Image Extraction by Wide Angle Foveated Lens for Overt-Attention

Sota Shimizu   Hao Jiang   Joel W. Burdick

Robotics Laboratory, Dept. of Mechanical Engineering, Division of Engineering and Applied Science
California Institute of Technology
104-44, Pasadena, 91125, CA, USA

e-mail: {sato, jianghao}@caltech.edu, jwb@robotics.caltech.edu

Abstract – **This paper defines Wide Angle Foveated (WAF) imaging. A proposed model combines Cartesian coordinate system, a log-polar coordinate system, and a unique camera model composed of planar projection and spherical projection for all-purpose use of a single imaging device. The central field-of-view (FOV) and intermediate FOV are given translation-invariance and, rotation and scale-invariance for pattern recognition, respectively. Further, the peripheral FOV is more useful for camera's view direction control, because its image height is linear to an incident angle to the camera model's optical center point. Thus, this imaging model improves its usability especially when a camera is dynamically moved, that is, *overt-attention*. Moreover, simulation results of image extraction show advantages of the proposed model, in view of its magnification factor of the central FOV, accuracy of scale-invariance and flexibility to describe other WAF vision sensors.**

Index Terms – Wide Angle Foveated lens, Log-Polar Mapping, Overt-Attention, Advanced WAF imaging, rotation and scale-invariance

I. INTRODUCTION

The human retina-like image sensor, that is, the *fovea* vision sensor, is quite applicable for all-purpose use. Such a space-variant image sensor realizes not only to observe much wider field-of-view (FOV) with much smaller number of data [1], but also to observe the central FOV more in detail than other parts of FOV. Log-polar (LP) mapping is used as the most typical model for its image representation. This model is inspired by analytic formulation from biological observation of the primate visual system [2]. Sandini applied this model to computer vision computationally [3] and he and other researchers have advanced this theory to produce the LP vision chip with CCD or CMOS technologies [4]-[6]. The LP mapping is not only effective for drastic image data reduction (as the human retina does) but also suitable for generating geometrical rotation and scale-invariant feature simply, because of its mathematical property. The latter is the most indispensable advantage.

As another method to acquire the log-polar image, an optical approach is also well-known [7]-[9]. This approach usually combines the special-made Wide Angle Foveated (WAF) lens with a commercially available conventional Cartesian vision chip, where photosensitive element is size-invariant, although the LP chip approach combines the special-made chip with logarithm size-variant photosensitive elements with a conventional lens. It is one of the strongest advantages that the optical approach can realize a more complicated combination of different coordinate systems more easily than the special chip. Another of strong advantages is that the WAF lens can provide much higher resolution in the central FOV because of its optical magnification factor. This is due to an assumption that the maximal density of photosensitive elements is equivalent in both cases.

What is the image representation more suitable for fovea vision system in view of its functionality? Because camera's view direction control is quite essential for the fovea vision system, we should take account of *overt-attention*, that is, a type of attention when the camera is dynamically moved (another type is *covert-attention* that is, attention when the camera is statically fixed). It is well-known that a rotation, scale and translation-invariant property is quite applicable for pattern recognition [10]-[12] (Fourier-Mellin Transform is well-known as an algorithm to generate such a property). However, generally, a Cartesian image is not rotation and scale-invariant, and a LP image is not translation-invariant (that is, translation causes geometrical deformation of projection in the LP coordinates [4]). The *overt*-vision system with the fovea vision sensor can combine such 2 types of image for reliable pattern recognition, because precise camera view direction control to a target (using the Cartesian image) reduces the deformation (in the LP image). In addition, if the FOV is represented by the spherical projection, it is useful for the camera's view direction control. Thus, a combination of the Cartesian image, the LP image and spherical projection in addition to planar projection is so rational that it is applicable for many tasks of central vision and peripheral vision in all-purpose use of a single fovea vision sensor.

We often have to transform the LP image acquired by the space-variant fovea sensors, into the Cartesian image [4][9], because a method such as Fourier-Mellin Transform needs Cartesian coordinates in order to extract rotation, scale and translation-invariant feature. However, this does not mean that the Cartesian image is more suitable as a representation of an original input image, because the Cartesian image remapped conversely from the space-variant input image has higher resolution in its central FOV than that in the opposite case (from the Cartesian image to the LP image). Therefore, space-variant image acquisition and representation like a foveated image have also such an inherent merit.

This paper defines a model of the WAF imaging by a combining Cartesian coordinates, LP coordinates and a unique camera model in Chapter II, examines the proposed WAF model comparing with other WAF sensors in Chapter III, and concludes in Chapter IV.

