Scale and Rotation Invariant Recognition Method using Higher-Order Local Autocorrelation Features of Log-Polar Image

Takio KURITA[1], Kazuhiro HOTTA[2], and Taketoshi MISHIMA[2]

[1] Electrotechnical Laboratory, 1-1-4, Umezono, Tsukuba City, 305, JAPAN
[2] Saitama University, 255, Shimo-Okubo, Urawa City, 338, JAPAN

Abstract. This paper proposes a scale and rotation invariant recognition method which uses higher-order local autocorrelation (HLAC) features of log-polar image. Linear scalings and rotations are represented as shifts in the log-polar image which is obtained by re-sampling of the input image. HLAC features of log-polar image become robust to the linear scalings and rotations of a target because HLAC features are shift-invariant. By combining these features with a simple classifier which uses linear discriminant analysis, we can design a scale and rotation invariant recognition system. Robustness to the scalings and rotations are confirmed by experiments on 2D shapes and face recognition. Robustness to the changes of backgrounds is also confirmed by experiments on face recognition.

1 Introduction

When people watch a target, they capture the target with the center of the visual field by eye movement. The density of the structure of human's vision system increases toward the center of the visual field and decreases from the center toward the periphery. The center of retina is called "fovea". Fovea has rich eyesight and its visual angle is only five degrees.

In the field of computer vision, the interest in active vision which acquires necessary information by moving the cameras is increasing[1, 2]. The one of the reasons of the interest is due to the developments of small, light, and fast camera which can be controlled by computer. It is known that space variant sensor like biological vision is effective in active vision[3, 4] . The mechanism which always keeps the target at the center of the visual field is called "gaze control". That is one of the important elements of active vision system.

Shift invariance is important for recognition system which uses usual static camera because the target changes the position within the image frame. On the other hand, scale and rotation invariance becomes more important in recognition system which uses active vision head with gaze control mechanism, because active vision system can maintain the target with the center of camera. In this paper we consider the second case, where we can use active vision head.

One of the simplest models of space variant sensor is log-polar transformation. In the log-polar image, higher weights are given for the central region of the original image than peripheral region. This property of the log-polar images is good for target recognition because the peripheral region often includes unnecessary information such as background. There is another important property. In the log-polar coordinate, linear scaling and rotation are represented as shifts if the origin of the polar plane does not move. Therefore, shift-invariant features extracted from log-polar images become invariant to scalings and rotations of the target.

In this paper, we use higher-order local autocorrelation (HLAC) features [5] which are inherently shift-invariant, additive, and computationally inexpensive. Kurita et al.[6, 7] developed a real-time face recognition system based on computation of HLAC features. Goudail et al.[8] obtained a peak recognition rate of 99.9% by using a large database of 11,600 images of 116 different faces. These attempts assume the images taken by the usual static camera. Here we propose the recognition method for images which are taken by active vision system. Since HLAC features of log-polar images become invariant to scalings and rotations, we can design a scale and rotation invariant recognition system by combining these features with a simple classifier which uses linear discriminant analysis.

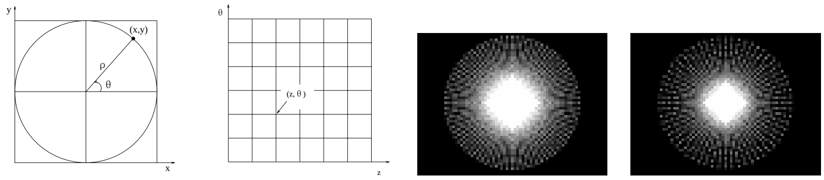
In section 2 we describe the property of log-polar transformation which is one of the simplest models of space variant sensor. Then HLAC features of log-polar image are introduced in section 3. In section 4 experimental results of 2D shapes and face images recognition.

2 Log-polar image

Input image is generally represented as a collection of pixel points on the Cartesian coordinate. Log-polar image can be constructed by the following transformations of the coordinates. At first, the point (x, y) on the Cartesian coordinate is transformed to the point $(\rho = \sqrt{(x^2 + y^2)}, \theta = \arctan(\frac{y}{x}))$ on the polar coordinate. Then the point on the polar coordinate is transformed to the point $(z = \log(\rho), \theta)$ on the log-polar coordinate by taking the logarithm of the scale ρ. Figure 1 (a) and (b) show Cartesian coordinate and log-polar coordinate.

