Proceedings of the 2005 IEEE
International Conference on Robotics and Automation
Barcelona, Spain, April 2005

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 -

Sota Shimizu and Shinsuke Shimojo
Psychophysics Laboratory, Division of Biology
California Institute of Technology
139-74, Pasadena, 91125, CA, USA
e-mail: sato@caltech.edu

Hao Jiang and Joel W. Burdick
Robotics Laboratory, Dept. of Mechanical Engineering
California Institute of Technology
104-44, Pasadena, 91125, CA, USA
e-mail: jianghao@caltech.edu

***Abstract* - This paper introduces a machine vision system, which is suitable for cooperative works between the human and computer. This system provides images inputted from a stereo camera head not only to the processor but also to the user's sight as binocular wide-angle foveated (WAF) information, thus it is applicable for Virtual Reality (VR) systems such as tele-existence or training experts. The stereo camera head plays a role to get required input images foveated by special wide-angle optics under camera view direction control and 3D head mount display (HMD) displays fused 3D images to the user. Moreover, an analog video signal processing device much inspired from a structure of the human visual system realizes a unique way to provide WAF information to plural processors and the user. Therefore, this developed vision system is also much expected to be applicable for the human brain and vision research, because the design concept is to mimic the human visual system. Further, an algorithm to generate features using Discrete Fourier Transform (DFT) for binocular fixation in order to provide well-fused 3D images to 3D HMD is proposed. This paper examines influences of applying this algorithm to space variant images such as WAF images, based on experimental results.**

Index Terms* - *fovea; stereo vision; 3D HMD; binocular fixation; feature generation; DFT; stereo matching; LGN device

I. INTRODUCTION

Many researchers have been interested in how the human being sees objects and how they look, using his or her own eyes for a long time. This kind of research has rather a long history due to biological interest, medical and industrial applications. Regarding how to see, we have often observed the human view directions measured from the iris positions by eye tracking devices, which use near infrared rays, and have observed a blink from EOG (Electro-Oculo-Gram). Further, we can know how much degree of focus an image is projected on the human retina, using an Auto-Kerato-Refractometer. On the other hand, regarding how the objects look, it has been more difficult. It has ever relied on electro-physiological ways to put electrodes based on a surgical operation, magnetic ways such as fMRI (functional Magnetic Resonance Imaging), MEG (Magneto-Encephalo-Graphy) and TMS (Transcranial Magnetic Stimulation), and electric ways such as EEG (Electro-Encephalo-Graphy). An optical way using near infrared rays such as OT (Optical Topography) attracts a great deal of attention recently [1]. These are results observed inside the human brain and visual system. Apart from the above ways, in order to get false images, artificial vision systems such as CCD cameras, which are fixed in front of the human face or on the top of head, are often utilized. These ways can be classified into two kinds of category. One is whether to extract internal information or external one, and the other is whether the extracted information is true or false. Therefore, the case of using an artificial vision system is located at external and false.

Although the images from CCD cameras are artificial and false, they have been more applicable. They are easier to handle, because a coding of images is originally known. The applications are not only limited within the biological investigations but also extended broadly to medicine and industry. However, it is obvious that robotics researches, which are related to artificial vision systems and attempt to build up artificial intelligence, aim also at investigating a mechanism of the human brain and visual system. The author think it is quite helpful for the brain and vision research to create an artificial vision system, which has many points in common with the human visual system. Further, if it can induct its visual information not only to processors but also to the human in the same way as the human does, such a system is likely to be able to replace a real human organ.

This paper introduces a machine vision system with an active stereo camera head and 3D head mount display (hereafter HMD) to get and provide binocular wide-angle foveated (hereafter WAF) information characterized by wide-angle special optics [2], in chapter II. Chapter III describes an algorithm to generate feature, which is used for detecting attention point and stereo matching to control stereo camera view directions in order to fixate at the same object and provide well-fused 3D images to 3D HMD. Moreover, in this chapter, experiments are carried out and the results are examined from a viewpoint of the influence when this algorithm is applied to WAF images, that is, distorted and space-variant images. Chapter IV summarizes this paper with future works.

II. WIDE ANGLE FOVEATED VISION SYSTEM

Many types of machine vision system [2]-[13] have been reported ever. They are applied broadly not only for automatic robot control but also for tele-existence, remote operating system, surveillance system and so on. Although the composition and functions are characterized by the purpose, each system has the image input device, which is usually composed of CCD camera, respectively. The input device of [3] has 10 cameras to get 3D information, Ref. [4] has a hyperbolic mirror in front of a single camera to get wide-view information, Ref. [6] has a high-speed camera to

*This work is partially supported by Wind & Biomass Energy R&D and Information Center to S.Shimizu

get a robot's own posture, Ref. [11] has stereo camera head with 6 DOF to get information efficiently, and Ref. [12] has zoom cameras to get detailed information. Generally, almost all image input devices are classified into getting wide-view information, getting detailed information, getting 3D information, getting posture or (and) further getting information efficiently by controlling cameras' view directions. TABLE I shows the classification.