II. IMAGE REPRESENTATION FOR OVERT

A. Modeling

Fig. 1 shows the appearance of the WAF lens, which can be attached with commercially available image device such as CCD camera. Its input image is shown in Fig. 2 comparing with the pinhole camera (PHC) lens. This PHC image has the same FOV, that is, the same view angle and the same number of pixels. The WAF lens has about 120 degrees wide FOV and adequate high resolution in the central FOV. This design concept is inspired much from functions of the human eye.
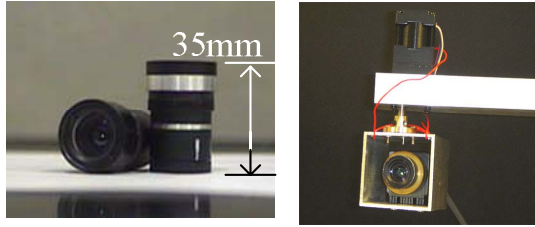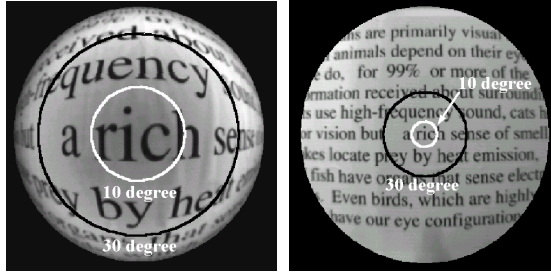

Fig. 1 WAF lens and its embedded view

(a)Input image by WAF lens   (b)The image by PHC lens
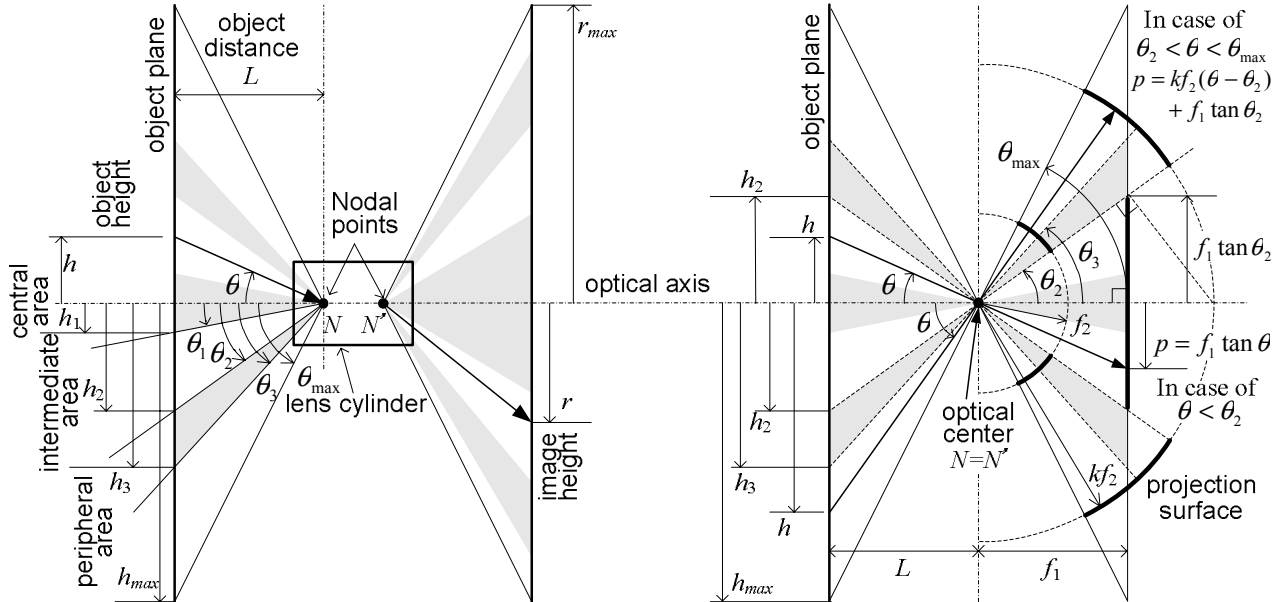Fig. 2 Images of WAF lens and pinhole camera lens