In this paper, we use the re-sampling method by the reverse transformation to obtain a log-polar image from a input image. To obtain the pixel value at the point (z_i, θ_j) on the log-polar image, the point is reversely transformed to the point $(\exp(z_i)\cos(\theta_j), \exp(z_i)\sin(\theta_j))$ on the Cartesian plane. Then the value of the point (z_i, θ_j) is estimated as the average intensity value of the neighboring points of the back-projected point $(\exp(z_i)\cos(\theta_j), \exp(z_i)\sin(\theta_j))$ on the input image. We can obtain a log-polar image by performing this estimation for all points on the log-polar image. Figure 1 (c) and (d) show sampling point of input image. It is noticed that sampling density increases toward the center of image and decreases from the center toward the periphery. The number of sampling points depends on the resolution of re-sampling to obtain the log-polar image.

Scalings of a target are represented as shifts along z (= $\log(\rho)$) axis on the log-polar image (coordinate). Rotations of a target are also represented as shifts along θ axis but the upper end and the lower end are connected because 0 and 2π are equivalent.



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

Fig. 1. Sampling points of input image to construct the log-polar image. (a) Cartesian coordinate. (b) Log-Polar coordinate. (c) The resolution of input image is 320×240. The resolution of log-polar image is 80×80. (d) The resolution of log-polar image is 60×60.

3 Higher-order local autocorrelation features of log-polar image

To obtain scale and rotation invariant features, we have to extract shift invariant features from log-polar image because the scalings and rotations are represented as shifts in log-polar image.

It is well known that autocorrelation function is shift-invariant. Its extension to higher orders is higher-order autocorrelation function. The Nth-order autocorrelation functions with N displacements $(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_N)$ from the reference point $\boldsymbol{r}$ are defined by

$$x^N(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_N) = \int I(\boldsymbol{r})I(\boldsymbol{r}+\boldsymbol{a}_1)\cdots I(\boldsymbol{r}+\boldsymbol{a}_N)d\boldsymbol{r} \quad (1)$$

where function $I(\boldsymbol{r})$ denotes a log-polar image and $\boldsymbol{r} = (z = log(\rho), \theta)$. The number of these autocorrelation functions obtained by the combination of the displacements over the image are enormous. Here we reduce them for practical application. At first, we restrict the order N up to the second (N = 0,1,2). Then, we also restrict the range of displacements to within a local 3×3 window, because the correlation within local region is much higher than the correlation between far points. By eliminating the displacements which are equivalent by the shift, the number of the patterns of the displacements are reduced to 35. Since these features are obviously invariant to shift, HLAC features extracted from log-polar images become robust to linear scalings and rotations of a target in the original image. By combining these features with a simple classifier which uses linear discriminant analysis, we can design a scale and rotation invariant recognition system.

4 Experiments

To show effectiveness of the proposed method, we have performed experiments using images with different scales and rotations of targets. Recognition targets are 2D shapes and human faces.

In the following experiments, we used a simple classifier based on linear discriminant analysis. New features $\boldsymbol{y}$ are computed by linear combinations of the HLAC features $\boldsymbol{x}$ as $\boldsymbol{y} = A^T\boldsymbol{x}+\boldsymbol{b}$. The discriminant criterion $J = \mathrm{tr}(\hat{\Sigma}_W^{-1}\hat{\Sigma}_B)$ is maximized to obtain the coefficient matrix A. To discriminate unknown image, we can use a nearest neighbor classifier which compares distances between the unknown input and each class means in the constructed discriminant space.

4.1 2D shapes recognition

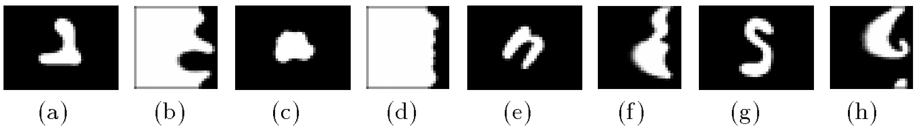


Fig. 2. Examples of 2D shape images and their log-polar images. (a) Shape 1. (b) Log-Polar image of (a). (c) Shape 2. (d) Log-Polar image of (c). (e) Shape 3. (f) Log-Polar image of (e). (g) Shape 4. (h) Log-Polar image of (g).

