https://doi.org/10.2352/issn.2169-2629.2020.28.53

Double Ring Model for Foveated Imaging

Hiroaki Kotera; Kotera Imaging Laboratory, Chiba, Japan

Abstract

Human visual system has a space-variant resolution nature. In the retinal receptive field, the resolution is not uniform but sampled finest in the central fovea and coarser in the peripheral. This variable resolution mapping function is born by the cerebral primary visual cortex V1. It has a clear visual field map of spatial information, and this spatial mapping structure is called Retinotopy. The forward mapping to visual cortex from retina is characterized with complex LPT (Log-Polar-Transform) by Schwartz. The retinal receptive field image is reconstructed by inverse projection LPT^{-1} from V1. This reconstructed process is called Foveated Imaging. Since the spatial information is concentrated in the center of the visual field, the Foveated Imaging is applied to image compression, pattern recognition, robot vision, and/or computer vision. The retinal receptive field image is suitable for material appearance expression with natural blurring due to peripheral vision.

However, the complexity of the inverse transform LPT^{-1} was a bottleneck. This paper proposes a Double-Ring-structured novel Foveated Imaging method using positive and negative Gaussian blur masks without using the inverse transform LPT^{-1} of Schwartz theory and reports the evaluation of reproduction errors.

Retina to/from Visual Cortex Mapping

Human vision has a high resolution in the fovea of the retina and a lower resolution as it goes to the periphery. This variable resolution mapping structure is called *Retinotopy*. This function is performed in the primary visual cortex V1. Of the cerebral visual cortex, V1 is the most studied. From the anatomical structure of the striate cortex composing V1, Talbot and Marshall [1] predicted that V1 has a clear map of spatial information. Schwartz [2] clarified that the retina to V1 mapping follows a space-variant Log-Polar Transformation (LPT).

The space-variant retinal receptive field image is reproduced by the inverse transformation LPT^{-1}. This regeneration process is called *Foveated Imaging*.

Since the *Foveated Imaging* condenses spatial information in the center of the visual field, it has led to the developments such as 1-chip retinal sensor, robot vision, wide-angle camera for vehicle, and is going ahead towards optical Active Vision by a combination with DMD (Digital Micromirror Device) [3]-[13].

Recently, it has also been applied to a glasses-type artificial reality display called Near-Eye Display [14].

Mapping from Retina to Visual Cortex (LPT)

The complex vector z pointing a pixel located at (x, y) in the retina is transformed to a new vector $log\,(z)$ by LPT as follows.

$$z = x + jy = \rho e^{j\theta};\ \rho = |z| = \sqrt{x^2 + y^2},\ \theta = tan^{-1}(y/x)$$
$$log(z) = u + jv = log(\rho) + j\theta;\ j = \sqrt{-1} \quad (1)$$

The retinal image is sampled at spatially-variant resolution on the polar coordinate (ρ, θ), that is, in the radial direction, fineest in the fovea but coarser towards peripheral according to the *logarithm of* ρ. While, in the angle direction, it's sampled at a constant pitch $\Delta\theta$, and mapped to the coordinate (u, v) of the striate cortex in V1.

In the discrete LPT system, (ρ, θ) is digitized to R number of rings and S number of sectors. The striate cortex image is stored in the new Cartesian coordinates (u, v) [15] [16] as given by

$$u \triangleq \rho(u) = \rho_0 a^u,\ v \triangleq \theta(v) = v\Delta\theta = (2\pi/S)v$$
$$where,\ a = exp\left[log(\rho_{max}/\rho_0)/R\right] \quad (2)$$

where, ρ_0 denotes the radius of blind spot and $\rho \geq \rho_0$ prevents for the points near origin not to be mapped to the negative infinite-point. This regulation is called CBS (Central Blind Spot) model [17].

***Figure* 1** illustrates how the image "sunflower" is sampled on an LPT grid placed virtually on the Retina and mapped to the coordinates (u, v) in the striate cortex of primary visual cortex V1.