TABLE I
CLASSIFYING IMAGE INPUT DEVICES

Categories	Reference numbers
Wide-view	[2][4][7][12]
Detailed	[2][7][12]
3D	[2][3][5][9] to [13]
Posture	[6]
Efficiently	[2][3][5] to [12]

With respect to category '3D', each device utilizes binocular parallax or convergence angle between plural cameras (usually two) except the case using motion parallax from a single camera. In addition to that, many of them have a function to control camera view directions. This design concept seems to be based on mimicking the functions of the human eyes. The author presumes that this is because many researchers attempt to carry out vision-based control of mobile machines like living creatures. At this point, Refs. [2], [7] to [10] and [12] are quite interesting. They introduce each fovea vision system to observe environment using space-variant images suggested by distribution of the human visual acuity (Fig.1), respectively. Refs. [8] to [10] reduce the data amount each image using a method to remap a space-variant image, that is, the resolution becomes lower as going to periphery [14]. They indicate a solution for a trade-off problem between wide-view and detailed images, that is, if we get wider-view information, it usually becomes less detailed. If we realize both simultaneously, they cause enormous increment of pixels each image. In order to solve this problem, Ref. [7] gets foveated images by fusing two images from two cameras, attached with telephoto lens and wide-angle lens, respectively, and Ref. [12] gets it by fusing images with different resolution from a single camera, attached with zoom lens. On the other hand, Ref. [2] gets 120 degrees WAF images using a special optics. These 3 devices can get more detailed information than [8] to [10]. Further, Ref. [2] has an advantage to get such more detailed and wider images using a single camera, on the same time.

Generally, almost all of these input devices are classified into plural categories. The devices described above, especially in the case as the vision system for a mobile robot, are usually required to get information applicable for all-purpose and provide it to the processor for automation. When we assume cooperative works between the human and computer, such as a remote surveillance, it is obvious that wide-view information have many large advantages for all purpose and it is also obvious that 3D and detailed information are helpful. Therefore, an active stereo WAF vision system is one of the most suitable image input devices for all-purpose.

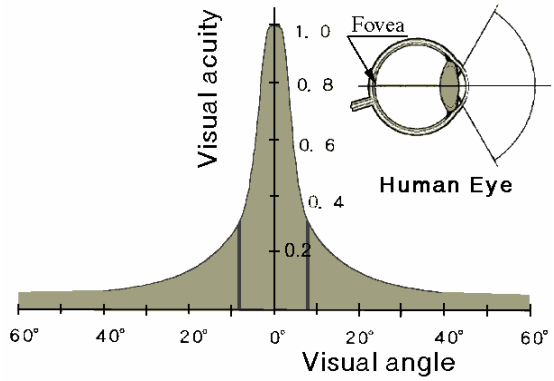


Fig.1 Human visual acuity and a structure of human eye ball

On the other hand, the machine vision system for applications of VR have image output device such as a computer monitor, HMD, 3D display and so on. The output device of [18] projects 3D information on HMD, Ref. [15] has a wide-view immersive display surrounding the user (operator), Ref. [19] has other LCDs to provide the user with more detailed images, and Ref. [21] reconstructs and displays foveated images by detecting view directions of the user (subject) who are watching a monitor using eye tracking device. The author classifies them as shown in TABLE II, that is, displaying 3D information, displaying wide-view information, displaying detailed information and displaying required information efficiently based on view directions.

The first 3 of these 4 categories are quite necessary for cooperative works between the human and computer in the same way as the image input device. Regarding 3D displays, they realizes to have presence. They display 3D images from left and right images on 2 LCDs in the HMD, or on a single panel display, thus an image input device with two cameras needs to be combined with such 3D image output devices. Wide-view information is not only quite effective to have presence but also suitable for all-purpose and detailed information improves the quality of the cooperative works. In order to realize both simultaneously, foveated images have often been applied also to image output devices. Refs. [19] to [21] utilizes foveated images. They display the data, which are reduced and reconstructed using the detected view direction. To reduce the image data is obviously helpful to displaying further much wider-view images. Therefore, the 3D display, which can provide the user with binocular WAF information, is one of the most suitable image output devices for all-purpose.

TABLE II
CLASSIFYING IMAGE OUTPUT DEVICES

Categories	Reference numbers
3D	[18][19]
Wide-view	[15] to [17][19][20]
Detailed	[19]
Efficiently	[18] to [21]

A. Image Input Device

The author realizes WAF images by distorting images based on special optics [2]. We call this special optics Wide Angle Foveated (WAF) lens. Figure 2 shows the appearance and the compound system. The WAF lens is composed of small 7 lenses, which include 2 aspherical

surfaces. The WAF lens is compactly designed to attach with commercially available CCD camera and so on (Elmo CN42H, here). The length is about 35mm and the diameter is about 22mm(at maximum). Figure 3 shows an input image of the WAF lens and the pinhole camera image with the same pixels and the same visual field. White and Black circle in each image correspond to incident angles with 10 and 30 degrees, respectively. Figure 4 shows image height of the WAF lens and the pinhole camera versus incident angle. We can notice a design concept of the WAF lens, suggested by the human visual acuity, is such wide-angle as possible and such high resolution at the center of visual field as possible without increasing pixels.