A correct LP image from the input image by the WAF lens is extracted based on camera calibration. Fig. 3 shows a geometrical sketch of the proposed camera model, which is based on a combination of planar projection (PP), that is, a perspective projection by tangent of incident angle, θ, to the lens optical center, and spherical projection (SP), that is, linear to θ. The projection height, p, of this camera model is defined as follows:

if the section of θ is $0 \leq \theta \leq \theta_2$,

$$p = f_1 \tan \theta, \tag{1}$$

else if $\theta_2 \leq \theta \leq \theta_{max} (=\pi/3 \text{ [rad]})$,

$$p = k \cdot f_2(\theta - \theta_2) + f_1 \tan \theta_2, \tag{2}$$

where f_1 and f_2 are focal lengths to the projection plane and the SP surface, respectively, and k is a correction factor for continuity of both projections,

$$k = \frac{f_1}{f_2 \cos^2 \theta_2}. \tag{3}$$

A model of the WAF image, namely Advanced WAF (AdWAF) imaging, is defined by the following equations, combining both PP by (1) and SP by (2) with both Cartesian and LP coordinates:

if $0 \leq \theta \leq \theta_1$,

$$r = r_{max} c_1 f_1 \tan \theta, \tag{4}$$

else if $\theta_1 \leq \theta \leq \theta_2$,

$$r = r_{max}(c_2 \log_a c_1 f_1 \tan \theta + d_1), \tag{5}$$

else if $\theta_2 \leq \theta \leq \theta_3$,

$$r = r_{max}\{c_3 \log_a(\frac{k \cdot f_2(\theta - \theta_2)}{c_1 f_1 \tan \theta_2} + 1) + d_2\}, \tag{6}$$

else if $\theta_3 \leq \theta \leq \theta_{max}$,

$$r = r_{max}\{c_4 f_2(\theta - \theta_3) + d_3\}, \tag{7}$$

Fig. 3 Camera model based on combination of planer projection and spherical projection

where r is the image height corresponding to θ, r_{max} is the maximum height when $\theta = \theta_{max}$, and c_i (i=1,2,3,4) is a scale modification factor for adjusting the height of partial images extracted from each section of θ. Further, d_i (i=1,2,3) is

$$d_1 = c_1 f_1 \tan\theta_1 - c_2 \log_a(c_1 f_1 \tan\theta_1) , \qquad (8)$$

$$d_2 = c_2 \log_a(c_1 f_1 \tan\theta_2) + d_1 , \qquad (9)$$

$$d_3 = c_3 \log_a\left(\frac{k \cdot f_2(\theta_3 - \theta_2)}{c_1 f_1 \tan\theta_2} + 1\right) + d_2 . \qquad (10)$$

Here, if $c_1 = c_2 = c_3 = c_4$, and r =1 when $\theta_{max} = \pi/3$, then

$$k = \frac{\theta_3 - \theta_2 + \cos\theta_2 \tan\theta_2}{\cos^2\theta_2 \tan\theta_1} . \qquad (11)$$

$$f_1 = bL , \qquad (12)$$

$$f_2 = \frac{f_1 \tan\theta_1}{\theta_3 - \theta_2 + \cos^2\theta_2 \tan\theta_2} , \qquad (13)$$

$$a = \exp(\frac{1}{f_1 \tan\theta_1}) , \qquad (14)$$

where L is the object distance, that is, a distance from the optical center to the object plane, and b is given by

$$b = \frac{1}{L \tan\theta_1} \bigg/ \left(\frac{\pi/3 - \theta_3}{\theta_3 - \theta_2 + \cos^2\theta_2 \tan\theta_2} \right.$$
$$\left. + \log\frac{\theta_3 - \theta_2 + \cos\theta_2 \tan\theta_2}{\cos^2\theta_2 \tan\theta_1} + 1 \right). \qquad (15)$$

In this case, not only (4)-(7) are continuous but also their derivatives are continuous.

Fig. 4 shows the image height r, the magnification factors dr/dh and r/h in the radial and tangential directions of the AdWAF image, respectively, in terms of the object height h, with those of other types of lens. The h_{max} and r_{max} are normalized to 1 (when $\theta_{max}=\pi/3$) to compare every type easily. The boundaries of FOV, h_1, h_2 and h_3 are 0.1 (=9.826 [°]), 0.4 (=19.107 [°]), and 0.6 (=34.715 [°]), respectively, in case of Fig.4. Other types of lens, that is, the LP lens, Fish eye (FE) lens, the PHC lens and the WAF lens are defined in the following.

LP lens:

if $0 \le \theta \le \theta_1$, which corresponds to $0 \le h \le h_1$,

$$r = r_{max} c_0 (h/h_1)/(1 - \log h_1) , \qquad (15)$$

else if $\theta_1 \le \theta \le \theta_2$, which corresponds to $h_1 \le h \le h_2$

$$r = r_{max} c_1 (1 - \log_{h_0} h) , \qquad (16)$$

where a basis, h_0, is decided as

$$h_0 = h_1 / e , \qquad (17)$$

such that (15) and (16) are continuous at $h=h_1$ when $c_0=c_1$, and their derivatives are also continuous.

