixels, the size of the image, and the viewing distance. In general, as the size of the image increases (holding pixel size and viewing distance constant), the potential level of compression also increases because the sampling density in the peripheral retina is very low. Note also that the compression obtained by using RRI's can be multiplicatively combined with the conventional data compression methods for higher data compression rates.

For RRI based applications, the viewing distance and the fixation point information is needed. Therefore, RRI's are mainly useful for a single viewer setting, in which an eye tracking device is used. An example would be a person sitting in front of a high resolution monitor, who sees a sequence of RRI's corresponding to his sequence of fixations. Without an eye tracker, the use of RRI's is limited to situations where the human fixation behavior is highly predictable. For example, when viewing digital video, humans have very predictable fixation behavior for certain frame sequences, which would allow RRI sequences to be constructed according to the predicted fixation points.

In the literature, the nonuniform sampling properties of the human retina have been reported in detail by Curcio *et al*. [1], and by Curcio and Allen [2]. The nonuniform filtering properties of the primate retina can be found in Croner and Kaplan's recent work [3], whereas the optical properties of the human eye dates back to Campbell and Gubish's work [4] published in 1966. The applications of the nonuniform resolution properties of the retina is a relatively new area, and there is some recent work on enhancing image classification schemes using retina-like preprocessing. Kuyel, Geisler, and Ghosh used a retinal coder with an artificial classifier in order to explain human texture segmentation behavior [5]. In a later study [6], the same retinal coder was used for projecting sequentially increasing retinal resolutions on a target in order to increase the classification speed. There is also a substantial amount of recent work on using retinal coding for low bandwidth video. Kortum and Geisler implemented a real-time video conferencing system based on retinal coding [7]. The reconstruction was done by simply changing the pixel size. In a more recent study, Geisler and Perry developed a multiresolution foveated coder and improved the reconstruction algorithm using blending functions [8]. Basu and Wiebe also developed a low bandwidth videoconferencing system using retina-like coding [9]. In their work they showed how multiple fovea can be used for multiple center of attentions and they demonstrated how retinal coding can be combined with existing image compression algorithms such as JPEG. Experimental results on sequential retinal fixations on video degradation were shown by Duchowsky and McCormick [10]. Another interesting application of retinal coding is the direct VLSI implementation of the coder, as a retinal image sensor. Among examples are the work of Pardo and Martinuzzi [11] and the work of Wodnicky and Roberts [12].

A major problem of the retinal coding and decoding algorithms is known to be the aliasing artifacts which occur in the peripheral region after the decoding [7], [8]. In our work, a smoothly decaying resolution is used for retinal coding (as opposed to decreasing the resolution by integer steps at certain eccentricities) and b-spline reconstruction is used to control the smoothness of retinally reconstructed images. Increasing the order of b-splines have the effect of smoothing the aliasing at the periphery without blurring the fixation region significantly.

An important question that needs to be answered is how much perceptually lossless compression is attainable using retinal coding. In this study (Section IV), neurophysiological results and sampling theory is used to calculate how much retinal coding based compression is possible using the existing display technology. Results show

that retinal coding will be more effective as the display resolution increases. When using existing displays such as TV and computer screens, humans tend to adjust their distances to the display until they obtain a good retinal resolution match. Due to this reason, retinal coding technology can improve video compression only to a certain degree. However, the compression obtained from retinal coding is still comparable to compression obtained from traditional algorithms such as motion compensation.

II. Retinal Coding of Images

This section outlines the nature of human retinal resolution and demonstrates how uniformly sampled images can be coded to have a sampling match with the retina. Downsampling techniques that enable a "balanced image degradation" with respect to retinal resolution are then described.

A. Overview of Primate Retinal Coding

To provide a background for the retinal model, we briefly describe the optical and neural processing which occurs in the human/primate eye. As light passes through the eye, it is modified by the transfer function of the optics of the eye. This optical transfer function is nearly optimal, under daylight conditions, and has nearly constant characteristics within 10° of the line of sight. Under daylight conditions, cone photoreceptors sample the light that falls onto the retina. The density of the cone photoreceptors is the highest in a small retinal region called the "fovea," and declines quickly with increasing eccentricity (angular distance from the line of sight). The fixation point projects onto the center of the fovea, where the sampling density of the cones is highest. The cone aperture (i.e., the effective area of light collection) also become larger with increasing eccentricity, and hence accomplishes more "light averaging." For the first few degrees of eccentricity, it is estimated that each cone makes an excitatory connection to a single on-center/off-surround type midget ganglion cell via an on-bipolar cell, and an inhibitory connection to a single off-center/on-surround midget ganglion cell via an off-bipolar cell [2]. (Note that the ganglion cells are the output cells of the retina; their axons form the optical nerve, and that the midget ganglion cells carry the high spatial resolution information.) Therefore, for the first few degrees of eccentricity, we assume that a single cone is largely responsible for the excitatory center response of a midget ganglion cell. However, due to the optics of the eye the effective size of the center is larger than a single cone. The sampling density of the ganglion cells is also very high near the foveal center and declines quickly with eccentricity (see Fig. 2). There are approximately 5 million photoreceptors in the average human retina, but only about 1 million ganglion cells. Furthermore, one half of the ganglion cells (the on cells) sample exactly the same spatial locations as the other half (the off cells). In the fovea, the sampling density of the ganglion cells is well matched to photoreceptors; in the periphery, the sampling density of the ganglion cells falls well below that of the cones. Overall, the ganglion cells substantially undersample the photoreceptors. Some useful recent measurements of human retinal photoreceptor and ganglion cell topographies have been reported by Curcio [1] and Curcio and Allen [2].