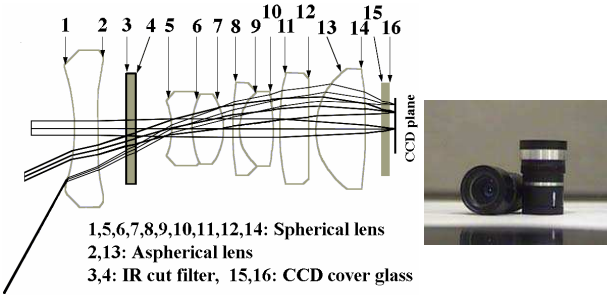
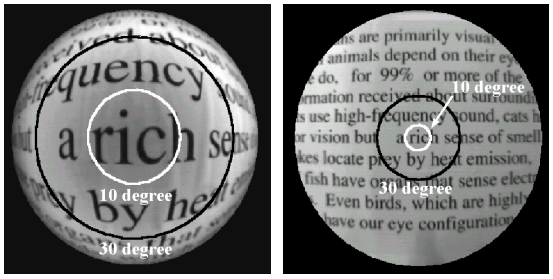


Fig.2 Compound system and a picture of WAF lens



(a)Input image by WAF lens (b)Pinhole camera image

Fig.3 Input images of WAF lens and pinhole camera

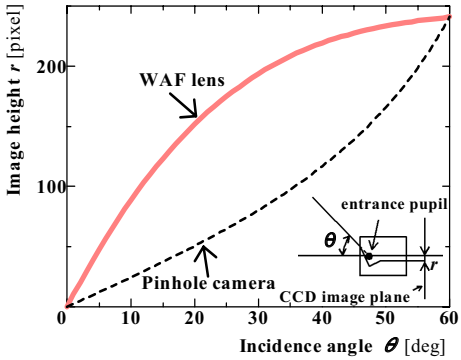


Fig.4 Image height of the WAF lens and the pinhole camera

B. *Camera View Direction Control Device*

A picture of Active Stereo Camera Head is shown in Fig.8. This device has 4 stepping motors (Oriental motor) to control camera view directions with 4 DOF. The base line length, that is a distance between both cameras' approximate optical centers is 300mm. Regarding DOF of Stereo Camera Head and the base line length, some improvement might be necessary because the human change their view directions in such a way and the average distance between both eyes is actually much smaller.

C. *Analogue Video Signal Processing Device*

An analogue video signal processing device, suggested from the structure of the human visual system, characterizes this fovea vision system still more. We call this newly developed device LGN Video Mixing Device (hereafter LGN device) based on its functions. This device outputs not only the left and right two cameras' original video signals but also the mixed video signals based on field multiplexing [5][13] and provides them to plural (two, in this paper) processors and the user. Figure 5 shows a picture of LGN device and the block diagram. LGN device includes video separators and a video field multiplexer to provide each signal. LGN, that is, Lateral Geniculate Nucleus, is a part of thalamus in the human brain, where receives monocular and binocular information from the eyes [22].

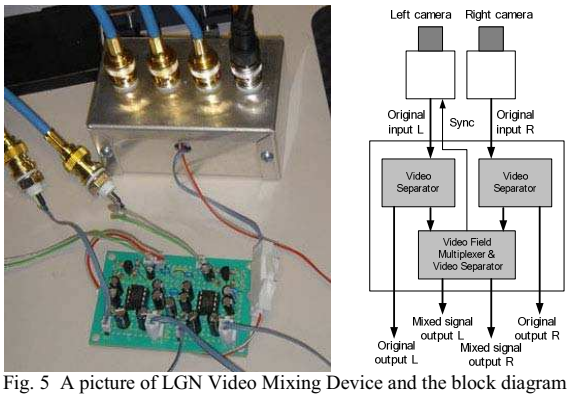


Fig. 5 A picture of LGN Video Mixing Device and the block diagram

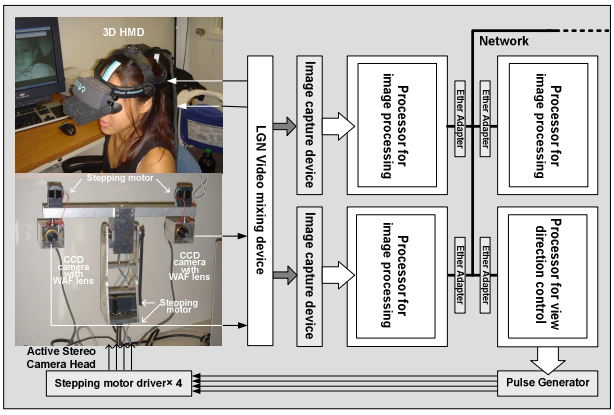


Fig.6 Block diagram of Active Stereo Wide-Angle Foveated Vision System with 3D HMD

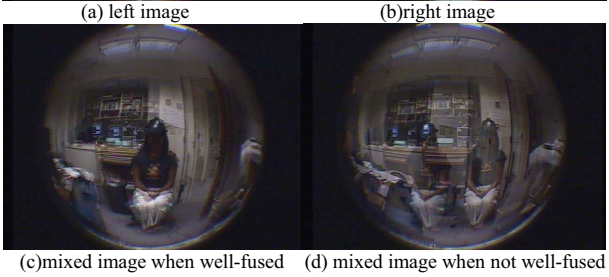
D. *Image Output Device and the System Overview*

A picture of 3D HMD (Virtual Research) is shown in the block diagram of Active Stereo WAF Vision system (ASWAFVS) of Fig.6. The user can observe binocular WAF information as 3D fused images through this HMD, because the LGN device can also output left and right original video signals, that is, WAF images from the WAF lens. On the other hand, the signal mixed at the LGN device is outputted into processors for image processing. These processors are connected with another processor for camera view direction control by network.