The height $h(u)$ and width $w(u)$ of a unit cell between u+1 and u are given by the following equations [18].

$$h(u) = \Delta\rho = \rho(u) - \rho(u-1) = \rho_0(a-1)a^{n-1}$$
$$w(u) = 0.5\{\rho(u) + \rho(u-1)\} = (\pi/S)(1+a)a^{n-1}\rho_0 \quad (3)$$

The aspect ratio $\gamma(u)$ of each fan-shaped sector is kept constant as

$$\gamma(u) = w(u)/h(u) = \left(\frac{a+1}{a-1}\right)(\pi/S) \quad (4)$$

Hence the area $\alpha(u)$ of unit sector cell increases exponentially with u as given by

$$\alpha(u) = \pi\left\{\rho(u)^2 - \rho(u-1)^2\right\}/S = \pi\rho_0^2\left(a^2-1\right)a^{2u}S^{-1} \quad (5)$$

Here, the number of sectors S is given by

$$S = \left\lfloor \pi\left(\frac{a+1}{a-1}\right) \right\rfloor \quad \lfloor\ \rfloor \text{denotes } floor \text{ function} \quad (6)$$

Since the maximum radius ρ_{max} in Eq. (2) is given by the half of short side of rectangle image size, S is determined when ρ_0 and the ring number R is set and the base a is calculated.

Based on these parameters, if we design the sampling grid to satisfy $\gamma(u) \cong 1$, we can get *Foveated Imaging* with less distortion.

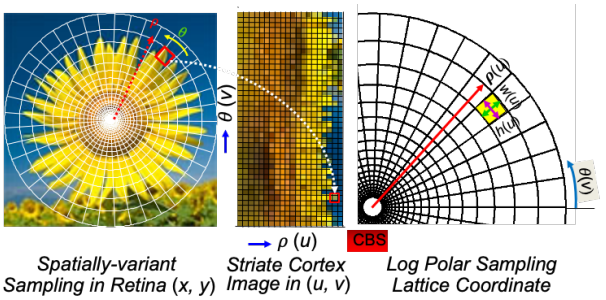


Figure 1. *"Sunflower" sampled in Retina and mapped to Striate Cortex*

Mapping from Visual Cortex to Retina (*LPT*$^{-1}$)

An optical image on the retina is sampled at variable resolution by LPT and mapped to the striate cortex in the cerebral visual cortexV1. Now, a key to *Foveated Imaging* is how to easily obtain the inverse transformation LPT^{-1}.

The relation between the coordinates (x, y) in retina and those of (u, v) in LPT is given for a view point (x_0, y_0) by

$$x(u,v) = \rho_0 a^u cos(2\pi v/S) + x_0$$
$$y(u,v) = \rho_0 a^u sin(2\pi v/S) + y_0 \quad (7)$$

The inverse mapped image quality after LPT^{-1} depends on the lattice design parameters (ρ, R, S). Here, the maximum radius ρ_{max} is determined from ρ and R by setting u=R in Eq. (3) as follows.

$$\rho_{max} = \rho(R) = \rho_0 a^R \quad (8)$$

In order that the outermost ring radius ρ_{max} does not protrude from the screen, it is necessary to put the viewpoint (x_0, y_0) inward by ρ_{max} from the four sides of the screen. Of course, the viewpoint (x_0, y_0) can be set freely.

To fit the forward mapped LPT image within the screen for the original image size (X, Y), conversely, ρ_{max} may be limited by the following equation.

$$\rho_{max} = Min\left[x_0, y_0, X - x_0, Y - y_0\right] \quad (9)$$

For example, in the case of x_0=X/2 and y_0=Y/2, the viewpoint is placed at the center of screen, then ρ_{max} is given by the half of short side of rectangle image.

***Figure* 2** outlines the process of transformation from the retina to the visual cortex V1 by LPT and the reverse transformation from the visual cortex V1 to the retina by LPT^{-1}.

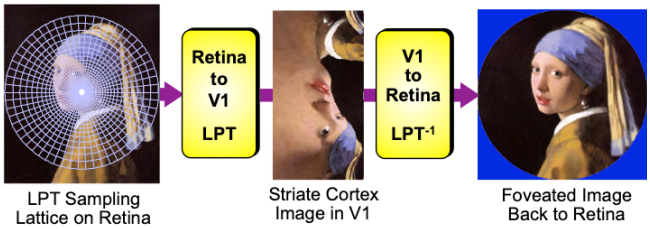


Figure 2. *Foveated imaging process Retina to/from visual cortex V1*

Conventional Method for Foveated Imaging

In practice, it is necessary to deal with blind spots and to suppress the mosaic artifacts in the inverse transform LPT^{-1}. So far, variety of foveated imaging methods have been developed, such as Super Pixel [19], Wavelet [20], Hermite [22], and Gaussian [23] [24]. Since it's difficult to express the space-variant blurring based on LPT^{-1} with a single operator, multiple filtering methods were introduced.

