Gradient descent approach for minimizing dissimilarity measure in log-polarimagery *

JIN Yong-jun ** **and JIANG You-yi**
College of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China

(Received 17 December 2005)

Log-polar mapping has been proposed as a very appropriate space-variant imaging model in active vision applications. There is no doubt about the importance of translation estimation in active visual tracking. In this paper an approach is presented, and its performances are evaluated. The approach uses a gradient descent for minimizing a dissimilarity measure. The experimental results reveal that this method is efficient for approaching active image translations.

CLC number: TP751 **Document code**: A **Article ID**: 1673-1905(2006)04-0302-03

The information received by the visual cortex is the result of a log-polar conformal mapping of the retinal stimulus[1]. Indeed, the log-polar transformation has been adopted in numerous applications and is the subject of current research interest[2]. Log-polar images offer a good method with reasonably high resolution, and significant date reduction in large visual field. Moreover, it is scale-and-rotation invariant in pattern recognition, which is a useful feature. However the log-polar mapping is a nonlinear discrete transform.

To keep the advantages of the log-polar transformation and still be able to estimate linear features, translation estimation to log-polar images is necessary[3-8].

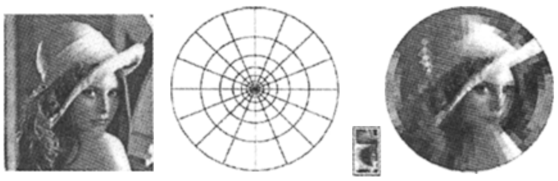
Oshiro *et al*[9] used log-polar images for binocular tracking. In the works, the target is segmented from the background by means of a zero disparity filter, and the centroid of the resulting target region is used as the tracking error signal. In contrast, we focus our work on monocular log-polar images, make no use of optical flow, assume no model of the target, and restrict to a translational motion model. Additionally, the methods presented here are conceptually simple, and one of them has the important advantage of being computationally very efficient and able to estimate considerably large target translations. This makes the approach very suitable for real-time applications, for example, traffic scenarios, video-conference, etc.

A log-polar mapping commonly used in literature defines the log-polar coordinates

$$(\zeta,\eta)=(\log_a(\frac{\rho}{\rho_0}),q\cdot\theta) \qquad (1)$$

where (ρ,θ) are the polar coordinates defined from the cartesian coordinates (x,y) as usual , i.e. $(\rho,\theta)=(\sqrt{x^2+y^2},\arctan(y/x))$. Because of the discretization, the continuous coordinates (ζ,η) become the discrete ones $(u,v)=(|\zeta|,|\theta|)$, $0\leq u<R$, $0\leq v<S$, with R and S being the number of rings and sectors of the discrete log-polar image, respectively, and $q=S/2\pi$ sectors/radian being the angular resolution. Having chosen R, ρ_0 (the radius of the innermost ring), and ρ_{max} (the radius of the visual field), the transformation parameter a is computed as $a=\exp(\ln(\rho_{max}/\rho_0)/R)$.

Given the parameters of the cartesian and log-polar geometries, the log-polar transformation builds the map λ, where $\lambda(i,j)$ is the set of log-polar pixels (u,v) intersecting the cartesian pixel (i,j). If the original cartesian image is sized $M\times N$, ρ_{max} is defined as $\rho_{max}=\frac{1}{2}\min(M,N)$ and the log-polar transform is centered at the foveate point: $(x_c,y_c)=(M/2,N/2)$. An example of a log-polar transformation is shown in Fig. 1.



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

Fig. 1 An example of a log-polar transformation: (a) original cartesian image (256 · 256); (b) log-polar grid (10 · 16), with receptive fields shown; (c) cortical image (32 · 64) computed from (a); (d) cartesian image (256 · 256) reconstructed from (c) by the inverse log-polar mapping

A translation in the x and y coordinate axis, (x_0,y_0), can be expressed in the cartesian plane as simply as the linear equation:

* Supported by National Natural Science Foundation of China (Grant No. 60575013).

