A 3D log-polar transform for pattern recognition

著者	Nakata Takayuki, Bao Yue, Fujiwara Naofumi
journal or publication title	Systems and Computers in Japan
volume	38
number	13
page range	49-57
year	2007-11-30
URL	http://hdl.handle.net/2297/7390

doi: 10.1002/scj.20564

A 3D Log-Polar Transform for Pattern Recognition

Takayuki Nakata,¹ Yue Bao,² and Naofumi Fujiwara³

¹Faculty of Engineering, Toyama Prefectural University, Toyama, 939-0398 Japan

²Faculty of Engineering, Musashi Institute of Technology, Tokyo, 158-8557 Japan

³Faculty of Engineering, Kanazawa University, Kanazawa, 920-1192 Japan

SUMMARY

The log-polar transformation is a model of the foveal sensor, which is often seen in the visual apparatus of living organisms. The log-polar transformation is widely known in the field of image recognition, and has often been used for facial image recognition. However, this detection method is limited to 2D environments, since it is invariant only to scaling and rotation. This paper proposes a new log-polar transformation which also detects affine transformation parameters. The proposed method can detect the posture of a 2D image in a 3D environment, as well as scaling and rotation in a 2D environment. © 2007 Wiley Periodicals, Inc. *Syst Comp Jpn*, 38(13): 49–57, 2007; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/scj.20564

Key words: log-polar transform; affine-invariant recognition; pattern matching; computer vision.

1. Introduction

The resolution of the foveal sensor of some organisms is highest at its center and lower at its circumference. The idea of such a foveal sensor is known to be useful for active vision systems. The log-polar transformation is a model of the simplest foveal sensor, and has many interest-

ing properties [1]. For example, a change of the size of an object in a log-polar image is represented by a parallel translation. This property is very useful for constructing recognition algorithms that are invariant to the size and rotation of an object. The log-polar transformation directly preserves shape features such as angles. That is, local image processing (such as filtering, edge detection, and simple template-matching calculations) can be applied directly to a log-polar image. Therefore, the log-polar transformation is widely known and has been applied to facial image recognition. However, the log-polar transformation can detect only changes of the size of an object and changes of its rotation. When a 2D image in a 3D environment is seen from two camera viewpoints, their projected images have a relationship that approximates the affine transformation [2]. Therefore, it is very important for 3D analysis of image in computer vision to establish a stable 2D image recognition method in the affine transformation. This paper proposes an expanded log-polar transformation method which can detect the affine parameters as well as the size and rotation of an object.

There have been various methods [3–7] of performing an affine transformation on an arbitrary-shaped 2D figure without changing its shape, since there are many applications (e.g., image search, parcel identification, and position or posture recognition of components on a production line). An improved log-polar method which can detect affine parameters by using a log-log transformation has been proposed [8]. In this method, the logarithmic representation is used for both transfer coordinates, so that an

inclination perpendicular to the optical axis of a camera is represented by a parallel translation. This transformation cannot continuously represent a rotation around the optical axis of the camera, and therefore, different log-log images must be used for each rotation angle. If the rotation angle is measured with a high accuracy, the memory capacity must be increased.

For the detection experiments in this paper, four images were used: two log-polar images (360×512 pixels), one template image T (256×256 pixels), and one input original image I . Since the log-log transformation image does not contain angular information, 360 transformation images (512×512 pixels) are required for each of I and T , together with the two original images (722 images in all) to achieve the same accuracy. The proposed algorithm offers a significant reduction of memory requirements.

2. The Log-Polar Transformation

Let the coordinates of the original image be (x, y) . The coordinates (ρ, ϕ) of its log-polar transformation are given by

$$(\rho, \phi) = \left(\log(\sqrt{x^2 + y^2}), \arctan\left(\frac{y}{x}\right) \right) \quad (1)$$

Figure 1 shows a standard image of the letter P and its log-polar transformed image. Figure 2 is the same as Fig. 1, but with the size of the letter increased. Note that the log-polar image has the same size and shape, but the position is parallel-shifted toward the ρ -axis. Figure 3 shows a case in which the letter is rotated, and its log-polar image is parallel-shifted toward the ϕ -axis. As these examples indicate, the log-polar transformation can detect an image regardless of its size and direction. These properties are useful for such applications as facial-image recognition [9, 10], although they are limited to a region around the center of the original image, and thus are not applicable to the affine transformation.

Fig. 1. Log-polar transform-standard.

