Translation Estimation in Log-Polar Images*

V. J. Traver, F. Pla
{vtraver, pla}@uji.es
Grup de Visió per Ordinador, Universitat Jaume I
E12071 Castellón (Spain)

Abstract

Log-polar mapping plays a fundamental role in active vision by allowing both a wide field-of-view and a high-accuity area. While log-polar mapping simplifies certain visual tasks, translation estimation becomes more difficult in this domain. However, given the importance of translational motion for mobile target tracking, solving this problem in log-polar space is of great significance. In this paper, we present some preliminary results of translational motion estimation using log-polar images.

Keywords: log-polar mapping, active vision, motion estimation, correlation.

1 Introduction

Visual target tracking has interesting real-world applications, such as visual surveillance, automatic video recording or video conferencing. However, the conventional uniform-resolution images impose a significant computational burden and can, therefore, jeopardize the real-time performance required in these applications. To overcome these limitations, space-variant imaging has been advocated as an adequate solution, which is able to simultaneously have a wide field-of-view and a high-resolution area (the *fovea*), by having a decreasing resolution towards the periphery [2].

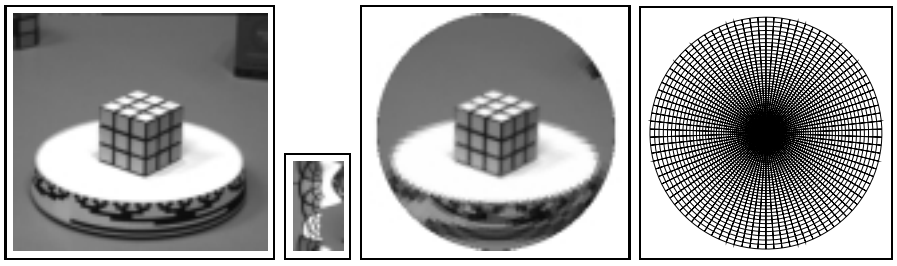
In space-variant active vision [6], one can detect peripheral activity only at a coarse detail. For a finer escrutinity of the potential object of interest, this will have to be observed with the highest resolution possible, which, in turn, requires a camera motion, to have the target centered. After the target *foveation*, it has to be actively tracked. To that end, the motion of the target must be estimated so that the proper command motion is issued for having the target centered again.

This work has been partially supported by projects GV97-TI-05-27 from the *Conselleria d'Educació, Cultura i Ciència, Generalitat Valenciana*, and CICYT TIC98-0677-C02-01 from the *Ministerio de Educación y Cultura*. The first author was partially supported by a predoctoral scholarship from *Conselleria d'Educació, Cultura i Ciència, Generalitat Valenciana*.

Thus, estimation of translation is a key issue in active tracking. Therefore, to achieve an effective object tracking and, at the same time, keep the advantages of log-polar images, the problem of translation estimation in the log-polar domain must necessarily be addressed [1, 4]). This is just the problem we tackle in this paper.

2 Log-polar mapping

In our implementation, the mapping proposed by Jurie [3] is used, which defines the log-polar coordinates $(\xi, \eta) = \log_a \left(\frac{\rho+\rho_0}{\rho_0}), k \cdot \theta\right)$ from the polar coordinates (ρ, θ), where a, ρ_0, and k are parameters of the transformation, that are functions of the chosen number of rings (R) and sectors (S). This mapping (see an example in figure 1) have two distinctive features. On the one hand, it properly addresses the well-known singularity of the logarithmic function at the origin. This prevents having a blind area at the center of the sensor. On the other hand, log-polar pixels are fractions of rectangular pixels. This results in a more flexible and accurate mapping than those forming log-polar pixels as a mere aggregation of cartesian pixels.



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

Figure 1: Example of log-polar transformation: (a) the original 256×240 cartesian image of a *Rubik* cube on a microwave turntable, (b) the *cortical* image after the log-polar mapping, (c) the *retinic* image, obtained by performing the inverse log-polar mapping, and (d) the grid layout for this particular mapping ($R = 50$, $S = 90$).

3 The motion estimation method

Translation in cartesian domain can be formulated quite easily, but it is rather less straightforward in (log-)polar coordinates. Panerai *et al.* [4] pointed out that the projection of a translation vector at p on radial line passing through p, results in a retinal shift $\Delta\rho$ (the magnitude of the projected vector), which is maximum along the direction of motion, and follows a sinusoidal function of the angle of the radial line, being constant for each radial direction.

3.1 The basic method

We can estimate $\Delta\rho(\eta)$ at each discrete radial line η of the log-polar image, by considering the log-polar images before, I_1, and after, I_2, the translation. Then, we have a discrete sinusoidal function, its amplitude and phase will allow us to estimate the vector of motion [4]. One way to proceed is to compute the discrete Fourier transform (DFT) of $\Delta\rho(\eta)$, which is a 1-D signal. The more similar the function $\Delta\rho(\eta)$ to a sinusoidal wave, the better the first harmonic resulting from the DFT will approximate it, and the more accurate the motion magnitude and orientation estimations will be. Let $\mathcal{D}\rho(k_\eta)$ be the DFT of $\Delta\rho(\eta)$. Then, the magnitude of the motion is given by the amplitude of the first Fourier harmonic, i.e., $\alpha_1 \cdot |\mathcal{D}\rho(1)|$, where α_1 depends on what exactly the DFT transform computes, and on the number of data used to compute it (in our case, $\alpha_1 = 2/S$). The motion direction is given by the phase of this first harmonic, i.e., $\arg(\mathcal{D}\rho(1))$.