** E-mail: jianggangan@nwpu. edu. cn

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \tag{2}$$

In the complex logarithmic plane, however, this simplicity disappears, and the motion model would read as follows:

$$\begin{pmatrix} \zeta' \\ \eta' \end{pmatrix} = \begin{cases} \log_a\left(\dfrac{\sqrt{(\rho_0 a^\zeta \cos(\eta/q) + x_0)^2 + (\rho_0 a^\zeta \sin(\eta/q) + y_0)^2}}{\rho_0}\right) \\ q\arctan\left(\dfrac{\rho_0 a^\zeta \sin(\eta/q) + y_0}{\rho_0 a^\zeta \cos(\eta/q) + x_0}\right) \end{cases} \tag{3}$$

Eq. (3) is obtained by expressing the cartesian coordinates ($x = x(\zeta, \eta)$, $y = y(\zeta, \eta)$) corresponding to the log-polar coordinates (ζ, η), then adding the motion displacement (x_0, y_0), and finally converting the result $(x + x_0, y + y_0)$ back to log-polar coordinates.

The correlation index λ for the vergence angle Ψ in a certain range, gives rise to a one-dimensional (1D) function, $\lambda(\Psi)$[3]. Our approach extends this idea to a 2D function (the correlation surface), which is dependent of the translational components in x and y direction with x_0 and y_0 expressed by $\lambda(x_0, y_0)$. We can compute the value between one image and version shifted by (x_0, y_0) with the correlation measure λ. Fig. 2 is the illustration of the cortical image deformation under cartesian translation Log polar images. Fig. 3 shows an example of the resulting surface $\lambda(x_0, y_0)$ computed in cartesian (Fig. 4 (a)) and log-polar (Fig. 4(b)) domains for the same images. Algorithm: Gradient-descent-based translation estimation in log-polar images.

Input: Two log-polar images, I_1 and I_2

Output: The estimated translation vector $(\hat{x}_0, \hat{y}_0)$

```
(x_0^0, y_0^0) ← (0,0)   {initial guess}
k ← 0   {iteration number}
δ ← 1   {step length}
while (δ > δ_min) Λ (k < k_max) do
k ← k+1
(x_0^k, y_0^k) ← (x_0^{k-1}, y_0^{k-1}) − g(∇λ)   {estimation update rule}
if minimum surpassed then
δ ← δ/2
end if
end while
(x̂_0, ŷ_0) ← (x_0^k, y_0^k)
```

As we have shown, that the minimum of the correlation surface computed on log-polar images occurs (approximately) at the correct displacement of the target. To estimate the translation parameters (x_0, y_0), our approach consists of finding the location of the minimum of the correlation measure $\lambda(x_0, y_0)$. The shape of the correlation surfaces suggests that a gradient-based search could be an adequate search algorithm. The estimation at iteration k, (x_0^k, y_0^k), is updated from the estimation at the previous iteration (x_0^{k-1}, y_0^{k-1}), by using the gradient, $\nabla\lambda$, of the correlation measure, as the most promising direction to move, i. e:

$$(x_0^k, y_0^k) = (x_0^{k-1}, y_0^{k-1}) - g(\nabla\lambda) \tag{4}$$

A common definition is $g(\nabla\lambda) = \delta \cdot \nabla\lambda / z.sfnc$, which the movement of the unit vector gradient $\nabla\lambda|$ in the opposite direction with a certain amount δ is considered. The issue of choosing the value for δ usually involves a trade-off. Then, an adaptive, rather than a fixed step, is called for. One possibility consists of moving 'large' steps when it is far from the minimum, and 'small' steps when it is in the contrary. This is our idea: initially $\delta = 1$ and whenever the minimum is surpassed the value of δ is halved. The search may be stopped using some criteria such as that δ is smaller than a given threshold $\delta_{\min}$, or that the search has reached a maximum number of iterations $k_{\max}$.

To evaluate Eq. (4) we need a way to compute the gradient $\nabla\lambda$. Let I_1 and I_2 be two log-polar images.