Fig. 2. Log-polar transform-magnification.

3. Relationship between Affine Transformation and Log-Polar Transformation

Let the template image (prepared beforehand as shown in Fig. 4) be T . Let an image taken by a camera (Fig. 5) be I . The relationship between I and T can be approximately described as follows: after an invariant axis is added to T , and the image is reduced only in a direction perpendicular to the invariant axis, the whole image is enlarged or reduced [7]. Let the angle of the invariant axis in T be θ . Let the invariant axis in I be ψ . Then, the relationship between T and I can be approximated by an affine transformation. Now, let us find a procedure by which the result of detection of the log-polar transformation is made to correspond to the affine transformation by using appropriate parameters.

Let the coordinates of the template image be (x_t, y_t) , and the coordinates of its input image be (x_a, y_a) . If these coordinates have an affine-transformation relationship with matrix C , the following equation holds:

$$\begin{pmatrix} x_a \\ y_a \end{pmatrix} = C \cdot \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \begin{pmatrix} x_m \\ y_m \end{pmatrix} \quad (2)$$

where x_m and y_m are the magnitudes of the parallel transformation. These parameters can be obtained in a log-polar transformation by detecting feature points M_p and M_i from

Fig. 3. Log-polar transform-rotation.

Fig. 4. Template image.

images T and I . In a facial image, a corner of image or a part at which the intensity suddenly changes is generally used as the feature point, although any point can be used for this purpose so long as its correspondence is correct. In this paper, it is assumed that such corresponding points are available, as these in the log-polar transformation.

The affine parameter C can be expressed in terms of a singular-value decomposition as

$$C = \begin{pmatrix} \cos \psi & \sin \psi \\ -\sin \psi & \cos \psi \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (3)$$

This equation shows that the affine transformation is equivalent to the enlargement or reduction of an arbitrary image with arbitrary magnifications of λ_1 and λ_2 in directions which are perpendicular and parallel to an arbitrary invariant axis. As Eq. (3) shows, the directions of scaling of λ_1 and λ_2 perpendicularly intersect the template and the input. Therefore, if these parameters are obtained, the affine parameters from T to I can be calculated.

Therefore, if θ , ψ , λ_1 , and λ_2 are obtained, affine-invariant detection of the template image and the input image can be performed.

Fig. 5. Input image.

4. Relationship between Affine Transformation and T in 3D Posture

Figure 6 shows the relationship between a template image in a 3D space and its input image seen from an arbitrary position. The affine transformation can be regarded as equivalent to a two-stage projection. First, T is obtained by normally projecting T onto plane P which is parallel to camera plane C . Then, T is centrally projected onto C so that the input image I is obtained. Note that the position and posture of the template image in a 3D space have 6 degrees of freedom.

In this paper, it is assumed that x_m and y_m (movements in the directions perpendicular to the camera axis) have already been detected. Therefore, z_m (movement in the direction of the camera axis) and three posture angles (4 degrees of freedom in all) are the parameters to be detected.

In this paper, the transform equation has 4 degrees of freedom, since the affine transformation consists of a 2×2 matrix. This corresponds to the 4 degrees of freedom to be detected.

Equation (3) can be rewritten as

$$C = \begin{pmatrix} \cos \psi & \sin \psi \\ -\sin \psi & \cos \psi \end{pmatrix} s \begin{pmatrix} 1 & 0 \\ 0 & \tau \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (4)$$

Fig. 6. Relation of T and I in 3D space.

where $s = \lambda_1$ is a scaling coefficient, and $\tau = \lambda_2/\lambda_1$. The angle of rotation around the invariant axis η is given by

$$\eta = \arccos \tau \tag{5}$$

where s depends on the distance z_m between the focal point of the camera and T , and θ, ψ, η are 3 degrees of freedom (the yaw, pitch, and roll of the posture angles) of T .

5. Method of Detection

Since the rotation and scaling of the original image appear as parallel translations in a log-polar transformation, their detection is easy. The affine parameters can be obtained by specifying θ, ψ, λ_2 , and λ_1 .