Figure 2 shows 2D shapes [9, 10] which are used in the experiment and their log-polar image. In this experiment, we collected images with different sizes (6 sizes) and rotations (12 angles) for each shape in Figure 2. The different sizes are obtained by changing the distance between the camera and the object. The rotation were changed about $\pi/6$ rad at a time by hand over the range from 0 to 2π rad. The number of images used in this experiment is 288 (4 classes $\times$ 6 sizes $\times$ 12 rotations). The resolution of these images is 320 $\times$ 240 pixels and each of images is binalized by Otsu's thresholding method [11, 12]. Since the resolution of input images and the resolution of log-polar images effect to the recognition rate, we measured them with different resolution of input images (320 $\times$ 240, 160 $\times$ 120, 80 $\times$ 60) and log-polar images (120 $\times$ 120, 80 $\times$ 80, 60 $\times$ 60, 40 $\times$ 40, 30 $\times$ 30, 20 $\times$ 20).

In this experiment, all of the images were used to construct discriminant space and the recognition rate were estimated by leave-one-out method. The recognition rates were 100% for all combinations of resolutions of input images and log-polar images. From these results, we can say that the proposed method has ability to recognize 2D shapes with different sizes and rotations.

Next, to show the robustness to scalings, we have performed the experiment in which images with only one size (4 classes $\times$ 1 size $\times$ 12 rotations) are used in learning and then the rest (4 classes $\times$ 5 sizes $\times$ 12 rotations) are used for evaluation of the recognition rate. For the comparison, we also calculated the recognition rates by using HLAC features extracted from original images directly. In this experiment, the resolution of input images was set to 160 $\times$ 120 pixels. The results are shown in Table 1 (a) and (b). The recognition rates obtained by HLAC features extracted from log-polar images are about 99%, while the best recognition rates obtained by HLAC features extracted from input images directly are at most 50 %.

Figure 3 (a) and (b) shows discriminant spaces constructed in this experiment. In Figure 3 (a) and (b), each of four classes is plotted with different symbols (Each symbol corresponds to each class of Figure 2). Figure 3 (a) and (b) shows discriminant space constructed from HLAC features extracted from log-polar image and HLAC features extracted from input images, respectively. Since there is the case which include noises when all eigenvalues are used, we restricted the dimension by a constraint in which cumulative proportion becomes more than 99%. Consequently, the dimension of discriminant space constructed from HLAC features extracted from log-polar image became three. However, the dimension of discriminant space of HLAC features of input image became two. It is noticed that Figure 3 (a) consists four group which corresponds to the original four classes, while each class is separated into exactly five groups which correspond to five scales of a target in Figure 3 (b). From these results, the proposed method becomes more robust to scalings of targets by using log-polar transformation.

Similarly, to show the robustness to rotations, we have performed the experiment in which only one rotation (4 classes $\times$ 6 sizes $\times$ 1 rotation) is used in learning and the rest (4 classes $\times$ 6 sizes $\times$ 11 rotations) are used for evaluation of the recognition rate. Again, for the comparison, we also calculated the recognition rates by using HLAC features extracted from original images directly. In this experiment, the size of input images was also set to 160 $\times$ 120 pixels. The results are shown in Table 1 (c) and (d). The recognition rate for HLAC features extracted from log-polar images is about 95 % except the case in which the resolution of the log-polar image is 40 $\times$ 40, but the best recognition rate obtained by HLAC features extracted from input images is only 55.68 %. From these experiments, the proposed method is very robust to the changes of scalings and rotations of 2D shapes.

HLAC features extracted from log-polar image are scale and rotation invariant but then do not shift invariant. To evaluate the sensitivity of the recognition rate to shift of a target, we have measured the recognition rates by changing distance between the center of the image and the center of gravity of the target. The resolution of input images is set to 160 $\times$ 120 pixels and the resolution of log-polar images is 20×20 pixels. The images without shifts were used in learning and recognition rates were estimated by right shifted images, respectively. The result is shown in Figure 3 (c). The horizontal axis shows distance between the center of image and the center of gravity of the target. The vertical axis shows the recognition rate. During the distance between center of image and the center of gravity of a target is less than 7 pixels, the recognition rate achieves more than

Table 1. Generalization to scalings and rotations (The resolution of input image is 160×120)

(a) HLAC features extracted from log-polar image

The resolution of log-polar image	Rate (%)
40×40	98.75
30×30	99.58
20×20	100.00

(b) HLAC features extracted from original image

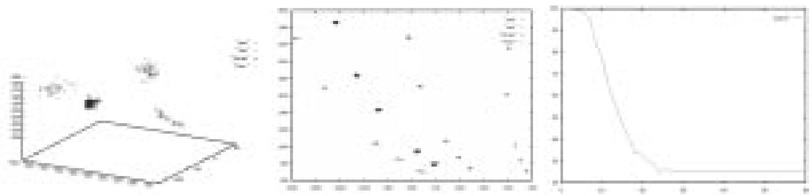
The resolution of input image	Rate (%)
80×60	47.92
40×30	50.00
20×15	37.08