Figure 7 shows results of the system implementation. In order to observe well-fused 3D information, both cameras should fixate at the same object correctly. Figure 7(a) and (b) show original images from the left and right cameras, when this system fixates at a person well. While those images are outputted to the 3D HMD, the mixed images as

shown in Fig.7(c) (well-fused) are inputted into processors, where image processing is carried out. Figure 7(d) shows a mixed image when the system doesn't fixate at the same point. In this case, the operator can't observe well-fused 3D information. Figure 8 shows signal flow when this system is implemented automatically for this application.

(a) left image (b)right image



(c)mixed image when well-fused (d) mixed image when not well-fused

Fig.7 Output images from LGN Video Mixing Device

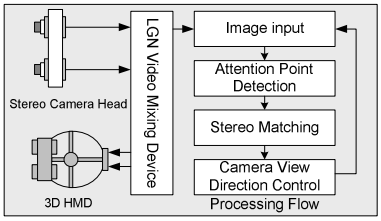


Fig.8 Signal flow of processing for this implementation

III. Feature Generation Algorithm

In order to provide the operator with well-fused 3D information, the features of some attention object, matching with each other, should be found correctly from the left and right images and the stereo camera head should fixate at it. This paper proposes a feature generation algorithm (hereafter FGA), using polar coordinates and Discrete Fourier Transform (hereafter DFT), in order to select the matching features.

A. Fundamentals

Feature generation is to extract and generate feature from candidate objects. The generated features should be robust to select a specific matching feature. The following method is to use DFT to generate features that are invariant to translation and rotation, thus being able to carry out classification task [23]-[25].

Figure 9(a) and (b) show some convex polygonal pattern with some rotation and translation. The former and latter assumed to be cases 1 and 2, respectively. The circle and cross in each figure show the origin of the coordinates (x, y) and moment center (MC). 2 kinds of range data are shown in Fig. 9(c) and (d). Actually they are extracted from the convex polygon, based on each origin. Even if they look different from each other, we can judge whether they correspond to the same pattern using the following way, where the range data have the polar coordinates (r_{org}, θ_{org}) and (r_{mc}, θ_{mc}), which is based on the origin and MC.

Step 1. calculating the moment center in each case
Step 2. transforming the range data from the origin to the moment center
Step 3. the new range data will be circular shift with each other, after transforming to the moment center
Step 4. applying DFT to the new range data as below to extract magnitude spectrum

Assume $\{r_1(n), n=0,\ldots N\text{-}1\}$ is a sequence with finite length N
Assume $\{r_2(n)= r_2[(n\text{-}m) \text{ nod } N], n=0,1,\ldots, N\text{-}1\}$ is the circular shift version of $r_1(n)$

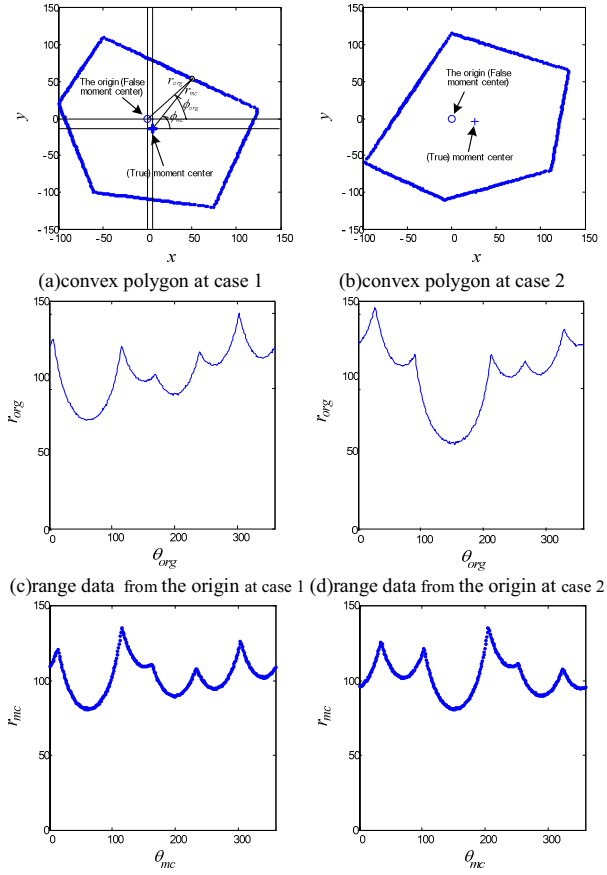
then $$R_1(k)=\mathrm{DFT}[r_1(n)]=\sum_{0}^{N-1} r_1(n)W^{nk} \quad (1)$$

$$R_2(k)=\mathrm{DFT}[r_2(n)]=\sum_{0}^{N-1} r2(n)W^{nk} \quad \text{with } W=e^{-j\frac{2\pi}{N}} \quad (2)$$

$$\Rightarrow \quad R_2(k)=R_1(k)W^{mk} \quad (3)$$

$$\Rightarrow \quad |R_2(k)|=|R_1(k)|, \ k=0,1,\ldots, N\text{-}1 \quad (4)$$

$\Rightarrow$ if $r_1(n)$ and $r_2(n)$ have a circular shift rotation, then $|R_1(k)|=|R_2(k)|$. This criterion can be used to test whether two sequences correspond to the same pattern.