Ganglion cell outputs are the retinal outputs and ganglion cell filtering properties for visual input primarily determine the filtering properties of the retina. The receptive field of a ganglion cell has been modeled by a center-surround, Difference of Gaussians (DOG) model (see [3]) where the peak of the center Gaussian is an order of magnitude stronger than the peak of the surround Gaussian and the ratio of the area under the surround region to the area under the center region is approximately 0.55. With increasing eccentricity, the receptive fields of ganglion cells increase and the peak filter gains decrease in such a way that the overall gain remains approximately constant.

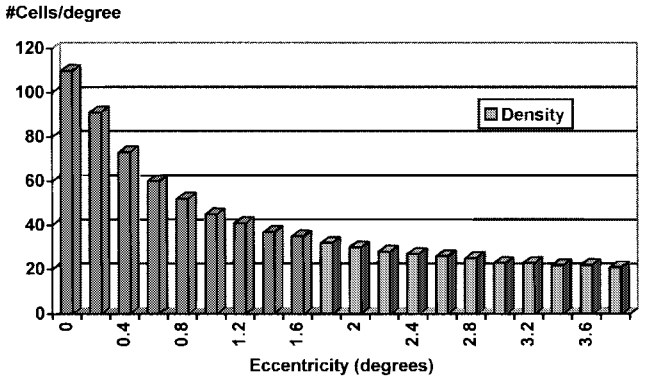


Fig. 2. Variation of the line density of human ganglion cells with eccentricity.

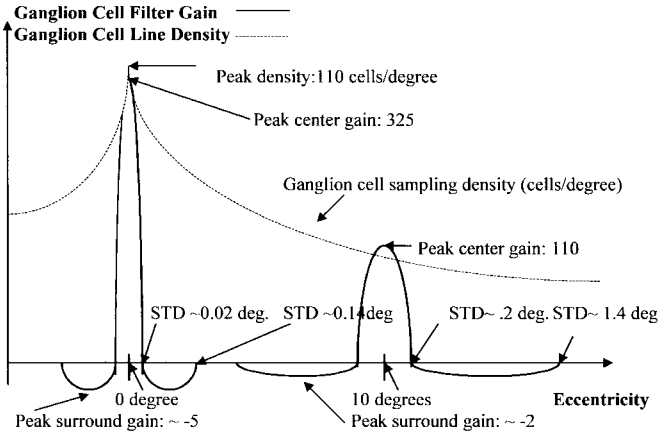
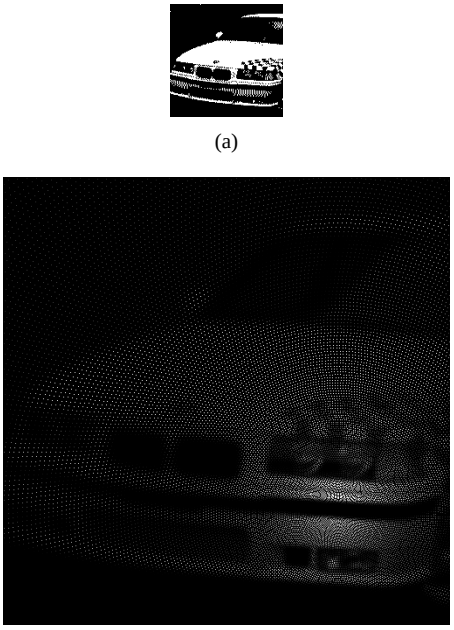


Fig. 3. A schematic of the data on human ganglion cell density and receptive field properties of the ganglion cell responses (not drawn to scale; see [3] for exact representations).

B. Retinal Coding Model

We have developed a model of the human retinal coding using the available data on primate retinal physiology and anatomy [1]–[3]. This model assumes circular symmetry around the fixation point. Given an input image, a fixation point and a viewing distance, the retinal coding model computes the locations of the ganglion cells onto which the image projects, as well as the outputs for these cells. We define the outputs and the locations of the ganglion cells as retinal codes. Note that ganglion cell locations can be computed from the fixation point and the viewing distance, and need not be stored explicitly. Once retinal codes are obtained, they can be backprojected and displayed on the original image grid, as shown in Fig. 4(b).