For example, in the Low-Pass method [21], a coefficient mask is prepared in advance to give a spatial blur distribution according to the pixel position, then continuous Gaussian blur is generated by a linear combination of locally blurred images with N sets of Gaussian functions using the coefficient in mask.

As another example, PSF method [24] prepares multiple point spread functions with different responses first. Next, a smooth Gaussian blur is realized as a set of convolution integrals with a Gaussian function with different σ depending on the pixel position. In practice, it's said that the local Gaussian blurring can be approximated by about 8 types of PSF and the efficiency has been improved significantly. These have the advantage that the spatial blur map can be freely set. Though, it's not intended to faithful reproduction of LPT^{-1}, but is aimed at visual display effects.

More recently, a high-speed method for obtaining a pseudo *F*oveated *I*maging has been proposed [25]. In this new method, an original image with $N_x \times N_y$ pixels sampled at regular intervals is resampled at the next logarithmic interval of

$$\Delta_x = 2{n_x}^{-1} log(N_x/2),\ \Delta_y = 2{n_y}^{-1} log(N_y/2) \quad (10)$$

Then, it is placed at the new coordinates $(x_k,\ y_k)$ as given by

$$x_k = exp(k\Delta_x)\ for\ k = 0, \cdots, n_x/2$$
$$y_k = exp(k\Delta_y)\ for\ k = 0, \cdots, n_y/2 \quad (11)$$

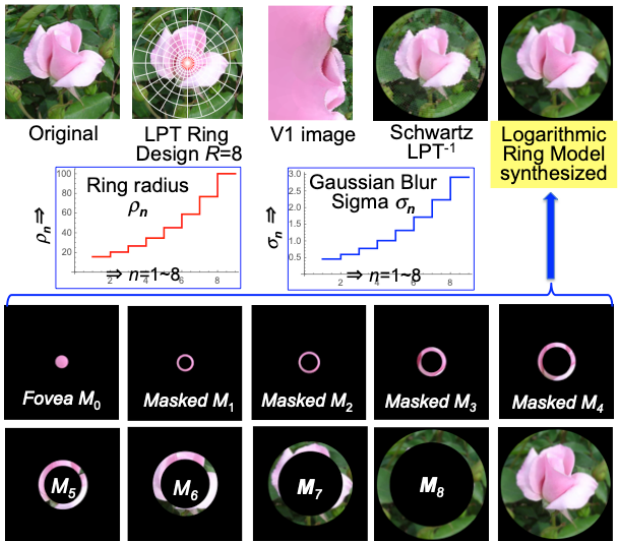
Where, the high speed operation is executed by a special hardware with 256 Pascal CUDA core GPUs. The foveated effect is realized with a simple method, and it is intended for application to movies. However, it has the limitation that the reproduced image size is reduced.

Logarithmic Ring Model

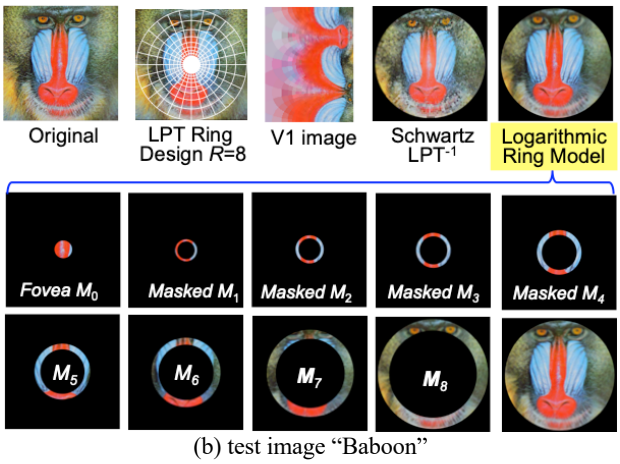
The inverse transformation LPT^{-1} can be approximated by a set of R rings whose width is given by the logarithm of ρ and increases radially outward from the center [18].

***Figure* 3** shows the image samples "Cherry shell" and "Baboon" reconstructed by the logarithmic ring model.

With the number of rings R=8, the radius ρ and the sigma σ of Gaussian blur increase stepwise, but no discontinuities are seen in the reproduced images. The ideal Schwartz theory could be approximated with high fidelity at ring number R=8.