(a)convex polygon at case 1 (b)convex polygon at case 2

(c)range data from the origin at case 1 (d)range data from the origin at case 2

(e)range data from true MC at case 1 (f)range data from true MC at case 2

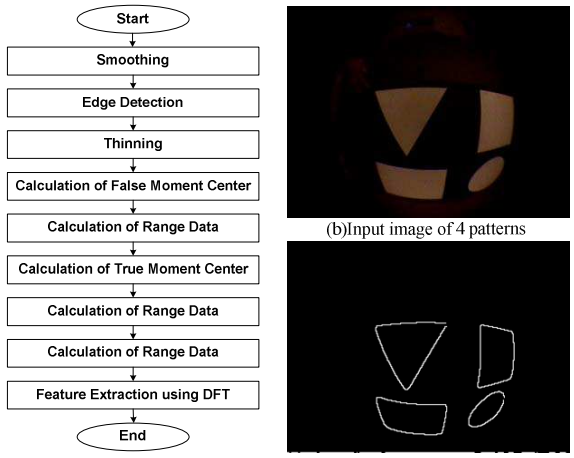
Fig.9 Fundamentals of FGA

B. Application to Space-Variant Images from WAF lens

If we apply the FGA to space-variant images from the WAF lens, we have to pay attention to the following points.

1) Distorion: Not only as an object is projected in the more peripheral area, but also as it is farther away from a camera and its size is larger, the shape of projected image becomes more distorted (see Fig.10(b)). We need to remove the distortion in order to apply the FGA.

2)Space-Variance: This property causes a phenomenon that the density of range data points does not become uniform (see Fig. 12). This makes worse the values of the moment center calculated from range data worse.



(a)Flow char of processing

(b)Input image of 4 patterns

(c)Image after thinning

Fig.10 Processing flow when FGA is applied to WAF image

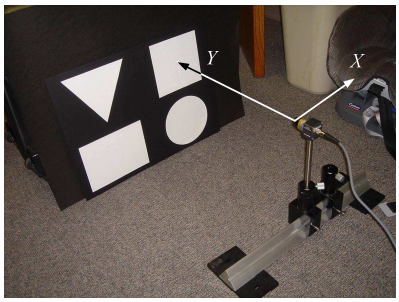


Fig.11 An experimental environment

In order to get sequential range data, the following processing is done, as shown in Fig. 10(a), where Sobel filter and Hilditch method are used for edge detection and thinning. Until deciding the false moment center (FMC) in the perspective coordinate system without distortion, all of basic image processing are done directly in the original coordinate system, that is, with distortion. We calculate the FMC from the edge points after thinning. These edge points are also transformed into perspective coordinates to calculate the range data from FMC. The Original coordinates are transformed to perspective coordinates using Eqs.(5) and (6). The true MC (TMC) and the range data from TMC are calculated from this range data from FMC using linear interpolation. This interpolation plays an important role to make the number of each range data constant, because we use DFT further to generate each feature. FMC and TMC correspond to the origin and (True) moment center in Fig.9(a) and (b) of section *A*., chapter III.

$$r_i = f_0^i\theta_i^3 + f_1^i\theta_i^2 + f_2^i\theta_i \ \text{[pixels]} \tag{5}$$

$$r_{per}{}^i = r_{\max}{}^i \frac{\tan\theta_i}{\sqrt{3}} \ \text{[pixels]} \tag{6}$$

,where each f_k^i (k=0,1,2) is a coefficient determined by camera calibration, and $r_{max}{}^i$ is image height of 60 degree by Eq.(5). Subscript i means left camera or right camera. Equation (5) is drawn in Fig.4.

C. Experiments

Experiments to extract and generate features from the WAF images using the FGA are carried out. Figure 11 shows an experimental environment. Features are generated from 4 kinds of patterns, that is, a circle with a radius of 8[cm], a square with a side 8[cm], a rectangle with sides of 15 and 20 [cm] and a triangle with a side 18[cm]. The screen with these 4 patterns is set with a distance of Y=50[cm] from the camera viewpoint and vertically to the optical axis. Further, we change the camera position horizontally, that is, X=20, 40 and 60[cm] from the center point of each pattern. These positions nearly correspond to incident angles θ=0, 20 and 40 degrees, respectively, that is corresponding to central area (0 to 10 degrees), intermediate area (10 to 30 degrees), and peripheral area (30 to 60 degrees) in the visual field (see Fig.3). The direction of the optical axis and the camera's sliding direction are defined as the Y axis and the X axis, respectively. In order to reduce the cameras' position error, this paper uses average values of 5 images as a generated feature in each condition.