(c) HLAC features extracted from log-polar image

The resolution of log-polar image	Rate (%)
40×40	76.14
30×30	95.83
20×20	96.59

(d) HLAC features extracted from input image

The resolution of input image	Rate (%)
80×60	37.12
40×30	55.68
20×15	26.89



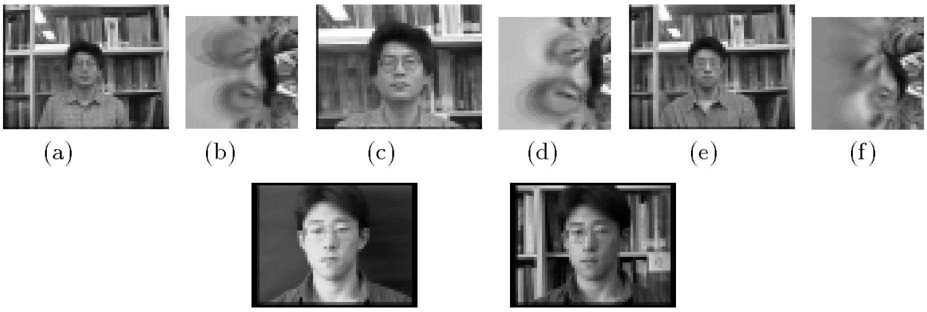
(a) (b) (c)

Fig. 3. Generalization to scalings: (a) HLAC features extracted from log-polar image.(The resolution of log-polar image is 20×20. The dimension of the discriminant space is three.) (b) HLAC features extracted from input image.(The resolution of input image is 40×30. The dimension of the discriminant space is two.) (c) Effect by changing distance from center.

about 95%. But, as the distance becomes more far, the recognition rate becomes suddenly very low. This means the recognition rate of the proposed method is very sensitive to the shift of the target. However, this sensitivity of shift of the target is not harmful for the recognition system combined with active vision. It is useful to identify the position of the target because we can do matching more accurately by checking the distances in discriminant space constructed by linear discriminant analysis. Currently, we are developing a face detection method using this property.

4.2 Face recognition

In the previous experiments, binary images are used but the proposed method can be applied to gray scale images (face images) without any changes. Examples of face images with different sizes and their log-polar images are shown in Figure 4 from (a) to (f). The face images were taken under fluorescent lights, namely no special lighting is used. We collected the face images whose scalings were



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

(g) (h)

Fig. 4. Examples of face images, their log-polar images, and face images with different background. (a) Face image 1. (b) Log-polar image of (a). (c) Face image 2. (d) Log-polar image of (c). (e) Face image 3. (f) Log-polar image of (e). (g) Face image 4. (h) Face image 5.

changed 7 steps by changing zoom parameter of camera, which can be controled by computer. The number of images are 1000 (5 persons $\times$ 200 images) from 5 persons. The resolution of original images is 320 $\times$ 240 pixels. Since both the resolution of input images and the resolution of log-polar images influence the recognition rates, we have performed experiment with changing the both resolutions. At first, all the images were used in learning and the recognition rates were estimated by leave-one-out method. We obtained good results. For some combinations of the resolution of input images and the resolution of log-polar images, the best recognition rates achieved over about 99.70%. Even the worst recognition rate achieved 97.90%. This means that the proposed method has ability to recognize face images with different sizes. In this experiment, the recognition rate was highest when the resolution of log-polar images was 60×60 pixels. The resolution of input images also affect to the recognition rate. In this experiment, highest recognition rate 99.70% was obtained when the resolution of input images was 160×120 pixels.

Since log-polar image gives higher weights for center region, namely face region, than peripheral region, namely background or clothes, it is expected that the recognition rate is also less influenced by the changes of background. To investigate such effect of different background, we have performed face image recognition with different background. Examples of face images with different background are shown in Figure 4 (g) and (h). The 1400 images (20 images $\times$ 7 scales $\times$ 2 background $\times$ 5 persons) are gathered. All the images were used in learning and the recognition rate was estimated by leave-one-out method. Almost all recognition rates were 100%. As expected, this shows that the proposed method is robust to the changes of backgrounds.