Let us give some details on how our retinal coding model is implemented. Curcio's density data is interpolated to determine the linear density of ganglion cells with varying eccentricity. This density data is inverted to find the distance between neighboring ganglion cells at each eccentricity. The first ganglion cell is placed at the fixation point. The ganglion cell spacing at the fixation point is used to determine the radius of the first ring of ganglion cells. Once this radius is determined, Curcio's data is used again to determine the intercellular spacing at this eccentricity. The ganglion cells are placed on this ring using the intercellular distance. This distance is also used



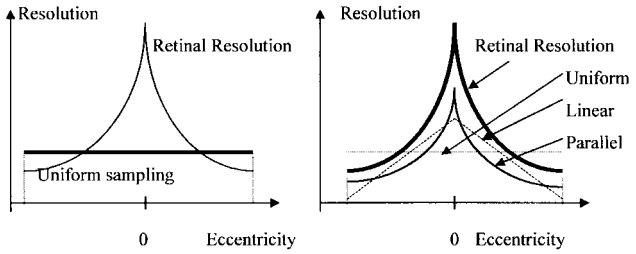
(a) (b)

Fig. 4. (a) A picture of a sports car, to be viewed at 25 cm viewing distance, (b) corresponding retinal codes (retinal outputs), backprojected on a 512 × 512 grid. Circular symmetry is assumed. Data from human and macaque retinal neurophysiology is used.

to determine how far the second ring will be from the first ring. As a rule, the intercellular spacing of kth ring is used as a radius increment to obtain the radius of the $(k+1)$th ring [see Fig. 4(b)].

The shift variant retinal filtering properties are determined with the help of Croner's and Curcio's data. The parametric form of the filter is DOG (see Fig. 3). The standard deviation of the center response at a certain eccentricity is assumed to be the same as the intercellular spacing at that specific eccentricity. The standard deviation of the surround response is determined to be seven times that of the center response [3]. The peak of the surround Gaussian is assumed to be 0.02 times that of the center Gaussian [3]. The receptive fields are computed for a single standard deviation of the surround Gaussian.

Equation (1) explains the nature of the nonuniform filtering involved in the computation of Fig. 4(b). Fig. 1 may be helpful in understanding the geometry involved in (1). Ganglion cell sampling lattice corresponding to the input image is formed in retinal coordinates using Curcio's data, and for each of these valid lattice points, the retinal output is computed using the DOG nature of the receptive fields. For a few degrees of eccentricity, a planar image can be considered to be a small patch of a sphere and spherical geometry can be used. The radius of the eye is neglected in comparison to the viewing distance. R is the viewing distance, ecc is the eccentricity, and the origin $(x=0,\ y=0)$ is considered to be the foveal center. $(R,\ ecc,\ \theta)$ triplet represents a unique retinal location as well as a unique point on the input image. $(x,\ y)$ is the projection of a valid ganglion cell location onto the image grid whereas $(x',\ y')$ is projection of any location within the ganglion cell receptive field. $\mathrm{Im}(x',\ y')$ is the value of the input image at the location $(x',\ y')$. $D(ecc)$ represents eccentricity dependent ganglion cell line density. $d(ecc)$ is the standard deviation of the center receptive field for a ganglion cell at eccentricity ecc, projected onto the input image grid. The standard deviation of the center receptive field is assumed to be equal to the intercellular spacing. $G(R,\ ecc,\ \theta)$ represents the output of a ganglion cell which is located at retinal location $(ecc,\ \theta)$. K is



(a) (b)

Fig. 5. (a) How a regular digital image downsamples the human retinal resolution. (b) Downsampling of retinal codes using a linearly decreasing resolution and a parallel decreasing resolution. The number of pixels (area under the resolution curve) is the same for the uniform, linear and parallel downsampling cases.

the normalizing factor for the DOG receptive field

$$G(R,\ ecc,\ \theta) = K \sum_{y'=y-7d(ecc)}^{y'=y+7d(ecc)} \sum_{x'=x-7d(ecc)}^{x'=x+7d(ecc)} \cdot \left(\exp\left[-\frac{(x-x')^2+(y-y')^2}{2d(ecc)^2} \right] - 0.02\ \exp\left[-\frac{(x-x')^2+(y-y')^2}{2(7d(ecc))^2} \right] \right) \cdot \mathrm{Im}(x',\ y') \tag{1}$$

$$x = R\ \tan(ecc)\cos(\theta),$$
$$y = R\ \tan(ecc)\sin(\theta),$$
$$d(ecc) \approx R\ \tan[1/D(ecc)].$$

Fig. 4 shows the outputs of our retinal model for a picture of a sports car. The source picture is a 512 × 512 uniformly sampled picture, corresponding to a 4 × 4 degree square. The size of the source image is 1.75 × 1.75 cm from 25 cm viewing distance [Fig. 4(a)]. We have printed an enlarged version of the backprojected retinal codes [Fig. 4(b)] on a 512 × 512 grid. Fig. 4(b) should be examined carefully because only the pixels that correspond to ganglion cell locations are defined. The brightness of a pixel that correspond to a ganglion cell represents the output of a ganglion cell, for the input image. The cumulative outputs for the ganglion cells represent the retinal outputs for this 4 × 4 degree picture of a sports car. One can notice that the fixation point is just below the driver side headlights. The sampling density drops away from the fixation point and the filtering properties become more low-pass. Undefined points on Fig. 4(b) are assigned a "black" color, and they are merely used for determining the location of actual sampling points. For this 4 × 4 degree image, there are approximately 47 000 retinal sampling points.