(a) flower image "Cherry shell"



(b) test image "Baboon"

Figure 3. *Reconstructed samples by Logarithmic Ring model*

However, this model still has a drawback of high computation cost, because we must prepare Gaussian-blurred images with different σ by the ring number R in advance. Moreover, it requires a troublesome logical mask operations using a donut-shaped ring whose width expands exponentially.

Now, the reconstructed image by logarithmic Ring model with R=8 became the target for the new proposed model .

Double Ring Model for Foveated Imaging

As a more simple new method to approximate spatially-variant retinal image, *Double Ring* model with high resolution in the fovea and low resolution in the peripheral is proposed. This naming may be misleading. Mathematically, a **ring** in geometric meaning not set theoretic, is a set equipped with two binary operations of addition and multiplication. Here, we tried the ***Foveated Imaging*** using only ***two masks*** composed of the ***convex*** and ***concave*** 2-D Gaussian functions $\boldsymbol{G}_{Posi}$ and $\boldsymbol{G}_{Nega}$.

Since each mask works similar to a **ring** to extract the fovea and peripheral regions with binary-like operation, multiplication and addition after, as follows. We called this *Double Ring* model.

$$\boldsymbol{G}_{Posi}\left(x,y,\sigma\right) = \frac{1}{2\pi\sigma^2}\, exp\left\{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}\right\}$$
$$\boldsymbol{G}_{Nega}\left(x,y,\sigma\right) = \frac{1}{2\pi\sigma^2}\left[1-exp\left\{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}\right\}\right] \quad (10)$$

Here, for the original image of $X\times Y$ pixels, we cut out a $W\times W$ square area equal to the short side W=*Min* [X, Y] from $\boldsymbol{G}_{Posi}$ and $\boldsymbol{G}_{Nega}$. Each area is used as the control masks $\boldsymbol{M}_{Posi}$ and $\boldsymbol{M}_{Nega}$ respectively with the size adjusted by Eq. (11).

$$\boldsymbol{M}_{Posi}\left(x,y,\sigma\right) = Mask\left\{\boldsymbol{G}_{Posi}\left(x,y,\sigma\right), size \rightarrow W\right\}$$
$$\boldsymbol{M}_{Nega}\left(x,y,\sigma\right) = Mask\left\{\boldsymbol{G}_{Nega}\left(x,y,\sigma\right), size \rightarrow W\right\} \quad (11)$$

The high resolution fovea area is extracted by multiplying the original image $\boldsymbol{I}_{mage}$ by the convex control mask $\boldsymbol{M}_{Posi}$ as

$$\boldsymbol{C}_{enter}\left(x,y,\sigma\right) = \boldsymbol{M}_{Posi}\left(x,y,\sigma\right)\cdot\boldsymbol{I}_{mage}\left(x,y\right) \quad (12)$$

On the other hand, in the low-resolution peripheral area, the peripheral blur outside the fovea is replaced by the local average in the neighborhood radius r surrounding each pixel.

First, the blurred image $\boldsymbol{I}_{blur}$ with the local average filter is generated as

$$\boldsymbol{I}_{blur}\left(x,y,r\right) = MeanFilt\left\{\boldsymbol{I}_{mage}\left(x,y\right), area \rightarrow r\right\} \quad (13)$$

Next, multiplying $\boldsymbol{I}_{blur}$ by the concave control mask $\boldsymbol{M}_{Nega}$ that suppresses blurring in the fovea and leaves the peripheral as it is, the donut-shaped peripheral blurred image is extracted by the following equation.

$$\boldsymbol{S}_{urrnd}\left(x,y,\sigma,r\right) = \boldsymbol{M}_{Nega}\left(x,y,\sigma\right)\cdot\boldsymbol{I}_{blur}\left(x,y,r\right) \quad (14)$$

Lastly the objective foveated image is obtained by synthesizing $\boldsymbol{C}_{enter}$ and $\boldsymbol{S}_{urrnd}$ as follows.

$$\boldsymbol{F}_{oveated}\left(x,y,\sigma,r\right) = \boldsymbol{C}_{enter}\left(x,y,\sigma\right) + \boldsymbol{S}_{urrnd}\left(x,y,\sigma,r\right) \quad (15)$$

Now. changing the two parameters (σ, r), the approximation error between the target image by the logarithmic ring model (R=8) and the foveated image in Eq. (15) was evaluated and the 2-D error map was generated. Thus we searched for the bottom point in error map and found out the optimum values of (σ, r) to give the minimum approximation error.

Finally, the reconstruction error is evaluated by the average color difference ΔE_{ab}* in CIELAB space.