References

1. J.Y.Aloimonos and I.Weiss, "Active vision," International Journal of Computer Vision, pp.333-356, 1988.
2. D.H.Ballard, "Animate vision," Artificial Intelligence, Vol.48, pp.57-86, 1991.
3. G.Sandini and V.Tagliasco, "An anthropomorphic retina-like structure for scene analysis," Computer Graphics and Image Processing, Vol.14, pp.365-372, 1980.
4. L.Massone, G.Sandini, and V.Tagliasco, "Form-invariant: Topological mapping strategy for 2D shape recognition," Computer Vision, Graphics and Image Processing, Vol.30, pp.169-188, 1985.
5. N.Otsu and T.Kurita, "A new scheme for practical flexible and intelligent vision systems," *Proc. IAPR Workshop on Computer Vision, Tokyo*, pp. 431-435 , 1988.
6. T.Kurita, N.Otsu, and T.Sato, "A face recognition method using higher order local autocorrelation and multivariate analysis," Proc. 11th IAPR International Conf. on Pattern Recognition, pp.213-216, 1992.
7. T.Kurita ,"A study on applications of statistical methods to flexible information processing (in Japanese) ,"Researches of the Electrotechnical Laboratory, No.957, November,1993.
8. F.Goudail, E.Lange, T.Iwamoto, K.Kyuma, and N.Otsu, "Face recognition system using local autocorrelations and multiscale integration," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.18, No.10, 1996.
9. I.Sekita, T.Kurita, and N.Otsu, "Complex autoregressive model for shape recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.PAMI-14, No.4, pp.489-496, 1992.
10. T.Kurita, I.Sekita, and N.Otsu, "Invariant distance measures for planar shapes based on complex autoregressive model," Pattern Recognition, Vol.27, No.7, pp.903-911, 1994.
11. N.Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Systems, Man, and Cybernetics, Vol.SMC-9, No.1, pp.62-66, 1979.
12. N.Otsu ,"Mathematical studies on feature extraction in pattern recognition (in Japanese)," Researches of the Electrotechnical Laboratory,No.818, July, 1981.
13. T.Kurita , K.Hotta , T.Mishima, "Scale and rotation invariant recognition of 2D shapes and face images using higher-order local autocorrelation features of log-polar image (in Japanese)," Techinical report of IEICE,PRMU96-212,pp.151-158,1997.
14. T.Kurita , K.Hotta , T.Mishima, "Scale and rotation invariant recognition of Face images Using Higher Order Local Autocorrelation Features of Log-Polar Image (in Japanese)," Trans. of IEICE, Informatoin and Communication Engineers D-II, Vol.J80-D-II No.8, pp.2209-2217,1997.
15. K.Hotta, T.Kurita, T.Mishima, "Scale and rotation invariant recognition of 2-D shape using higher-order local autocorrelation features of log-polar image (in Japanese),"Proceedings of the 1997 IEICE general conference,D-12-229,pp.436,1997.
16. O.Hasegawa, K.Itou, T.Kurita, S.Hayamizu, K.Tanaka, K.Yamamoto, N.Otsu, "Active agent oriented multimodal interface system," Proc. of IJCAI'95, pp.82-87, 1995.
17. M.Kirby , L.Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," IEEE Trans. Patt. Anal. and Mach. Intell., Vol.12, pp.103-108,1990
18. N.Kita ,"Active vision systems using human vision as inspiration ", Trans. of Information Processing Society of Japan , Vol.36, No.3, pp.264-270,1995
19. Y.Kuniyoshi, N.Kita, S.Rougeaux, and T.Suehiro, "Active stereo vision system with foveated wide angle lenses," Proc. of 2nd Asian Conf. on Computer Vision, Vol.I, pp.359-363, 1995.
20. L.Berthouze, S.Rougeaux, F.Chavand, and Y.Kuniyoshi, "Calibration of a foveated wide-angle lens on an active vision head," Proc. of CVPR'96, pp.183-188, 1993.
21. H.Yamamoto, Y.Yeshurun, Martin D.Levine, "Foveated vision system with attentional mechanisms (in Japanese) ,"Trans.IEICE D-II, Vol.J77-D-II,, No.1, pp.119-129, 1994.
22. J.Van der Spiegel, G.Kreider, C.Claeys, I.Debusschere, G.Sandini, P.Dario, F.Fantini, P.Bellutti, G. Soncini, "A foveated retina-like sensor using ccd technology,"Analog VLSI and Neural Network Implementations, Boston,1989
23. M.Tistarelli and G.Sandini, "On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.15, NO.4, pp.401-410, 1993.
24. M.Kreutz, B.Völpel, and H.Janβen, "Scale-invariant image recognition based on higher-order autocorrelation features," Pattern Recognition, Vol.29, No.1, 1996.

This article was processed using the LaTeX macro package with LLNCS style