FE lens:

$$r = \frac{3}{\pi} r_{max} c_0 \theta \qquad (0 \le \theta \le \theta_{max}) , \qquad (18)$$

PHC lens:

$$r = \frac{1}{\sqrt{3}} r_{max} c_0 \tan\theta \qquad (0 \le \theta \le \theta_{max}) , \qquad (19)$$

WAF lens:

$$r = r_{max}(a_0 \theta^3 + a_1 \theta^2 + a_2 \theta) \qquad (0 \le \theta \le \theta_{max}) . \qquad (20)$$

A bold solid line shows the WAF lens. The distribution of its image height and magnification factors (M.F.) is characterized by the design concept of the WAF lens, that is, acquiring much wider FOV and higher resolution locally in the central FOV. Its M.F in the radial direction is much higher than the PHC lens (a bold broken line) and the FE lens (a fine broken line) in small incident angles, on the other hand, lower than them in large incident angles. Comparing with the LP lens (a solid line), one notes that the proposed AdWAF model (a fine solid line with circle) can acquire higher M.F. in the central area ($0 \le h \le h_1$), by the same number of data, because of its lower M.F. in the intermediate and peripheral areas. The modification factor c_i is used for the M.F. of the AdWAF image in order not to exceed that of the WAF lens. Additionally, if r_{max} is 0.93 in the AdWAF model, the modificative M.F. is nearly equivalent to the M.F. of the LP lens in the central area, as in Fig. 4(b). Therefore, this means the AdWAF model can reduce more pixels in the peripheral FOV than the LP lens.

(a)image height r

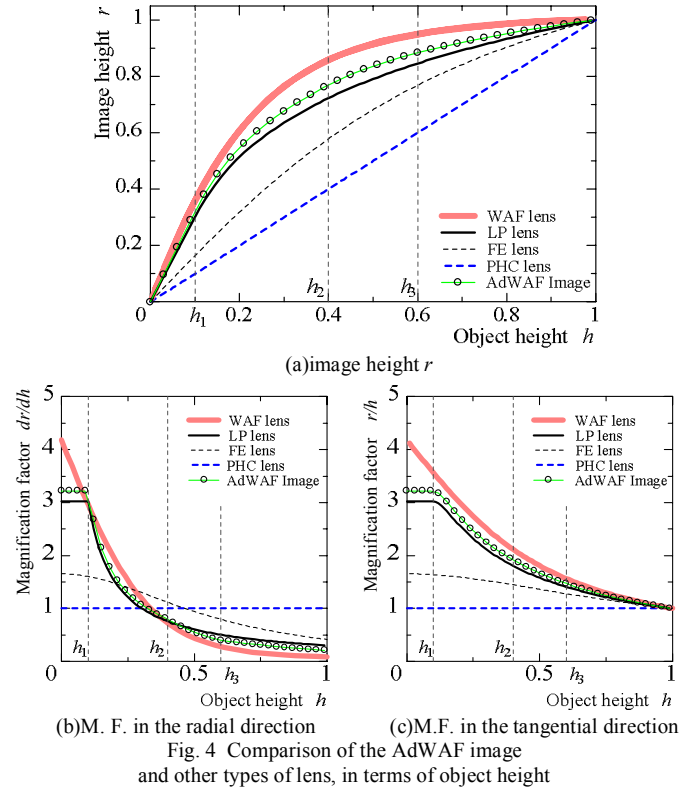(b)M. F. in the radial direction   (c)M.F. in the tangential direction

Fig. 4 Comparison of the AdWAF image and other types of lens, in terms of object height

The entire FOV is divided into three areas, that is, the central area ($0 \le \theta \le \theta_1$), the intermediate area ($\theta_1 \le \theta \le \theta_3$) and the peripheral area ($\theta_3 \le \theta \le \theta_{max}$) as shown in Fig. 3. The AdWAF model divides these three areas further into 4

areas by a boundary θ_2 between the PP and the SP. The PP and SP parts in the LP area are named PPLP area and SPLP area, respectively. The PPLP area should be used for central vision such as pattern recognition, and the SPLP area should be used more for peripheral vision. The central area (fovea) is a planar Cartesian area, because the image height r is linear to the object height h in this area. On the other hand, the peripheral area (periphery) is a spherical Cartesian area, because the r is linear to the incident angle θ.