The log-polar transformation can be applied to the whole image. In this paper, however, the affine parameters were detected by taking a circular region around an object. Let us explain the effect of taking the circular region. Figure 7 shows a test template having a ring pattern. Figure 8 shows the input image to which an affine transformation was applied. The ring (circle) in the template is converted to an ellipse in the input image after the affine transformation. Figure 9 shows the log-polar transformation of the template image (ring), with the vertical axis being the θ -axis and the horizontal axis being the ρ -axis. Figure 10 shows the log-polar transformation of the input image, with the vertical axis being the ψ -axis and the horizontal axis being ρ . Figure 10 shows a wave pattern, with the intervals of the large and small ellipses being equal in the direction of the ρ -axis. The lengths of all the radiating lines (in Fig. 8) appear to be equal in Fig. 10, and they are also equal to the radiating lines in Fig. 7.

It is possible to find features of affine-invariance textures in a log-polar transformed image, so long as it is in a direction parallel to the ρ -axis. Note that there are distortions

Fig. 8. Oval input image.

Fig. 9. Ring template transformed by log-polar.

Fig. 7. Ring template image.

Fig. 10. Oval input transformed by log-polar.

tions between the θ -axis (Fig. 9) and the ψ -axis (Fig. 10), and that the intervals of straight lines do not necessarily agree. However, the angles of the invariant axis (directions of λ_1 and λ_2) are not changed in the affine transformation. Therefore, it is possible to identify such parts in each of Figs. 9 and 10, and θ and ψ can be detected. In Fig. 8, the direction of the invariant axis is the direction of the long axis of the ellipse, and the rightmost line of the wave shape in Fig. 10.

As explained above, affine parameters for the transformation from T to I can be detected by setting a ring in the two images, and detecting the corresponding parts.

6. Algorithm

6.1. LMedS estimation

After the input image I has been obtained, the following procedures are performed. Since all the matching operations are not necessarily successful in practice due to such factors as noise, least median estimation (LMedS), which is a robust method, is applied. Data which are distant from the finally detected ellipse by more than $\pm\delta$ are defined as exceptional values. The least square method (LMS) in which the estimated value is the mean square difference from the whole body of data influences all the detection results. LMedS has the advantage that the detected results are not influenced by exceptional values unless they exceed 50%, since the median is the estimated value [11]. However, unlike LMS, LMedS is a nonlinear equation and is thus very difficult to solve. Therefore, a temporary solution is obtained by random sampling and evaluation. Practical procedures are as follows.

(i) F data are selected randomly from a total of n data. A total of n data means n sets of corresponding points between the template image and an input image.

(ii) The parameters of the function are estimated by using the F data and the least-squares method.

(iii) The similarity of the parameters to all the data is estimated by LMedS.

(iv) Process (i) to (iii) is repeated q times, and the parameter showing the best similarity is chosen as the solution.

The ellipse to be detected is given by

$$\left(\frac{x \cos \zeta + y \sin \zeta}{a}\right)^2 + \left(\frac{x \sin \zeta - y \cos \zeta}{b}\right)^2 = 1 \quad (6)$$

where ζ is the rotation angle, a is the long axis, and b is the short axis.

The estimation ellipse can be determined by obtaining three more points of matching marks on the ellipse, since this equation has three unknowns.

Let the ratio of the number of exceptional values to the number of all the data be ϵ . The probability P that no exceptional value is contained in q random samples is given by

$$P = 1 - (1 - (1 - \epsilon)^F)^q \quad (7)$$

P is equal to the probability that all F sampling data are within $\pm\delta$.

P can be increased by reducing F so that q increases. Therefore, F was chosen as 3 in this paper: this is the minimum number by which an ellipse can be determined. In this algorithm, $n = 18$ and $q = 6$ were empirically selected.

6.2. Detection procedure

(1) The log-polar transformation is applied to T and I . Let their images be T_L and I_L , respectively.

(2) A line having a certain length is drawn parallel to ρ on T_L , with angle θ_T being random.

(3) Based on the ordering colors on the line (in Procedure 2), template matching is applied to I_L using the successive residue method, and the matched parts are marked.

(4) This process is repeated F times with different θ_T , and marks are applied to I_L .

(5) The marks on I_L are inversely transformed. If the inversely transformed marks contain no noise, they lie on the ellipse which is to be detected. The ellipse is approximated by the ellipse obtained by the LMS method.

(6) The estimated ellipse is evaluated by using LMedS. In a single practical operation, the square differences of the ellipse (estimated from the selected F points using LMS) and the n points are calculated. The median of this quantity is regarded as the LMedS evaluation for this operation.

(7) Operations (1) to (6) are repeated q times. The ellipse which is estimated from the F data (sampled from the least mean square differences in LMedS) is regarded as the final approximated ellipse. The long axis and short axis of this ellipse are denoted ρ_n and ρ_n , respectively. Let the angle of the long axis be ψ . The set of angles $(\theta_1, \psi_1) \dots (\theta_F, \psi_F)$ is recorded.