The above procedure is summarized in ***Figure*** **4**.

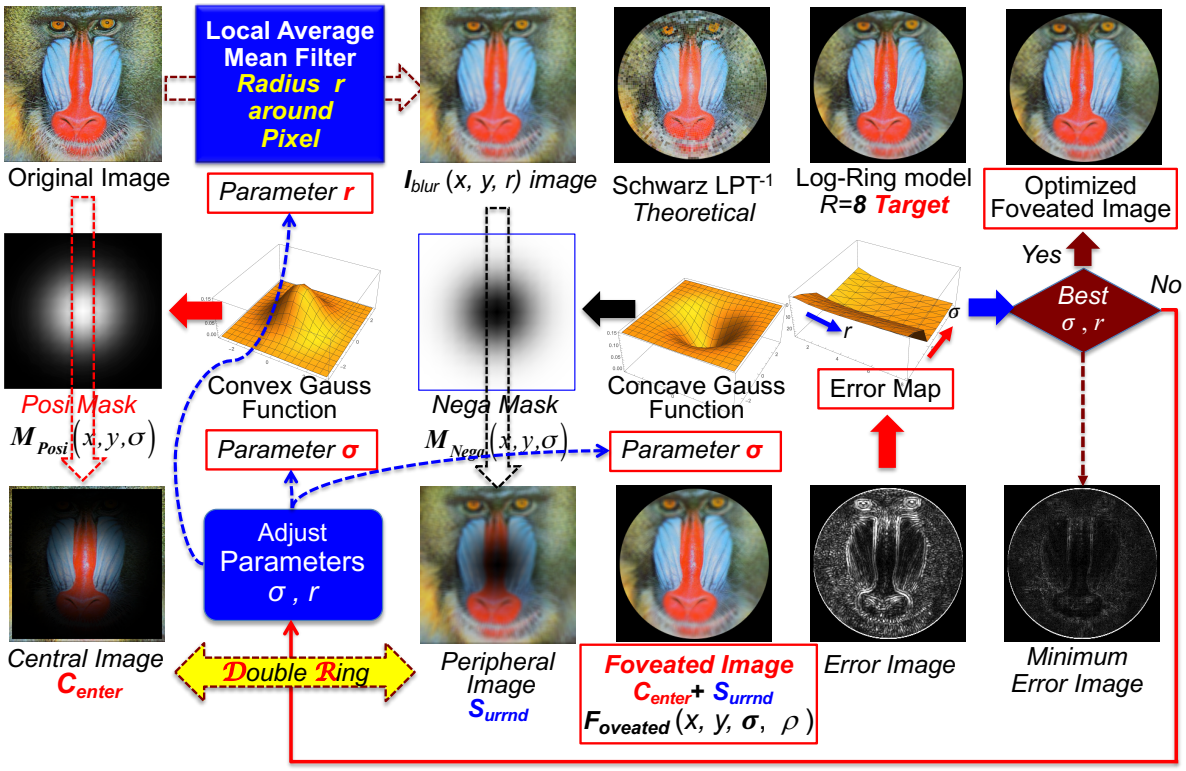


Figure 4. *Overview of Optimizing Procedure in the proposed Double-Ring Foveated Imaging model*

Experimental Results and Evaluation

Approximation Accuracy

***Figure* 5** shows optimal approximation results for typical four images. Although the error images are displayed with raw data, the difference between the target and the optimum approximated image is so small that cannot be visually discriminated. The minimum point in the error map differs slightly depending on the tested image, though, sufficient approximation accuracy can be ensured even if the average value of a representative test image is set as the default. Therefore, in practice, if the standard parameters (σ, r) are used as defaults, the minimization process may be omitted and the process speed can be much increased.

Comparison with other models

***Figure* 6** shows a comparison with the accurate & efficient model by T. Popkin et al [24], and another comparison with the LPT-based model by R. M. Narayaman et al [26]. Since both models work excellent, the error images are compared with raw data. Although the differences are so small, the proposed model is even better as clearly noted in the mean CIELAB ΔE^*_{ab}.