B. Image Extraction

Fig. 5 simulates the AdWAF image comparing with other types of lens, by the whole view, in conditions of r_{max}=64[pixel], θ_{max}=π/3, h_{max}=1, h_1=0.1, h_2=0.4, and h_3=0.6. Each image is extracted from the original image (Fig. 5(a)) with 512 x 512[pixel^2]. One notes that the AdWAF image obviously has higher resolution in its central area than the LP lens. On the other hand, the resolution of its peripheral area is between the WAF lens and the LP lens.

(a)original image with 512 x 512[pixel^2]

(b)WAF lens image   (c)LP lens image   (d)FE lens image

(e)PHC lens image   (f)AdWAF image   (b)WAF lens image
Fig. 5 Simulation of WAF lens, LP lens, lens, PHC lens and AdWAF image

Fig. 6 shows the AdWAF image, actually extracted from the WAF lens in the same condition as the above simulation. Figs. 6(a), (b) and (c) are the extracted AdWAF image by the whole view, the PPLP image, that is, by (5), and the fovea image by (4), respectively. The rotation and scale-invariant property of the LP image is well-known [4]. On the other hand, the fovea image with Cartesian coordinates has only translation-invariant property.
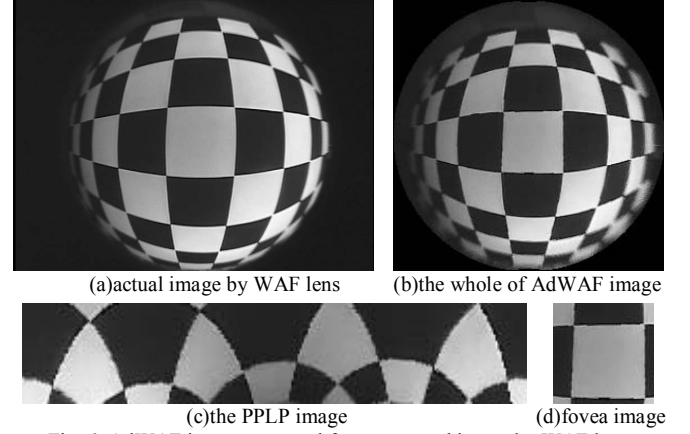

(a)actual image by WAF lens   (b)the whole of AdWAF image

(c)the PPLP image   (d)fovea image
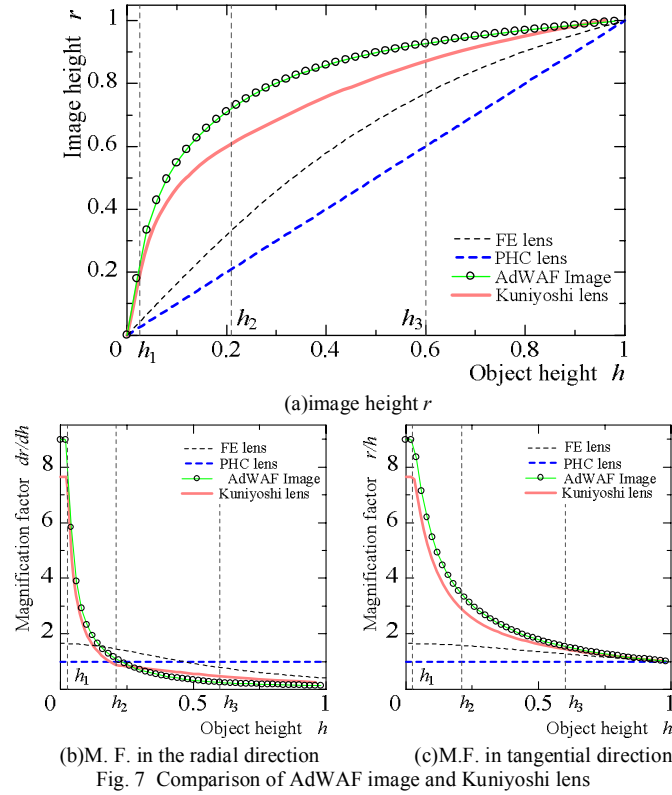Fig. 6 AdWAF image extracted from an actual image by WAF lens

III. EXAMINATION

A. Relation to Present LP Sensors

The AdWAF model defined by (1)-(15) can describe the model of Sandini's LP vision chip [4], by combining a wide-angle lens. For example, assume the lens is composed of the PHC lens' curve and the FE lens' curve. Both curves are connected continuously. When their boundary between both curves exists in this vision chip's LP area, Sandini's case is equivalent to the AdWAF model with θ_3= θ_{max}. Additionally, assume another case that the lens is a M.F.-variant special lens like the WAF lens. When the WAF lens' fovea part is projected by fitting the entire FOV of Sandini's vision chip, the AdWAF model makes it possible more smoothly because of both logarithmical descriptions of PP and SP. That is, the proposed model helps a practical camera calibration on such a unique combination and a design of another WAF senor.