Since the retinal codes are backprojected onto a 512 × 512 uniform grid in Fig. 4(b), there are some mapping errors. Around the fixation point, the number of ganglion cells may exceed the number of grid points and mapping cannot be done properly. Aliasing effects occur due to mapping on a uniformly sampled grid. However, at high eccentricity, mapping becomes almost error free and circular rings of equally spaced ganglion cells can be observed. If viewed from a distance, the individual dots on Fig. 4(b) blur into greyscale regions, and the contrast of the picture reduces with eccentricity. This is a misleading observation. The error is due to assigning a constant color to the undefined points in the figure. Fig. 4(b) should be observed very carefully from a very close distance to compare the outputs of individual ganglion cells. Fig. 4(b) is not a retinally reconstructed

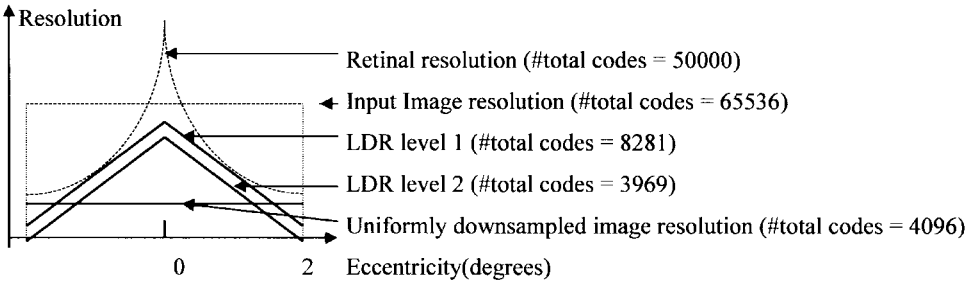


Fig. 6. Linearly decreasing resolution model for downsampling the retinal codes (not drawn to scale).

image. It is a series of retinal codes (retinal outputs) backprojected on a uniformly sampled grid.

The number of retinal codes obtained for a given image viewed at a given eccentricity is usually very high (47 000 for a 2° solid angle!). Speed requirements can limit the computation of a full retinal sampling lattice. A computationally more feasible technique is to down-sample the retinal lattice. It is beneficial to do downsampling in such a way that the resolution curve of the downsampled lattice is similar to the resolution curve of the retinal lattice. In doing so, a constant resolution loss at all eccentricities can be achieved. Other downsampling strategies, which give different resolution losses at different values of eccentricity are also possible. An example of this sort of downsampling is linear downsampling. Figs. 5 and 6 explain this downsampling process graphically.

III. Retinally Reconstructed Images (RRI's): Decoding of Retinal Codes on a Uniformly Sampled Grid

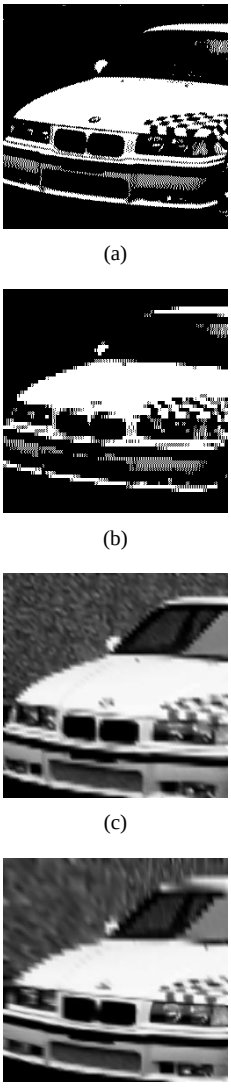
When backprojected on a uniformly sampled grid which has the same size with the image, the retinal codes represent an optimum sampling array which matches with the sampling lattice on the retina. However, only some of the image points are actually "defined." The problem of retinal decoding is to obtain the intensity values for every pixel of a uniformly sampled output image, given a set of retinal codes. It is known that an exact reconstruction from nonuniformly spaced samples is not possible [13]. In the case of nonuniform sampling, major signal processing tools like the Poisson's sum formula becomes invalid and convolution, the most important tool of linear shift invariant system theory, does not apply. In this work, we have made an approximation to the exact reconstruction by using three-dimensional (3-D) B-spline surfaces. B-Splines are known to be better approximators than truncated sinc functions [14], which also provides motivation for our reconstruction scheme.

For computational simplicity, we used a "linearly decreasing resolution" (LDR) model at two levels of resolutions. For a 2° solid angle, these resolutions used 3969 (63^2) and 8281 (91^2) sampling points in comparison to approximately 50 000 samples at retinal resolution.

For image reconstruction, we treat the retinal codes in 3-D (x, y, z) space. The x and y coordinates represent the backprojected location of a retinal code, whereas the z coordinate represents the intensity of that specific retinal code. These 3-D retinal codes are assigned to be the "control coefficients" of the B-Spline surface interpolation. The Cox De-Boor recursion [15] is used for evaluating the B-spline basis functions. The B-Spline interpolation evaluates points on a 3-D surface and the properties of this surface can be controlled by changing the "control coefficients," B-Spline order, and the knot vector. The smoothness of the B-spline surface is controlled by the B-spline order. We want the reconstructed surface to interpolate the end points of the image, therefore, we have used an "end point interpolating" knot vector. This type of knot vector is also called "uniform periodic." When the B-Spline reconstruction is performed, a series of interpolated points can be obtained. While these points can be made sufficiently dense by increasing the grid size, they are nonuniformly spaced. We solved this problem using linear interpolation (triangulation) in three dimensions.