D. Results and Examination

Figure 12 shows 2D pattern plotted in the perspective coordinates, range data from TMC, interpolated range data and magnitude of 360 points DFT (Only 25 points are shown) from the range data, of one image in each case. We see that each pattern has peculiar distributions of the range data and DFT magnitude. However, as a horizontal position, X, increases, they include more noise. It is difficult to distinguish between square and rectangle at X=40, especially (between Figs.12 (f) and (i)).

In order to estimate generated features and examine influences of applying FGA to space-variant images from the WAF lens, we investigate feature correlation, ρ, defined in Eq.(7). The ρ is calculated from the average of five 360-dimensional feature vectors of DFT magnitude.

$$\rho = \frac{\langle R_1, R_2 \rangle}{\|R_1\| \|R_2\|} \tag{7}$$

Figure 13 shows feature correlation between position at X=0 and the other positions in the same pattern. We see that the ρ of each pattern decreases as the X increases. Figure 14 shows feature correlation between each pattern with position at X=0 and all patterns with each position. Circle, triangle and square in each graph correspond to X=0, 20 and 40, respectively. Each pattern has the highest ρ when it is compared with the same pattern. Figures 13 and 14 indicate seven points in the following.

1) Each pattern has stable values of ρ, when compared in the same pattern, although the dispersion increases as the X increases (e.g. circle).

2) Except when X=40, all patterns have very stable values of ρ in the same pattern.

3) Except when X=40, almost all patterns have good features to distinguish the 4 patterns.

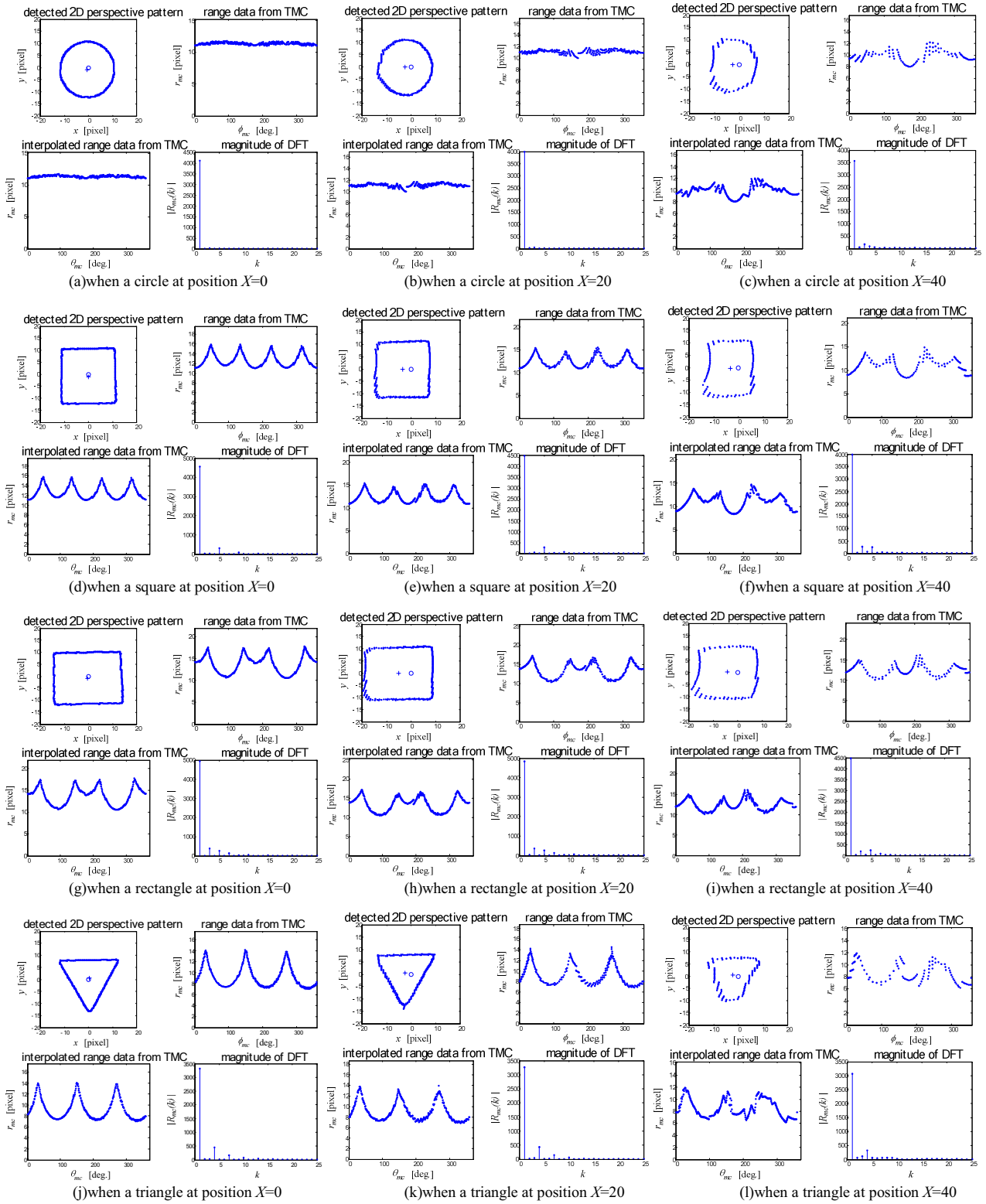
4) When X=40, no feature with X=0 is good to distinguish all 4 patterns

5) Features of circle and triangle at X=0 are good to distinguish itself from other 3 patterns.