Fig. 7 shows a comparison of the AdWAF model and Kuniyoshi's lens (K lens) [8], in conditions of r_{max}=1, θ_{max}=π/3, h_{max}=1, h_1=0.026, h_2=0.21, and h_3=0.6, which are determined by the distribution of M.F. of the actual K-lens. The values of h_1 and h_2, respectively, correspond to incident angles, θ_1 (=2.584 [°]) and θ_2 (=20.0[°]), from boundaries in the K lens' FOV. The FE lens, the PHC lens, the K lens and the AdWAF model are shown by a fine broken line, a bold broken line, a bold solid line, and a fine solid line with circle, respectively. The K lens changes the image height logarithmically of θ in $h_1 \leq h \leq h_2$, and linearly to θ in $h_2 \leq h \leq h_{max}$. That is, the AdWAF model includes the model of K lens, inherently. In the case when the first boundary h_1 (=θ_1) is equal with the second boundary h_2 (=θ_2) and the third boundary h_3 (=θ_3) is equal with the second boundary of the K lens model (given by the same symbol, θ_2, in [8]), they are

consistent (although Kuniyoshi assumes that two focal lengths, f_1 and f_2, to the PP and SP surfaces have the same value, exactly speaking). One notes that this condition of the boundaries gives the AdWAF model higher M.F. in its central area than the K lens, by the same number of data. On the other hand, if it has the same M.F. in its central area as the K lens, the r_{max} is 0.85. Therefore, the AdWAF model in this case can reduce the number of data by about 28 percent.

(a)image height r

(b)M. F. in the radial direction   (c)M.F. in tangential direction

Fig. 7 Comparison of AdWAF image and Kuniyoshi lens

B. Simulation Results

Fig. 8 shows test images with each different scale (α =0.75, 1 and 1.5). Figs. 8(d), (e) and (f) are the LP lens image, the AdWAF image, and the K lens image simulated by the distribution of M.F. of the actual K-lens. Both are extracted from Fig. 8(a) in conditions of $\theta_{max}=\pi/3$, $h_{max}=1$, h_1=0.026, h_2=0.21, and h_3=0.6. These two images bear out the property as mentioned in the previous section.

Both of the AdWAF model and the Kuniyoshi's model can acquire rotation and scale-invariant feature from the LP area. Figs. 9 and 10 show a comparison of the LP lens (left) and the AdWAF model (right), and a comparison of the LP lens (left) and the K lens (right), respectively, by their polar coordinate images extracted from test images in Figs. 8(a), (b) and (c). Both comparisons are simulated in the same conditions of $\theta_{max}=\pi/3$, $h_{max}=1$, h_1=0.026, h_2=0.21, and h_3=0.6. A bright part in each image shows the LP area composed of PPLP and SPLP. On the other hand, upper and lower dark parts, where brightness of each pixel is reduced half, show the planar Cartesian area (fovea) and the spherical Cartesian area (periphery). One notes that the AdWAF model acquires the wider fovea with translation-invariance than the LP lens, in

spite of being the periphery (the lower dark part), because the SPLP area reduces the image size of its corresponding part. In this condition, the AdWAF model acquires the wider fovea and a wider LP area than the K lens.
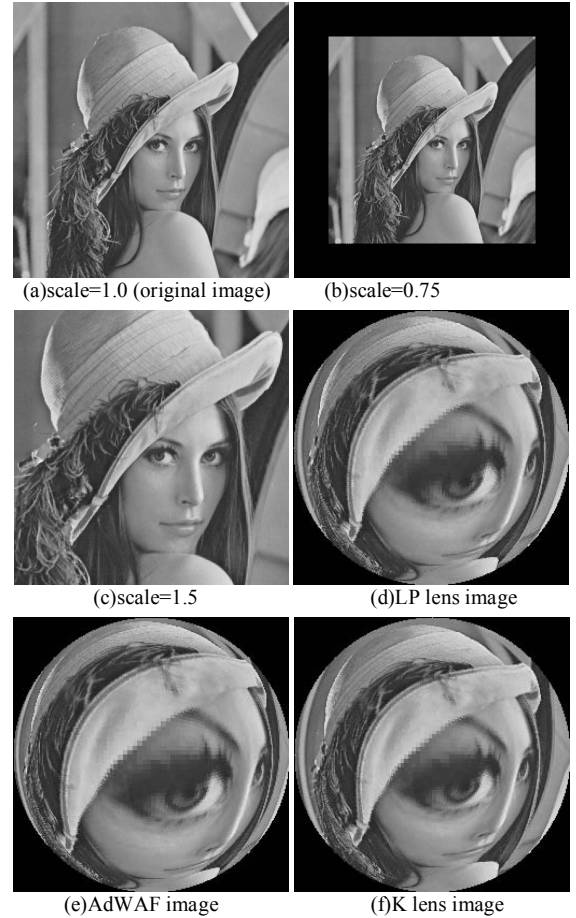