Fig. 7(a) is the source image for retinal reconstructed imaging. It has 256 $\times$ 256 resolution and it represents a 4 $\times$ 4 degree image. Thus it should be viewed from approximately 50 cm. In Fig. 7(b), a uniform "zero order hold" reconstruction example is given. The 256 $\times$ 256 source image is uniformly sampled at every 4th point in x and y directions and reduced to a 64 $\times$ 64 grid. This 64 $\times$ 64 image is then mapped back onto a 256 $\times$ 256 grid using bigger pixels (4 $\times$ 4 constant intensity blocks). In Fig. 7(c), a second order B-spline reconstruction example is given. First, the 256 $\times$ 256 source image is retinally coded into 3969 B-spline coefficients which are spaced uniformly on the image grid. Then, these coefficients are used to reconstruct a 256 $\times$ 256 B-spline image. The resulting reconstruction is much superior to zero order hold reconstruction. Fig. 7(d) is an example of an RRI. Second order b-spline reconstruction is used to reconstruct the 256 $\times$ 256 image out of 3969 linearly downsampled retinal codes which are used as B-spline control coefficients. The fixation point is slightly below the driver side headlight. This retinally reconstructed image gives reasonably high resolution (for a 64 $\times$ 64 image), as long as the fixation point is on the driver side headlight.

Fig. 8 is similar to Fig. 7 except that a higher resolution is used in the linear downsampling of retinal codes (8281 control coefficients) for the same car image. Fig. 8(a) is the source image having 256 $\times$ 256 resolution. Fig. 8(b) is a reconstructed image from 8281 uniformly spaced control coefficients using second order B-Splines. Fig. 8(c) is a retinally reconstructed image. 8381 second order B-spline coefficients are nonuniformly distributed over the image to give more emphasis on the fixation point. The resolution on the driver side headlight increases at the expense of a resolution loss away from the headlight. Some aliasing effects are visible on the passenger side edges of the car. Fig. 8(d) is a Gaussian filtered version of Fig. 8(c) to remove the aliasing effects which occur away from the fixation point. The filter corrects for aliasing but it considerably blurs the high resolution region around the fixation point. Fig. 8(e) is a retinally reconstructed image using sixth-order B-splines. If we compare Fig. 8(e) to Fig. 8(c), we can observe that the aliasing effects are corrected and the fixation region is not considerably blurred. Fig. 8(e) is a clear improvement over Fig. 8(d). When sixth order B-Splines are used, the interpolation is done using five neighboring control coefficients. This results in a small amount of smoothing in the fixation region, because all the control coefficients are very close to each other. Away from the fixation point, the five neighboring control coefficients result in a stronger smoothing because they are spread over a large area. This is the reason why the aliasing effects in Fig. 8(c) can be corrected without considerably distorting the fixation region. Fig. 8(f) is a uniformly sampled B-spline reconstruction using



(a) (b) (c) (d)

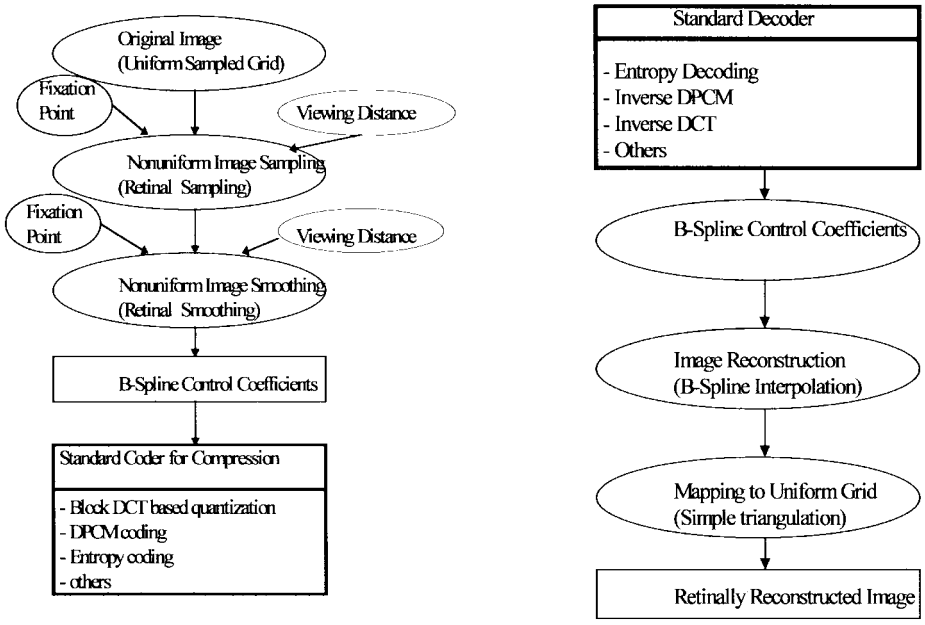
Fig. 7. (a) A 256 × 256 source image; (b) a 64 × 64 "zero order hold" reconstruction; (c) second-order B-spline reconstruction onto the 256 × 256 grid from 3969 uniformly spaced B-spline control coefficients; (d) retinally reconstructed image. Second order B-spline reconstruction is used to obtain the 256 × 256 image pixels from 3969 nonuniformly spaced control coefficients. The fixation point is on the driver side headlight. As long as the viewer keeps fixating on this headlight, this image will appear as if it has considerably high resolution. However, there is 16 times data reduction.