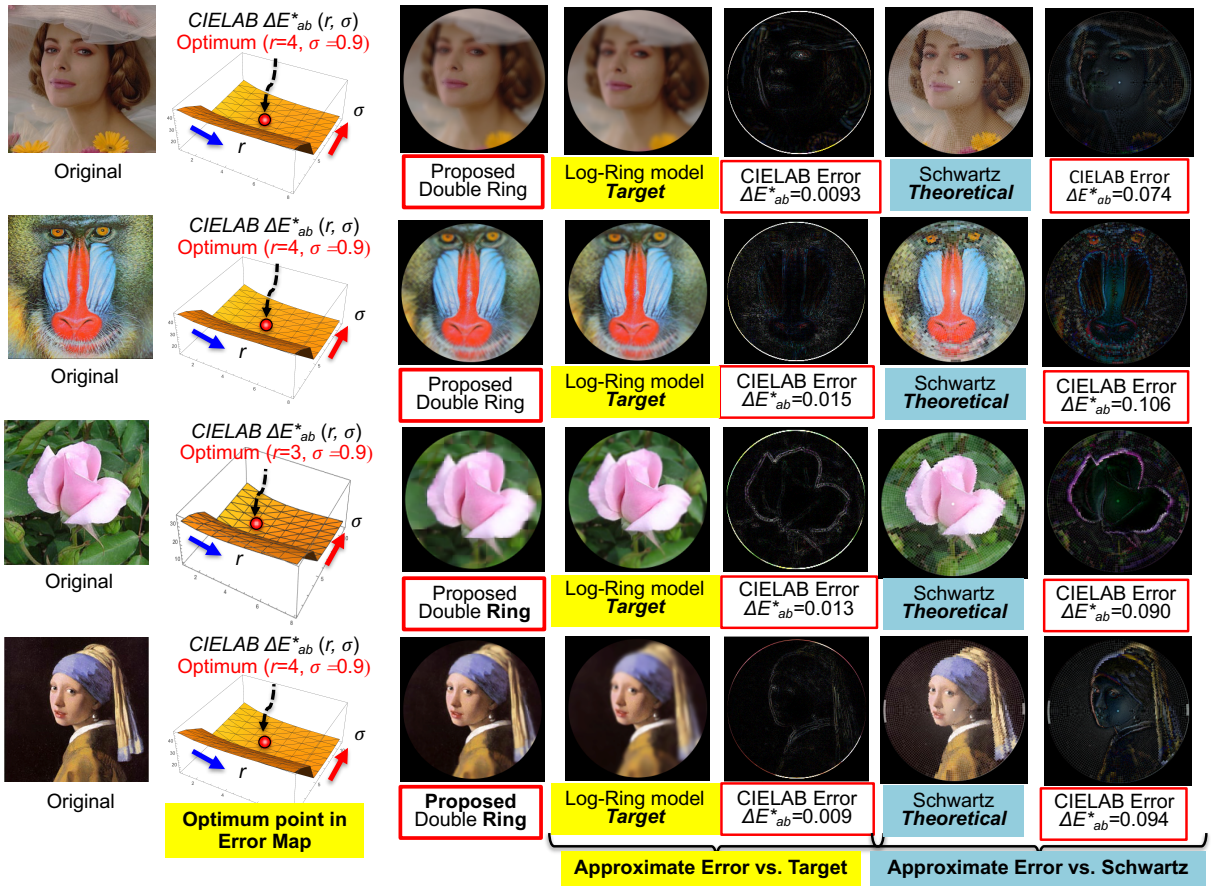


***Figure* 5.** *Foveated images back to Retina and approximation accuracy by the proposed model*

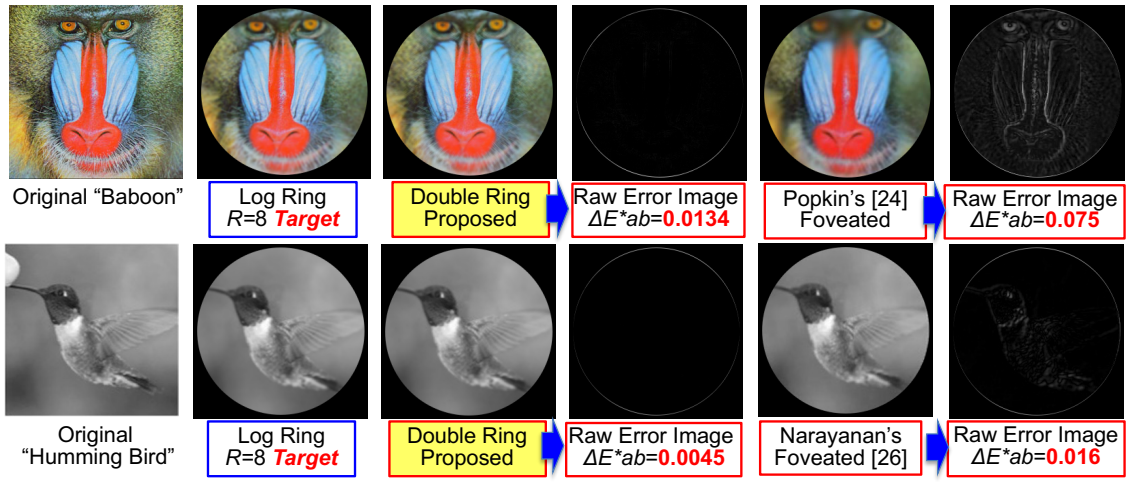


***Figure* 6** *A Comparison with accurate & efficient model by T. Popkin et al* [24] and by *R. M. Narayaman et al* [26]

Evaluation of Processing Speed

Finally, the processing speed is evaluated. This paper focuses on algorithm proposal. The proposed algorithm is programmed in a high-level language and executed in an interpreter method. We measured it as a measure of the relative simplicity of the model, not the gain or loss or efficiency. The proposed model takes time for iterative optimization, but sufficient accuracy can be obtained even if the default parameters are used, because the optimum values of (σ, r) are located around σ_{min}=0.9, r_{min}=3 ~ 4 as shown in ***Figure 5***. The error from the approximate target is 0.015 or less in CIELAB average color difference, and the high accuracy that cannot be visually discriminated is obtained.
On the other hand, since the fine mosaic remains in the Schwartz theory, the errors are slightly higher than those from the approximate target of logarithmic ring model. Still, since the CIELAB average color difference is around 0.1, similarly, it retains high accuracy that cannot be visually identified.

The processing speed is comparable to that of the logarithmic ring mask model, which was the approximate goal.

Table 1 Processing Speed (second /frame) : 256 X 256 Pixels

Method / Image	Schwartz Theoretical Model	Log-Ring (R=8, k=0.25) Default	Double Ring	
			(σ=0.9, r=4) Default	Including Optimization Iteration
Baboon	13.6	0.33	0.31	16.9
Pink Shell	10.7	0.29	0.24	12.1
Girl with pearl earrings	13.9	0.33	0.38	18.6

Conclusions

Many research results have already been reported on the *Foveated Imaging*. Applications are expanding from familiar surveillance cameras to intelligent robot vision, and real-time *Foveation Sensor Chips* are being developed.