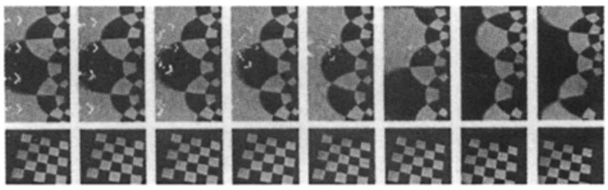


Fig. 2 illustration of the cortical image deformation under cartesian translation Log polar images in the first row, cartesian images are in the second row (shown much more smaller than they are). Columns 2-8 are versions of the original image (in first column) shifted by ($x_0 = s$, $y_0 = -S$), with S equal to 1, 3, 5, 10, 20, 30, and 40, respectively

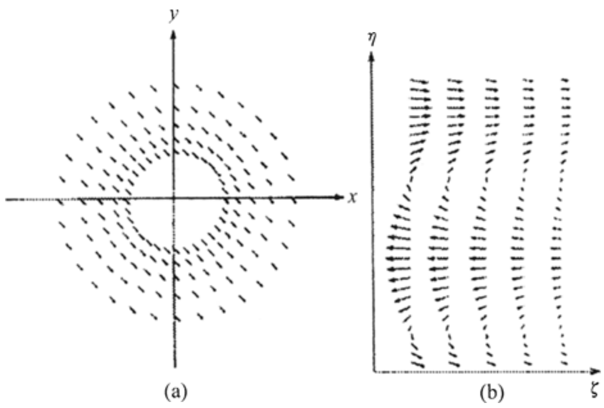


Fig. 3 Translation vectors in the (a) retinal and (b) cortical planes

In the case of the well-known SSD (sum of squared differences) correlation measure:

$$\lambda(x_0, y_0) = \sum_{(\zeta,\eta)\in\lambda} (I_2(\zeta', \eta') - I_1(\zeta, \eta))^2$$

the gradient $\nabla\lambda=(\lambda_{x_0},\lambda_{y_0})$ becomes:

$$\begin{pmatrix}\lambda_{x_0}\\ \lambda_{y_0}\end{pmatrix}=\begin{bmatrix}\dfrac{\partial\lambda}{\partial x_0}\\ \dfrac{\partial\lambda}{\partial y_0}\end{bmatrix}=2\begin{bmatrix}\sum_{(\zeta,\eta)\in\lambda}\{(I_2(\zeta',\eta')-I_1(\zeta,\eta))\cdot I_{2x_0}(\zeta',\eta')\}\\ \sum_{(\zeta,\eta)\in\lambda}\{(I_2(\zeta',\eta')-I_1(\zeta,\eta))\cdot I_{2y_0}(\zeta',\eta')\}\end{bmatrix}\quad(5)$$

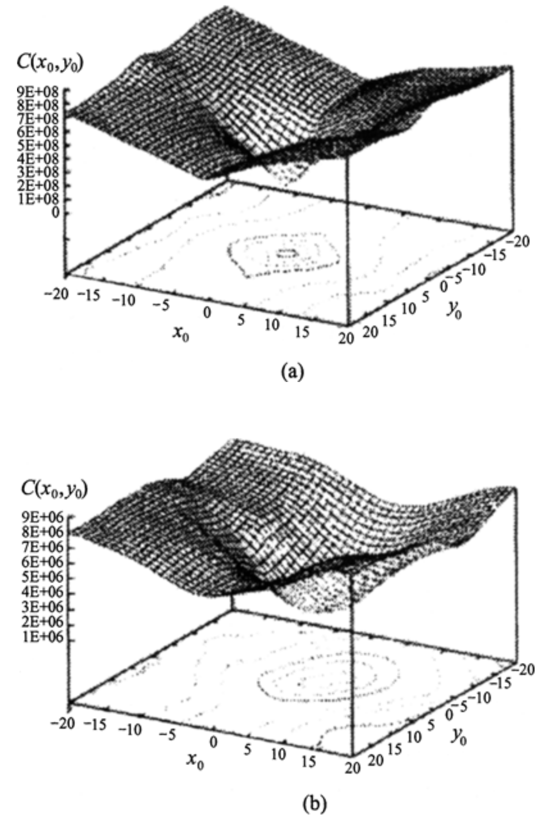


Fig. 4 Example of correlation surface in the (a) cartesian and (b) log-polar spaces