(8) λ_1 and λ_2 are given by

$$\lambda_1 = \frac{\rho_{I1}}{r_T}, \quad \lambda_2 = \frac{\rho_{I2}}{r_T} \quad (8)$$

where r_T is the radius of the circle which corresponds to the approximated ellipse from I on T .

(9) The rotation angle of the template is calculated by using $(\theta_1, \psi_1) \dots (\theta_F, \psi_F)$.

$$\theta = \frac{1}{F} \sum_{n=1}^F \left(\arctan \left(\frac{\lambda_1 \sin(\psi_n - \psi)}{\lambda_2 \cos(\psi_n - \psi)} \right) - \theta_n \right) \quad (9)$$

Let us next examine the relationship between the accuracy and the line drawn on T_L . The accuracy is independent of the length of the line, but depends on the resolution of the right end of the line. Therefore, the maximum length of the line is set in this paper. Theoretically, the distance to the left end of the line should be small, since a longer line can avoid occlusions or erroneous double recognition. In practice, however, it is necessary to use a distance to the left end greater than the detection error, since the detection error δ must be considered for matching on I_L . The maximum detection angle η_{max} around the invariant axis is determined by the ratio when the left-end value of the matching line on T_L and the left-end value on I_L differ most. This relationship is represented by

$$\eta_{max} = \arccos \lambda_{max} \quad (10)$$

In this experiment, the length of the line on T_L was chosen as the distance from the 32-nd pixel to the 128-th pixel (both counted from the center of the image). The log-polar image was made from the 8-th pixel to the 128-th pixel, with $\delta = 3$ pixels to provide a safety factor. Therefore, if matching occurs at the leftmost position on I_L , the left end is on the 8-th pixel and the right end is on the 32-th pixel, and $\eta_{max} = 75.5^\circ$.

7. Experiments and Discussion

The proposed algorithm was tested experimentally using real images. Figure 11 shows T formed by imaging a piece of packaging with a digital camera (Fujifilm FinePixF402). Figure 12 shows the input image I which was made by imaging the same wrapping paper from an arbitrary direction. Both T and I have a size of 256×256 pixels. An Athlon 1.6 GHz processor was used for the PC.

Fig. 11. Template image for experiment.

Fig. 12. Input image for experiment.

Figures 13 and 14 show log-polar transformed images of the above images, with matching lines for detection. The white lines on T_L and I_L are correlated, for example, the four white lines indicated by "1" in Figs. 13 to 16. The white line used for the correlation is matched with the waveform in Fig. 14. Comparing Figs. 16 and 15 (in both images, the inversely transformed white lines appear) successful steady matching with the affine transformation is evident, since the white lines and the approximated circle and ellipse are drawn in approximately the same position.

Figure 17 shows a template image modified in accordance with the parameters detected by the proposed algorithm. Figure 18 shows an image made by superimposing the image in Fig. 17 (with semitransparency) on the input image. There is no double image in the area where the two images are superimposed, indicating almost perfect agreement. The fact that the arcs and letters along the

Fig. 13. Template image transformed by log-polar.

Fig. 14. Input image transformed by log-polar.

Fig. 15. Template image and detected ring.

Fig. 16. Input image and detected oval.

Fig. 17. The image strained template with detection parameter.

boundary of the template image and the input image are smooth validates the detected affine parameters.

Thus, the experiments were successful.

7.1. Detection of flat textures

The texture of the template in the previous experiments is evenly distributed on the whole area, so that pattern matching is easier and success is achieved without errors.

However, the experiment described in this section is expected to withstand even erroneous matching due to such factors as flat texture or occlusions, since LMedS estimation (a robust approximation method) is applied. In this experiment, data points which match the detection ellipse within an error of $\pm\delta$, taking n data-points, are regarded as correct (without error). The ratio of the number of correct data points to the total number of data points is defined as the correct matching rate.

Fig. 18. Superposition image.

Fig. 19. Example of flat texture—Template.

Fig. 21. Example of flat texture—Superposition image.