The latest most notable application is in *Foveated Rendering*, which realizes high-end VR with ultra-low load. When viewing a real-time VR image on a head-mounted display (HMD), its peripheral visual field utilizes the visual characteristic that the image quality is not significantly degraded even if the drawing resolution is much reduced. Therefore, *Foveated Imaging* is regarded as an essential technique for the second-generation HMD VR. Since HMD VR is mainly applied to games, it works in conjunction with eye-tracking to display only a narrow area of the fovea with high resolution, and other areas with a fixed low resolution. Hence, the exact Foveation is not intended.

As well, in the proposed Double Ring *Foveated Imaging*, the peripheral visual field is replaced by a blurred image based on the neighborhood average of the radius r around the pixel. Roughly speaking, the processing looks similar, but the approach is fundamentally different in that the target image is set based on the Schwartz theory and optimal approximation is performed.

This paper proposed a simpler approach to approximate the spatially-variant sampling on retinal receptive fields without using the most troublesome inverse transform LPT^{-1}.

The parameters (σ, r) that give the minimum point of the error map are located in the vicinity of $\sigma \approx 0.9$ and $r \approx 3 \sim 5$ for the typical test images. Hence, sufficient approximation accuracy can be obtained even if (σ, r) are fixed to around these values as default. We would like to carry out additional verification for more different images.

The *Foveated Imaging*, which models the mapping from the retina to the visual cortex and its inverse, is closely related to C/S (Center/Surround) processing, which is the key to human visual perception for the outside world.

We hope that further elucidation of spatially-variant visual characteristics as well as the color appearance will lead to "material appearance", which is a major concern in recent years.

References

[1] S. A. Talbot and W. H. Marshall, "Physiological studies on neural mechanisms of visual localization and discrimination", American J. Ophthalmology, 24, 1255 (1941).

[2] E. L. Schwartz, "Spatial Mapping in the Primate Sensory Projection: Analytic Structure and Relevance to Perception", J. Biological Cybernetics., 25, 181 (1977).