(a)scale=1.0 (original image)   (b)scale=0.75

(c)scale=1.5   (d)LP lens image

(e)AdWAF image   (f)K lens image

Fig. 8 Test images with different scale, LP lens image, AdWAF image, and K lens image by distribution of M.F. of the actual K lens

(a)when scale=0.75

(b)when scale=1.0

(c)when scale=1.5
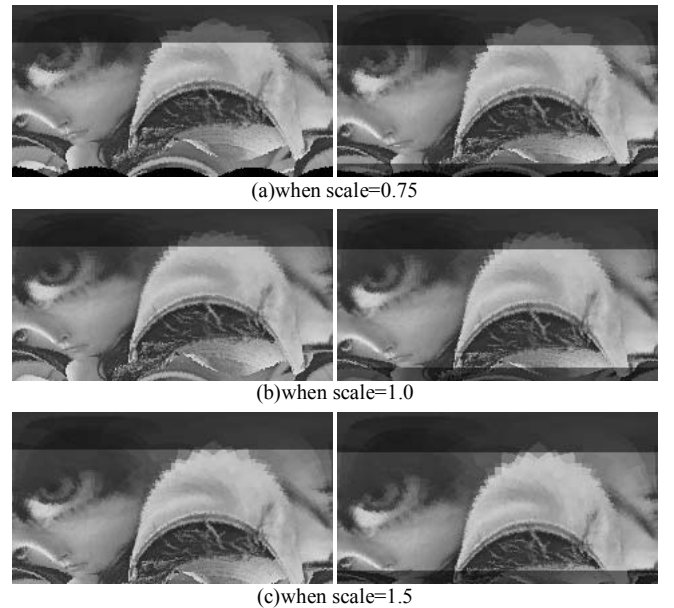
Fig. 9 LP images from the LP lens (left) and AdWAF image (right), when $\theta_{max}=\pi/3$, $h_{max}=1$, h_1=0.026, h_2=0.21, and h_3=0.6

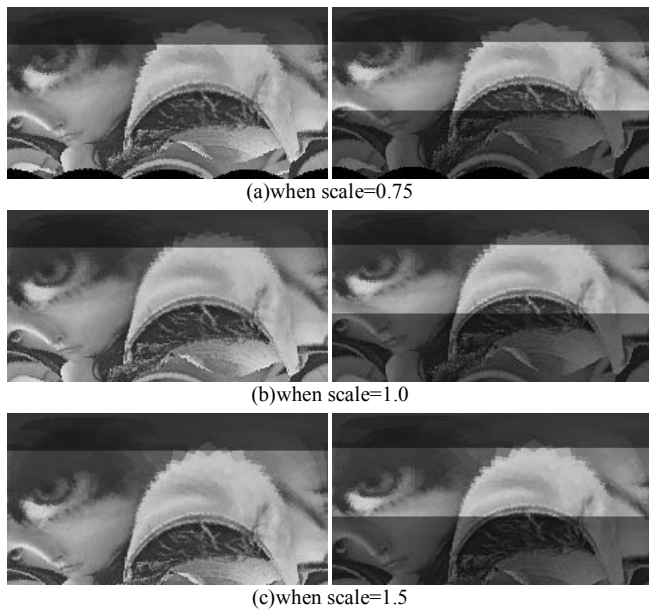(a)when scale=0.75

(b)when scale=1.0

(c)when scale=1.5

Fig. 10 LP images from the LP lens (left) and K lens image (right), when $\theta_{max}=\pi/3$, $h_{max}=1$, $h_1=0.026$, $h_2=0.21$, and $h_3=0.6$
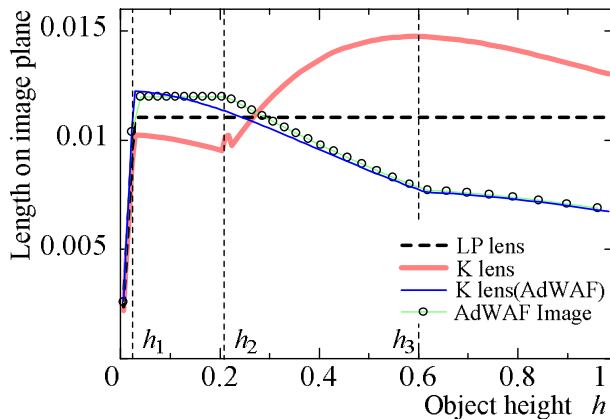