8. Conclusions

This paper proposes an extension of the log-polar transformation, which in its original form can detect only stable features of scaling and rotation angles, so that affine transformations can also be detected. This paper confirms experimentally that the long and short axes and rotation angle of an affine-transformed ellipse can be detected by using the fact that a pattern on each angle line undergoes no change. Only four sets of memory are required for this algorithm: T , I , and the log-polar transformation of each image. Test experiments were successfully carried out using this algorithm on real images taken by a digital camera. The algorithm can perform template matching which is invariant in the affine transformation, in addition to the detections of scaling and rotation angles, provided that the log-polar transform can be applied. This is due to the fact that this algorithm can be used under the same conditions as the log-polar transform.

The recognition of an image which is invariant to affine transformation allows recognition of a 2D image set in a 3D environment. If a 2D image is on a plane of a 3D object, it is possible to detect its position and posture with a single-eye camera by the proposed algorithm.

Complex texture on a template significantly affects the matching of white lines; the more complex the texture, the better the results. For example, if an image has fine texture over its area, all the white lines match at close to the correct positions. Therefore, if the proposed algorithm is applied to matching of an unchanged affine transformation, the complexity of the texture used determines its success. The user can freely choose the size and position of a template in an object, even if the object or texture is chosen beforehand. Therefore, it is possible to designate a part of an object which has a complex texture and is most suitable for detection as a template. Complexity of texture does not create any problems in matching.

Fig. 20. Example of flat texture—Detecting result.

Detection of feature points involves some problems. Since the log-polar transformation is not robust to changes of position, a corresponding point must be found beforehand. The construction of a single-eye object-detection system for 3D environment by solving this problem is a future objective of this project.

REFERENCES

1. Swain MJ, Stricker MA. Promising directions in active vision. *Int J Comput Vision* 1993;11:109–126.
2. Xu G, Tsuji S. Three-dimensional vision. Kyoritsu Shuppan; 1998.
3. Miyatake T, Matsuyama T, Nagao M. Affine transform invariant curve recognition using Fourier descriptors. *Trans Inf Process Soc Japan* 1983;24:64–71.
4. Bebis G, Georgiopoulos M, da Vitoria Lobo N, Shah M. Learning affine transformations. *Pattern Recognition* 1999;32:1783–1799.
5. Lamdan Y, Schwartz JT, Wolfson HJ. Affine invariant model-based object recognition. *IEEE Trans Robot Autom* 1990;6:578–589.
6. Kimura A, Watanabe T. Generalized Hough transform to be extended as an affine-invariant detector of arbitrary shapes. *Trans IEICE* 2001;J48-D-II:789–798.
7. Bao Y, Nakata T, Akasaka H. Perspective projection invariant pattern recognition without gravity center calculation. *Trans Inst Image Electron Eng Japan* 2003;32:659–666.
8. Jezekiel B-A, Wang Z. Pictorial recognition of objects employing affine invariance in the frequency domain. *IEEE Trans Pattern Anal Mach Intell* 1998;20:604–618.
9. Kurita T, Hotta K, Mishima T. Scale invariant recognition of face image using high-order local autocorrelation of log-polar images. *Trans IEICE* 1997;J80-D-II:2209–2217.
10. Massone L, Sandini G, Tagliasco V. Form-invariant: Topological mapping strategy for 2D shape recognition. *Comput Vis Graph Image Process* 1985;30:169–188.
11. Rousseeuw RJ. Least median of squares regression. *J Am Stat Assoc* 1984;79:871–880.

AUTHORS (from left to right)

Takayuki Nakata received his B.E., M.E., and D.Eng. degrees from Kanazawa University in 1998, 2001, and 2004. From 2002 to 2004 he was a special research student at Yokohama National University. Since 2004, he has been a research associate in the Faculty of Engineering, Toyama Prefectural University. His research interests include recognition of 3D object, 3D display, and optical networking.

Yue Bao received his Ph.D. degree from Kanazawa University in 1996 and became an assistant professor there. From 1998 to 2004, he served as a lecturer on the Faculty of Engineering at Yokohama National University and then an associate professor at the Graduate School of Social Environment and Information. Since 2004, he has been a professor on the Faculty of Knowledge Engineering and Graduate School of Engineering Science, Musashi Institute of Technology. His research interests involve image processing, 3D image display, laser-applied measurements, and autonomous moving robots.

Naofumi Fujiwara graduated from the Department of Aeronautical Engineering at Osaka Prefectural University in 1965 and became a research associate there, subsequently being appointed a lecturer and an associate professor. He moved to Kanazawa University in 1989 as a professor. He is engaged in research on vehicle automation which concerns positioning of mobile objects using lasers and machine vision, and automated vehicle guidance.