with λ being a certain set of image pixels (usually, the entire image), and where $I_{2x_0}=I_{2x_0}(\zeta',\eta')$ and $I_{2y_0}=I_{2y_0}(\zeta',\eta')$ are:

$$\begin{pmatrix}I_{2x_0}\\ I_{2y_0}\end{pmatrix}=\begin{pmatrix}\zeta'_{x_0} & \eta'_{x_0}\\ \zeta'_{y_0} & \eta'_{y_0}\end{pmatrix}\cdot\begin{pmatrix}I_{2\zeta'}\\ I_{2\eta'}\end{pmatrix}\quad(6)$$

The common notation for partial derivatives, $f_z=\partial f/\partial z$, is used. On the other hand, (ζ',η') depends on the space-variant way of the translational displacement (x_0, y_0). For the sake of simplicity, we use the following approximation, rather than the exact expression in Eq. (3):

$$\begin{pmatrix}\zeta'\\ \eta'\end{pmatrix}\approx\begin{pmatrix}\zeta\\ \eta\end{pmatrix}+\begin{pmatrix}\zeta_x & \zeta_y\\ \eta_x & \eta_y\end{pmatrix}\cdot\begin{pmatrix}x_0\\ y_0\end{pmatrix}\quad(7)$$

Deriving ζ' and η', as defined in Eq. (7), with respect to x_0 and y_0 yields: $\zeta'_{x_0}=\zeta_x$, $\zeta'_{y_0}=\zeta_y$, $\eta'_{x_0}=\eta_x$, $\eta'_{y_0}=\eta_y$ which are then used in Eq. (6).

Finally, by taking the partial derivatives of ζ and η, as defined in Eq. (1), with respect to x and y, we get ζ_x, ζ_y, η_x and η_y as follows:

$$\begin{pmatrix}\zeta_x & \zeta_y\\ \eta_x & \eta_y\end{pmatrix}=\begin{bmatrix}\dfrac{\partial\zeta}{\partial x} & \dfrac{\partial\zeta}{\partial y}\\ \dfrac{\partial\eta}{\partial x} & \dfrac{\partial\eta}{\partial y}\end{bmatrix}=\frac{1}{\rho}\begin{bmatrix}\dfrac{\cos\theta}{\ln a} & \dfrac{\sin\theta}{\ln a}\\ -\sin\theta & \cos\theta\end{bmatrix}\quad(8)$$

A brief presentation of the log-polar mapping has been given. Then, the importance of translation estimation and its difficulty in cortical images has been shown. Next, an approach (GDS) that effectively deals with translation estimation in the complex logarithmic domain has been introduced.

This strategy has been proven to be efficient. On the other hand, carrying out some off-line computations has been shown to be crucial in order to run these algorithms in real time. It makes sense in active visual tracking applications.

References

[1] E. L. Schwartz, *Biological Cybernetics*, **25**(1977), 181.

[2] N. Oshiro, N. Maru, A. Nishikawa, and F. Miyazaki, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Osaka Japan, **2**(1996), 791.

[3] A. Bernardino and J. Santos-Victor, *Robotics and Autonomous Systems*, **25**(1998), 137.

[4] C. F. Weiman and R. D. Juday, *SPIE*, 1989, 843.

[5] H. Tunley and D. Young, *Dynamic Fixation of a Moving Surface using Log Polar Sampling*, *British Machine Vision Conference*, vol. 2, BMVA Press, 1994, 579.

[6] F. Panerai, C. Capurro, and G. Sandini, *Space variant vision for an active camera mount*, *Technical Report TR I/95*, *LIRA*, *DIST*, University of Genova, Italy, February 1995.

[7] V. J. Traver and F. Pla, *Spanish Symposium on Pattern Recognition and Image Analysis* (*SNRFAI*), Castelló n, Spain, **2** (2001), 281.

[8] I. Ahrns and H. Neumann, *Proceedings of Artificial Intelligence*, 1998, 89.

[9] N. Oshiro, N. Maru, A. Nishikawa, and F. Miyazaki, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Osaka Japan 2(1996), 791.