6) Feature from triangle by FGA is the most prominent among these 4 patterns.

7) On the other hand, regarding triangle, the dispersion of ρ generated by FGA is the largest, especially when X=40.

Remember the two viewpoints, described in section B, chapter III. It seems that the dispersion of ρ is mainly caused by space-invariance, because the calculation in the



(a)when a circle at position X=0 (b)when a circle at position X=20 (c)when a circle at position X=40

(d)when a square at position X=0 (e)when a square at position X=20 (f)when a square at position X=40

(g)when a rectangle at position X=0 (h)when a rectangle at position X=20 (i)when a rectangle at position X=40

(j)when a triangle at position X=0 (k)when a triangle at position X=20 (l)when a triangle at position X=40

Fig.12 Plotted 2D pattern (top left), range data from TMC (top right), interpolated range data (bottom left) and magnitude of DFT (bottom right) in each case

polar coordinates to remove distortion using Eqs.(5) and (6). Of course, the influence of when removing distortion, which means a precision of Eq. (5) decided by camera calibration, cannot be ignored. However, the phenomenon that the dispersion of ρ regarding triangle is larger than that regarding other patterns seems to prove it. In other words, the other three patterns seems to be more similar to each other than triangle as going to periphery with lower resolution, because the low resolution removes high frequency components, that is, a feature of square and rectangle, compared with circle. On the other hand, the periphery is more sensitive to the camera calibration error. Conversely, this means the confusion between square and rectangle occurs more easily as shown in Fig.14, because the features generated by FGA is rotation-invariant, that is, they are not influenced by circular shift. Therefore, we should pay attention to the results more carefully, because the distortion and space-variance in WAF images from the WAF lens are more closely related. The results of this time conclude that the WAF lens is applicable enough for feature selection in the visual field inside 30 degrees.

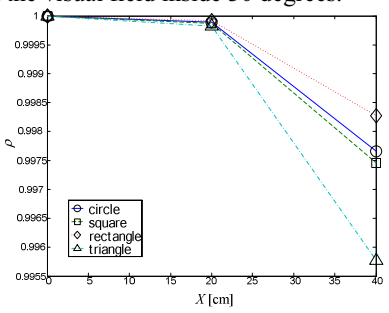


Fig 13 Feature correlation between position at X=0 and the other positions

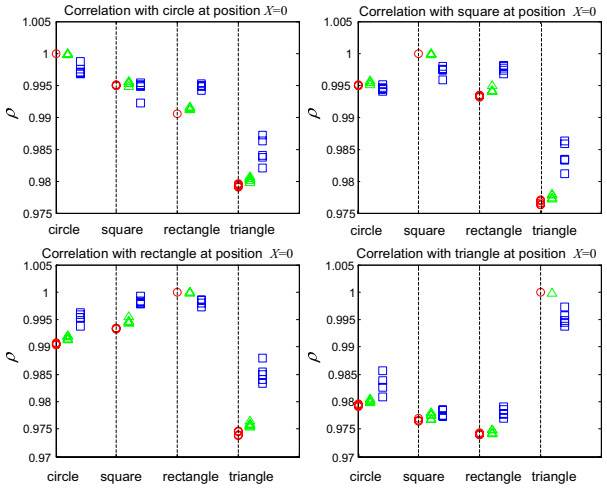


Fig.14 Feature correlation between each pattern with position at X=0 and all patterns with each position

IV. Conclusion

This paper is summarized in the following.

(1) ASWAFVS has been introduced with the design concept to get and provide binocular WAF information for cooperative work between the human and computer.

(2) FGA, which generates features from range data of each pattern using DFT, has been proposed and applied to WAF information to find features for detecting attention point and stereo matching. The experimental results show the visual field within 30 degrees is applicable enough for feature selection.

(3) The influence, which space-variance of WAF image causes to FGA, has been examined based on the experimental results.

As future works, we are planning to progress further both ASWAFVS and FGA thorough some applications. In addition, we would like to focus more on space-variant characteristics of WAF image from the WAF lens.

Acknowledgement

The author thanks to Ryusuke Hayashi and the other members in Prof. Shinsuke Shimojo's laboratory, and thanks to Prof. Pietro Perona in California Institute of Technology, and his laboratory's members for their assistance in carrying out this project. Sota Shimizu was partially supported by a grant from Japan Society for the Promotion of Science, and Wind & Biomass Energy R&D and Information Center.