[3] A.S. Rojer and E.L. Schwartz, Design considerations for a space-variant visual sensor with complex-logarithmic geometry, Proc. Intl Conf., PR, pg.278. (1990).

[4] C. W. Shin and S. Inokuchi, A New Retina-Like Visual Sensor Performing the Polar Transform, Proc. MVA, IAPR, pg.52. (1994).

[5] N. Oshiro et al, Binocular Tracking using Log Polar Mapping, Proc. IEEE Intl Conf., Intelligent Robots and Systems., 2, pg.791. (1996).

[6] D. Comaniciu et al, "Adaptive Resolution System for Distributed Surveillance", J. Real-Time Imaging, 8, 5, 427 (2002).

[7] G. Metra et al, "Learning to track colored objects with log-polar vision", J. Mechatronics, 14, 989 (2004).

[8] V. J. Traver and A Bernardino, "A review of log-polar imaging for visual perception in robotics", J. Robotics and Autonomous Systems, 58, 378 (2010)

[9] L. Itti, "Automatic Foveation for Video Compression Using a Neurobiological Model of Visual Attention", IEEE Trans. IP., 13, 10, 1304 (2004).

[10] S. Shimizu, "Wide-Angle Foveation for All-Purpose Use", IEEE/ASME Trans. Mechatronics, 13, 5, 587 (2008).

[11] Hashizume and Shimizu, "Development of wide-field sensor with fovea", ASTE A19 Annul Rept. RISE., Waseda Univ. (2011).

[12] L. Hardesty, "A faster single-pixel camera", MIT News Office, Mar.29. (2017).

[13] D. B. Phillips et al, "Adaptive foveated single-pixel imaging with dynamic super sampling". Science Advances Res. Article, 3, April. (2017).

[14] G. Tan et al," Foveated imaging for near-eye displays", J. Optics EXPRESS, 26, 19, 25076 (2018).

[15] V. J. Traver and F. Pla, Designing the lattice for log-polar images, Proc. 11th Intl Conf. Discrete Geometry for Computer Imagery, pg.164. (2003).

[16] V. J. Traver and F. Pla, "Dealing with 2D translation estimation in log-polar imagery", J. Image and Vision Computing, 21, 145 (2003).

[17] V. J. Traver and F. Pla, "Log-polar mapping template Design From task-level requirements to geometry parameters", J. Image and Vision Computing, 26, 1354 (2008).

[18] H. Kotera, "Visual Display Effect by Spatially-variant Resolution Transform", JJCSAJ, 37, 141 (2013) [In Japanese].

[19] R. S. Wallace, et al, "Space variant image processing", Intl. J. Computer Vision, 13, 71 (1994).

[20] E. C. Chang and C. Yap, A Wavelet Approach to Foveating Images, Proc. 13th ACM Symp. Computational Geometry, pg. 399. (1997).

[21] J. Portilla and R. Navvaro, Efficient Method for Space-Variant Low-Pass Filtering. Proc. VII Spanish Symp., Pattern Recognition and Image Analysis (1997).

[22] A. Krylov and D. Kortchaginc, Hermite Foveation, Proc. 14th Graphic Conference, pg.166. (2004).

[23] S. Tan et al, "Performance of three recursive algorithms for a fast space-variant Gaussian filtereing", J. Real-time Imaging, 9, 215 (2003).

[24] T. Popkin et al, "Accurate and Efficient Method for Smoothly Space-Variant Gaussian Blurring", IEEE Trans., IP, 19, 1362 (2010).

[25] U. Jaramillo-Avila and S. R. Anderson, Foveated image processing for faster object detection and recognition in embedded systems using deep convolutional neural networks. Proc. Biomimetic and Biohybrid Systems, Living Machines, pg.193. (2019).

[26] R. M. Narayanan et al, Considerations and Framework for Foveated Imaging Systems, Proc. 2018 Photonics, *5*, 18, pg.1. (2018).

Author Biography

Hiroaki Kotera *joined Panasonic in 1963. He received Ph. D from Univ. of Tokyo. After worked at Matsushita Res. Inst. Tokyo during 1973-1996, he was a professor at Chiba University. He retired in 2006 and is collaborating with Chiba University. He received 1993 IS&T honorable mention, 1995 SID Gutenberg prize, 2005 IEEE Chester Sall award, 2007 IS&T Raymond. C. Bowman award, 2009 SPSTJ and 2012 IIEEJ best paper awards. He is a Fellow of IS&T and IIEEJ.*