8281 control coefficients and sixth order B-splines. This figure is given for comparison with Fig. 8(b), to see the effects of changing the B-Spline order on reconstruction using uniform sampling.

The computational complexity of B-Spline based image reconstruction is substantial, but can be performed in real time using a DSP chip. The number of multiplications required for the retinal coding of an $N \times N$ image is approximately $N \times N$, and for a 256 × 256 image, the retinal coding for digital video at 30 frames/s will approximately cost 2 MFLOPS/s. On the other hand, the number of multiplications required for a uniform B-spline surface evaluation is $4(n-1)^2 S^2$ where n is the B-spline order, and S is the number of samples on the B-spline surface. For a 256 × 256 third-order B-spline surface, decoding of retinal codes onto RRI's at 30 frames/s costs approximately 30 MFLOPS/s. Triangulation is also computationally intensive, however, it can be replaced by a simpler mapping algorithm such as choosing the nearest B-spline surface point.

(a) (b) (c) (d) (e) (f)

Fig. 8. The image in Fig. 8(a) has 256 × 256 resolution the rest of the images have 91 × 91 resolution; (a) the source image, (b) second order B-Spline reconstruction from 8281 uniformly spaced control coefficients, (c) retinally reconstructed image. The fixation point is on the driver side headlight. Linearly decreasing resolution is used for coding the 8281 control coefficients. Second order B-Splines are used for reconstruction. d) A 7 × 7 Gaussian filter is used on Fig. 8(c) to get rid of the aliasing on passenger side edges of the car. However, when this is done, foveal region (driver side headlight) becomes distorted. (e) The order of B-spline reconstruction is raised from 2 to 6 for Fig. 8(c). The result is much better than Fig. 8(d). The blur in the foveal region is negligible, however, the aliasing effects on the passenger side edge of the hood is completely removed. (f) Sixth-order B-spline reconstruction using 8281 uniformly spaced B-Spline coefficients. Compare this figure to Fig. 8(b) to observe the effect of changing the B-Spline order on uniform reconstruction.



(a) (b)

Fig. 9. (a) Transmitter-end schematics of an image/video transmission scheme using retinal coding and (b) receiver-end schematics based on retinal codes and RRI's.

IV. Image/Video Compression Using Retinal Codes and RRI's

It is possible to take advantage of the nonuniform resolution properties of the human retina in digital video transmission and storage. Uniformly sampled images can be retinally coded as described in Section II. The retinal codes can be a downsampled version of the real retinal codes to suit the computational capabilities of the encoder. The key issue in retinal coding is to use a varying resolution pattern which is similar to that of the retina. The retinal codes which are encoded from the source image can be used in transmission or storage. At the receiver, retinally reconstructed images can be computed using the received retinal codes. One important aspect of this foveated image encoding/decoding scheme is the need for the fixation information and the viewing distance. The transmitter needs to know what the viewing distance is and where the fixation point is, in order to construct retinal codes around that fixation point. Therefore, the receiver needs to send the fixation information and the viewing distance information to the transmitter. This can be achieved through various ways. Using an eye tracker is currently an expensive option but the prices are dropping rapidly [16]. For digital video, a video clip can be "marked" in advance and high resolution can be assigned to spatio-temporal regions drawing the most attention. Foveated video applications are promising because at 30 frames/s, the eye does not have time to look at arbitrary regions of a single frame. Fig. 9(a) and (b) represent the flow diagrams of a transmitter and receiver using retinal encoding and retinally reconstructed images.

In the far periphery, the line density of retinal sampling points drop to almost one twentieth of the foveal density. Therefore, if a 40–50° field of view occurs, it is possible to obtain two orders of magnitude data compression using retinal coding. However, this high compression level is very unrealistic because current display technology uses a much smaller field of view and a lower resolution than the maximum retinal resolution. A TV screen viewed from a normal viewing distance fits in a few degrees of eccentricity. A computer video is usually played in a relatively small window which would also fit in a few degrees of eccentricity. Perhaps the only exception to the small-field-of-view display technology is the relatively uncommon and very expensive IMAX (maximum image) movie technology where the aim is to display in a wide field of view, so that the viewer can choose multiple regions to look at, as if he or she is in a real environment. Because the current display technology is mainly limited to TV screens and computer monitors, we will actually compute the maximum possible perceptually lossless compression that can be obtained by retinal coding for these display systems. To determine the maximum possible compression level, it is very important to know at which value of eccentricity the retinal resolution drops below the image resolution. In Fig. 1, this value of eccentricity is labeled "X°."