References

[1] T. Takeda, Brain Engineering, IEICE, (2003.3)
[2] S.SHIMIZU, Multi-Functional Application of Wide Angle Foveated Vision Sensor in Mobile Robot Navigation, Journal of Robotics and Mechatronics, vol.14, No.4, (2002,8)
[3] K. OGAWARA, S.IBA, T. TANUKI, Y. SATO, A. SAEGUSA, H. KIMURA and K. IKEUCHI, Recognition of Human Behaviour using Stero Vision and Data Gloves, Seisankenkyu, Vol.52, No.5, pp.225-230, (2000.5)
[4] K. Yamazawa, Y. Yagi, and M. Yachida, HyperOmni Vision: Visual Navigation with an Omnidirectional Image Sensor, Systems and Computers in Japan, Vol.28, No.4, pp.36-47, (1997.4)
[5] Y.Mitsuya, A. Konno, Y. Abe and M. Uchiyama, Design and Development of Binocular Vision System for a Quadrupedal Walking Robot, 188th SICE Tohoku Research Meeting, Source Number 188-14, (2000.6)
[6] R. Kurazume, S. Hirose, Development of image stabilization system for remote operation of walking robots, Proc. of IEEE Int. Conf. on Robotics and Automation, pp.1856-1861, (2000)
[7] J. H. Elder, Fadi Dornaika, Bob Hou and Ronen Goldstein, Attentive wide-field sensing for visual telepresence and surveillance, In, L. Itti, G. Rees & J. Tsotsos (Eds.), Neorobiology of Attention, Acdemic Press/Elsevier, San Diego, (2004.5)
[8] H. Yamamoto, Y. Yeshurun, and M. D. Levine, An active foveated vision system: Attentional mechanisms and scan path convergence measures, Computer Vision and Image Understanding, Vol. 63, pp.50-65, (1996.1)
[9] R.A Peters II and M. Bishay, Centering Pripheral Features in an Indoor Environment Using a Binocular Log-Polar 4DOF Camera Head, Robotics and Autonomous Systems, Vol.18, pp.271-281,(1996)
[10] A. Bernardino and J. Santos-Victor, Vergence Conyrol for robotic heads using log-polar images, Proc. of 1996 IEEE/RJS International Conference on Intelligent Robots and Systems, pp.1264-1271,(1996)
[11] H. Kim, G. York, G. Burton, E. Murphy-Chutorian and J. Triesch, Design of an Anthropomorphic Robot Head for Studying Development and Learning, Proc. of IEEE International Conference on Robotics and Automation, (2004)
[12] Imai, Furukawa, Tsuji, Stereo vision with zoom lens, Proc. of IEICE Annual Conference , D-536,(in Japanese), (1992)
[13] Y. Matsumoto, T. Shibata, K. Sakai, M. Inaba and H. Inoue, Real-time Color Stereo Vision System for a Mobile Robot based on Field Multiplexing, Proc. of IEEE Int. Conf. on Robotics and Automation, pp.1934-1939, (1997)
[14] G.Sandini and V.Tagliasco, An anthropomorphic retina-like structure for scene analysis, Computer Graphics and Image Processing, 14, pp.365-372, (1980)
[15] S. Morita, K. Yamazawa, and N. Yokoya, Internet telepresence by real-Time view-dependent image generation with omnidirectional video camera, Proc. SPIE Electronic Imaging, Vol. 5018, pp. 51-60, (2003.1).
[16] M. ETO, K. YAMAZAWA and N.YOKOYA, Networked Immersive Telepresence with High-resolution Omnidirectional Video Streams, TECHNICAL REPORT OF IEICE, PRMU2002-155, (2002.12)
[17] K.Tanaka, J.Hayashi, Y. Kunita, M. Inami, T. Maeda and S, Tachi, The Design and Development of TWISTER II: Immersive Full-color Autostereoscopic Display, Proceedings of the ICAT2001, pp. 56-63. (2001)
[18] T. Shibata，T. Kawai and etc., Development of stereoscopic 3D display system with optical correction, Proceedings of XVth Triennial Congress of the International Ergonomics Association,（2003）
[19] K.Iwamoto and K.Tanie, Development of an Eye Movement Tracking Type Head Mounted Display -Capturing and Displaying Real Environment Images with High Reality-, Proc. IEEE ICRA97, pp.3385-3390, (1997)
[20] Kuno, Uchikawa, Yagi, Shikakutokusei wo Koryoshita VR no Kaihatsu (the 1st report), Proc. of the 35th SICE Annual conference, pp.215-216(in Japanese), (1996.7)
[21] J.S. Perry and W. S. Geisler, Gaze-contingent real-time simulation of arbitrary visual fileds, In, B. Rogowitz and T. Pappas (Eds.), Human Vision and Electronic Imaging, SPIE Proceedings. (2002)
[22] E. Kandel, J. Schwartz and T. Jessell, Essentials of Neural Science and Behavior, Appleton&Lange Norwalk, CT, Section VI. Preception, Chapter23. Perception of Form and Motion, pp.425-451, (1995)
[23] G. Shen and T.D. Bui, Invariant Fourier-Wavelet descriptor for pattern recognition, Pattern Recognition, Vol.32, No7, pp.1083-1088, (1999)
[24] S. Theodoridis and K Koutroumbas, Pattern Recognition (2nd edition), Elsevier Science (USA), (2003)
[25] J. Zheng, Q Yang and W. Yang, Signal & Systems (2nd Edition), Higher Education Press (China), (2000)