Fig. 11 Accuracy of scale-invariance by LP lens, K lens, and AdWAF model

IV. Conclusion

This paper has defined the AdWAF imaging as a more advanced model of the WAF image, that is, combining Cartesian and LP coordinates, the PP and SP, taking account of all-purpose use of the *overt*-vision system. The AdWAF imaging from the undistorted test image and the input image distorted by the WAF lens has actually been implemented. The simulation results has proved the following advantages of the AdWAF model: (a) the AdWAF model can acquire higher M.F. in the central area than the LP lens model, (b) it can acquire a more exact scale-invariant area in the PPLP area and can acquire a wider translation-invariant area in the planar Cartesian area, and (c) it can describe other WAF vision sensors more flexibly, because it is defined more exactly.

References

[1] S.Shimizu, "Multi-Functional Application of Wide Angle Foveated Vision Sensor in Mobile Robot Navigation", Journal of Robotics and Mechatronics, vol.14, No.4, pp.382-389, August 2002
[2] E.L.Schwartz, "Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception", Biological Cybernetics, Vol.29, pp.181-194, 1977
[3] G.Sandini and V.Tagliasco, "An anthropomorphic retina-like structure for scene analysis", Computer Graphics and Image Processing, 14, pp.365-372, 1980
[4] F.Berton, G.Sandini and G.Metta, "Anthropomorphic Visual Sensors, Encyclopedia of Sensors", Edited by C.A. Grimes, E.C.Dickey and M.V.Pishko, in press
[5] J.Van der Spiegel, G.Kreider, C.Claeys, I.Debusschere, G.Sandini, P.Dario and et al., "A foveated retina-like sensor using CCD technology", Analog VLSI Implementations of Neural Networks. Kluwer, C.Mead and M.Ismail, Boston, 1989
[6] R.Wodnicki, G.W.Roberts and M.D.Levine, "A foveated image sensor in standard CMOS technology", In Custom Integrated Circuits Conference, Santa Clara, California, May 1995
[7] S.Shimizu, et.al., "Vision Sensor with Wide angle and High Distortion lens", Video proceedings of IEEE International Conference on Robotics and Automation, Visual Sensing 3, May 1995
[8] Y.Kuniyoshi, N.Kita, K.Sugimoto, et.al., "A Foveated Wide Angle Lens for Active Vision", Proc. of IEEE International Conference on Robotics and Automation, pp.2982-2988, May 1995
[9] S.Shimizu, H.Jiang and J.Burdick, "Machine Vision System to Induct Binocular Wide-Angle Foveated Information into Both the Human and Computers -Feature Generation Algorithm based on DFT for Binocular Fixation-", Proceedings of IEEE International Conference on Robotics and Automation, pp.805-811, April 2005
[10] D.Casasent, and D.Psaltis, "Position, Rotation and Scale-invariant Optical Correlation", Applied Optics, Vol. 15, pp.1795-1799, 1976
[11] B.S.Reddy and B.N. Chatterji, "An FFT-based Technique for Translation, Rotation and Scale-Invariant Image Registration", IEEE Transactions on Image Processing, Vol.5, No.8, pp.1266-1271, 1996
[12] A.Baumberg, "Reliable Feature Matching Across Widely Separated Views", Proceedings of Computer Vision and Pattern Recognition, Vol.1, pp.774-781, 2000

Fig. 11 shows a length on the image plane to indicate accuracy of scale-invariance in terms of the object height, h. This length is a difference between an image height corresponding to each h and another image height corresponding to 95 percent of the h, where $r_{max}=1$. A broken line, a bold solid line and a fine solid line with circle show the LP lens, the K lens and the AdWAF model, respectively. Scale-invariance means that a gradient of each line is continuously zero. One notes that all are not scale-invariant in the planar Cartesian area ($0 \leq h \leq h_1$), consequently. In PPLP area, the K-lens is not scale-invariant, exactly. A local peak around $h=h_2$ of the K lens means the value of h_2 is not necessarily appropriate in view of recognizing patterns projected near the boundary h_2. A fine solid line (without anything) shows the simulated K lens drawn by the AdWAF model in a different condition of $\theta_{max}=\pi/3$, $h_{max}=1$, $h_1=0.026$, $h_2=0.026$, and $h_3=0.21$. This line changes more smoothly than the original K lens, although PPLP area is not scale-invariant. Thus, since the definition of the AdWAF model is more exact, it can describe other WAF vision sensors, more flexibly.