Let us consider a specific example of displaying images on a 14 in 1024 × 768 SVGA computer monitor at 50 cm viewing distance, and compare the total number of sampling points for the human retina and for the monitor. Our retinal model estimates that, within a circular region of 0.5° radius, there are approximately 4900 ganglion cells. This is equivalent to a 70 × 70 lattice, if the total number of sampling points is considered. This half a degree radius corresponds to 0.43 cm from 50 cm reading distance. A 14 in SVGA 1024 × 768 computer monitor has approximately 753 pixels within this radius, which is equivalent to a 28 × 28 lattice. The total number of retinal sampling points clearly exceeds that of the monitor within the first 0.5° of eccentricity. Within a circular region of 2° radius, there are approximately 47 000 ganglion cells, which is equivalent to a 216 × 216 lattice. On the SVGA monitor, this region with a 2° radius has approximately a 112 × 112 sampling points. Only after 4° of eccentricity, the total number of sampling points of the retina drops below that of the computer monitor.

For the same SVGA monitor example, let us now compute the eccentricity at which the sampling rates of the retina and the monitor are equal (X° in Fig. 1). The spacing between the pixels of the monitor is 0.032° from 50 cm viewing distance and our retinal model predicts that the ganglion cell spacing drops to this value at approximately 2.01° of eccentricity. This approximately corresponds to a 3.5 cm × 3.5 cm image on the monitor, which has 112 × 112 pixels. We can say that the retina oversamples a circle of 2° radius

on this monitor. At 50 cm viewing distance, within the first 56 pixel radius of the fixation point, the retinal sampling is redundant. Since it is not possible to increase the monitor resolution, nothing can be done within this region in order to obtain a sampling match with the retina. At exactly the fifty-sixth pixel from the fixation point, resolution of the monitor and resolution of the retina are the same. After the first 56 pixels, the monitor sampling becomes redundant and "perceptually lossless" data compression becomes possible by accordingly decreasing the image resolution to match the retinal resolution. This gives some insight on how much "perceptually lossless" compression can be obtained using retinal coding, on a 14-in SVGA monitor viewed at 50 cm. Apparently, for monitor images smaller than 112 $\times$ 112 "perceptually lossless" compression is not possible because the retinal resolution exceeds the full resolution of the monitor. For 256 $\times$ 256 images, approximately two to three times perceptually lossless compression is possible. For an MPEG-1 video frame in standard input format (352 $\times$ 240), three to four times compression can be achieved. If the full monitor images are considered (1024 $\times$ 768), the compression level can reach an order of magnitude. This compression level will further increase if a high resolution monitor is used (such as a 17-in monitor with a 1600 $\times$ 1200 pixel resolution).

There are existing digital satellite reception systems (DSS and DirectTV, etc.) that perform real time decoding of digital MPEG video to analog NTSC systems. Therefore, it is of practical importance to investigate how much foveated perceptually lossless compression can be obtained using a TV display in a single viewer setting. Let us consider a 28-in TV set using NTSC system (453 $\times$ 340 effective pixels), viewed at 5 m distance. The maximum viewing eccentricity is approximately 4° when the fixation is at the center of the screen. Our retinal model predicts that the region within a 44 pixel radius (corresponds to 0.8° of eccentricity) of the fixation point is oversampled by the retina. Also according to our retinal model, the foveated perceptually lossless compression will be approximately 3.2 times.

These compression levels made possible by retinal coding are promising for digital video applications because they are comparable to the practical compression levels obtained by stand-alone algorithms such as motion compensation. Moreover, for emerging display technologies such as "paper quality displays" or HDTV, the foveated compression levels can be much higher because such displays use much higher resolutions than current displays.

V. Conclusion

In this paper, the nonuniform resolution properties of the human retina have been used to determine coding and decoding strategies for data compression. Given an image, a point of visual fixation and a viewing distance, we have determined the number of retinal sampling points allocated to view that particular image. We have also determined how the sampling points are distributed on the retina and what their filtering properties are. Data from primate retinal neurophysiology is used in our computations. The computation of the position and the intensity of the full set of retinal sampling points (retinal codes) can be intensive, therefore, we have suggested balanced ways of subsampling the retinal codes. We have also developed a B-Spline based method to obtain retinally reconstructed images (RRI's) from "retinal codes." RRI's are actual images, having a uniform sampling grid and a monotonically decreasing resolution with increasing eccentricity. We have demonstrated how retinal codes and RRI's can be used for data compression for image/video transmission and storage. We have also demonstrated under which conditions "perceptually lossless" image compression is possible using retinal coding and current digital display technology. The data compression ratios (2–10 times compression) obtained by stand-alone retinal coding are already promising for current display technologies, and these ratios will significantly improve for higher resolution future display technologies such as HDTV or "paper quality displays."

In RRI based compression, the need for prior knowledge on the fixation point and the viewing distance can be satisfied by an eye tracking device. Depending on the technology involved, eye tracking can be extremely expensive, and may require wearing special purpose lenses, eye electrodes, fixing the head position, etc. However, there are also relatively inexpensive and more comfortable eye trackers which are based on pattern recognition using infrared CCD camera input. These devices are transparent to the user. However, their temporal and spatial resolutions are limited by the frame rate and the spatial resolution of the CCD camera. Spatial accuracies in the order of one tenth of degree, and temporal accuracies in the order of 20 ms have been reported [17], [18]. One tenth of a degree eye tracking error has virtually no effect on RRI's. Regular saccadic movements of the human eye is no faster than every 150 ms [18], [19], therefore, the frame rate of the eye tracker has enough temporal resolution to compensate for saccades. However, the calibration and recognition algorithms of the eye tracker have to work in real time with respect to the frame rate.

Even though eye tracking is an important part of RRI based compression, there are situations where it may not be essential. It is known that, for still images, humans fixate longer on image regions with the most information content. The information content of an RRI region depends on the information content of the source image and the RRI resolution at that region. If the high resolution foveal region of an RRI coincides with a high information region of the source image, the human eye will spend more time on fixating at this region. For a sequence of such RRI's, the human eye is likely to follow the foveal regions automatically. This claim is also justified by the human tendency to fixate on the high interest spatio-temporal regions of digital video [20]. If such regions are premarked and retinally coded as foveal regions, substantial compression can be obtained without using an eye tracking device. The human eye will automatically fixate at the sequence of high interest foveal regions, and at 30 frames/s rate, it will not have enough time to scan through the low-resolution-low-interest regions of any single video frame.

References

[1] C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson, "Human photoreceptor topography," *J. Comparative Neurol.*, vol. 292, pp. 497–523, 1990.

[2] C. A. Curcio and K. A. Allen, "Topography of Ganglion cells across human retina," *J. Comparative Neurol.*, vol. 300, pp. 5–25, 1990.

[3] J. L. Croner and E. Kaplan, "Receptive fields of P and M Ganglion cells across the primate retina," *Vision Res.*, vol. 35, no. 1, pp. 7–24, 1995.

[4] F. W. Campbell and R. W. Gubish, "Optical quality of the human eye," *J. Physiol.*, vol. 186, pp. 558–578, 1966.

[5] T. Kuyel, W. S. Geisler, and J. Ghosh, "A nonparametric statistical analysis of texture segmentation performance using a foveated image preprocessing similar to the human retina," in *Proc. IEEE SSIAI-96*, 1996, pp. 207–212.

[6] T. Kuyel and J. Ghosh, "Sequential resolution nearest neighbor classifier," in *IASTED Signal and Image Processing '97*, 1997, pp. 441–446.

[7] P. Kortum and W. S. Geisler, "Implementation of a foveated image coding system for image bandwidth reduction," in *Proc. SPIE Human Vision, Electronic Imaging*, 1996, vol. 2657.

[8] W. Geisler and J. Perry, "A real-time foveated multiresolution system for low-bandwidth video communication," in *Proc. SPIE*, 1998, vol. 3299, pp. 294–305.

[9] A. Basu and K. J. Wiebe, "Enhancing videoconferencing using spatially varying sensing," *IEEE Trans. Syst., Man, Cybern. A*, vol. 28, pp. 137–148, Mar. 1998.

[10] A. T. Duchowski and B. H. McCormick, "Gaze contingent video resolution degradation," in *Proc. SPIE*, 1998, vol. 3299, pp. 318–329.

[11] F. Pardo and E. Martinuzzi, "Hardware environment for a retinal CCD sensor," in *EU-MCM SMART Workshop*, Apr. 1994.
[12] R. Wodnicki, G. W. Roberts, and M. D. Levine, "A foveated image sensor in standard CMOS technology," Tech. Rep., 1995.
[13] F. Marvasti and M. Analoui, "Recovery of signals from nonuniform samples using iterative methods," *IEEE Trans., Acoust., Speech, Signal Processing*, vol. 39, 1991.
[14] S. Moni and R. L. Kashyap, "Multisplines, nonwavelet multiresolution and piecewise polynomials," in *Proc. SPIE*, 1995, vol. 2569, pp. 393–404.
[15] C. deBoor, "On calculation with B-Splines," *J. Approx. Theory*, vol. 6, pp. 50–62, 1972.
[16] A. Joch, "What the eye teaches computers," *Byte Mag.*, pp. 99–100, July 1996.
[17] M. Bach, D. Bouis, and B. Fisher, "An accurate and linear occulometer," *J. Neurosci. Meth.*, vol. 9, pp. 9–14. 1983.
[18] S. Mannan, K. H. Raddock, and D. S. Wooding, "Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images," *Spatial Vision*, vol. 9, no. 3, pp. 363–386, 1995.
[19] H. Weber, "Presaccadic processes in the generation of pro and anti saccades in human subjects—A reaction time study," *Perception*, vol. 24, pp. 1265–1280, 1995.
[20] M. Tekalp, *Digital Video Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1995.
[21] W. S. Geisler and M. S. Banks, "Visual performance," *Handbook of Optics.* New York: McGraw-Hill, 1995.
[22] E. Cohen, "Algorithms for degree raising of splines," *ACM Trans. Graph.*, vol. 4, pp. 171–181, 1985.
[23] W. Tiller, "Rational B-Splines for curve and surface representation," *IEEE Comput., Graphics, Applicat.*, vol. 3, no. 6, pp. 61–69, 1983.
[24] D. F. Rogers, *Mathematical Elements for Computer Graphics.* New York: McGraw-Hill, 1990.
[25] C. deBoor, *A Practical Guide to Splines.* New York: Springer-Verlag, 1978.
[26] R. Navarro and J. Portilla, "Duality between foveatization and multiscale local spectrum estimation," in *Proc. SPIE*, 1998, vol. 3299, pp. 306–316.