For the estimation of $\Delta\rho(\eta)$ for each η, some correlation measure can be used. Let $\Delta^*\rho(\eta)$ be the estimation of the true $\Delta\rho(\eta)$. It can be computed with some 1-D correlation $\mathcal{C}$ along the radial direction at orientation η as:

$$\Delta^*\rho(\eta) = \arg\min_{\Delta\rho} \mathcal{C}\left(I_1(\xi,\eta), I_2(\xi + \Delta\xi(\Delta\rho), \eta)\right)$$

3.2 Variations to the method

In [4], the basic idea of the method is just introduced, and no experimental results are given. We have implemented, not only this basic algorithm, but also several variations aimed at increasing the robustness and accuracy of the basic formulation. We have carried out extensive experimentations in order to assess the behavior of the basic as well as our extensions, which are summarized in this section.

Phase correlation. Regarding the correlation measure $\mathcal{C}$, Panerai *et al.* suggest the use of the well-known sum of squared differences (SSD). An inconvenient with this kind of correlation is that one has to choose the number of $\Delta\rho$ values to be tested, as well as the set of particular values. An interesting alternative we have

tried is the use of phase correlation [5], a well-know method based on the shift theorem of Fourier transform, which estimates $\Delta\rho$ in a given direction η by finding the peak in the inverse Fourier transform of the normalized cross power spectrum,

$$\overline{C}_{1,2}(k_\xi) = \frac{\mathcal{I}_2^\eta(k_\xi)\mathcal{I}_1^{\eta *}(k_\xi)}{|\mathcal{I}_2^\eta(k_\xi)\mathcal{I}_1^{\eta *}(k_\xi)|},$$

where $\mathcal{I}_i^\eta(k_\xi)$ is the Fourier transform of $I_i^\eta(\xi) = I_i(\xi,\eta),\ i = 1,2$.

Notice that, for a correct application of the phase correlation method, we have to sample $I_i^\eta(\xi)$ *linearly*, not logarithmically. Therefore, in the discrete case, $I_1^\eta(\xi_j+\Delta\rho)$ is actually computed as $I_1^\eta(\xi_j + \Delta\xi_j(\Delta\rho))$. To try to minimize the boundary effects and spectral leackage of the DFT, a Gaussian-like weighting is applied [5].

Double sector and supporting sectors. On the other hand, to make the method more robust to noise and resolution effects, we have made experiments correlating not only on a radial sector η, but considering also the sector just in the opposite direction (i.e., η and $\eta + \eta(\theta(\eta) + \pi/2)$, thus forming an "extended" sector of $2R$ pixels. Moreover, when performing the correlation at η, we also consider its neighboring sector $\eta + a$ (tipically, $a \in \{-1, 1\}$). These extensions can be applied with the phase correlation as well.

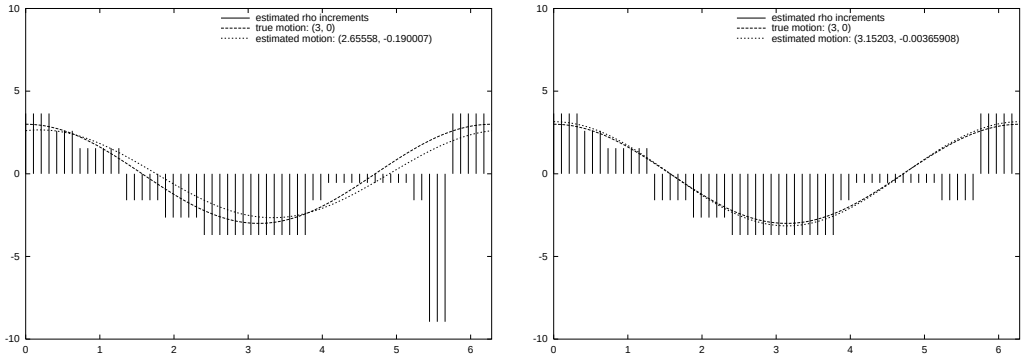
Smoothing. Finally, the discrete nature of the log-polar mapping, causes that $\Delta^*\rho(\eta)$ sometimes differ from the theoretically correct $\Delta\rho(\eta)$. To deal with this, two common filtering techniques have been tested and applied prior to the application of the DFT: the well-known median filtering, with a window width w_M and a Gaussian filter with a window width w_G.

4 Experimental results

To test the method described in previous section and the proposed extensions, we apply a translation to a cartesian image, and the log-polar transformation of both cartesian images (the original and the translated one) are taken as the input for the method. The magnitude ρ_T and direction θ_T of this "motion" will be taken as the ground-truth values. Their units are cartesian pixels and radians, respectively. The method will yield estimates, $\hat{\rho}_T$ and $\hat{\theta}_T$, of these true values.

Figure 2 show the estimation $\Delta^*\rho(\eta)$ for each η. The ground-truth and the estimated motions are also ploted as sinusoidals functions, the amplitude of which is the magnitude of the motion, and the phase of which is the direction of motion. In this case, the *Rubik* cube (figure 1) has undergone a motion ($\rho_T = 3, \theta_T = 0$). The log-polar images used here have 40×60 pixels, i.e., around 4 times less data than the original cartesian images. Figure 2(a) corresponds to the original method,

which results in an estimated motion of $(\hat{\rho}_T = 2.66, \hat{\theta}_T = -0.19)$. This estimation has an error of 11% in magnitude and 11° in orientation. We can appreciate in figure 2(b) that some noisy data is affecting the estimation. Then, after applying a median filtering to $\Delta^*\rho(\eta)$, the motion estimation shown in figure 2(b) becomes $(\hat{\rho}_T = 3.15, \hat{\theta}_T \approx 0.00)$. This time the error is only 5.1% in magnitude and 0.2° in orientation. In figure 2(b) it can be easily seen how the two sinusoidal functions almost overlap.



(a) without smoothing

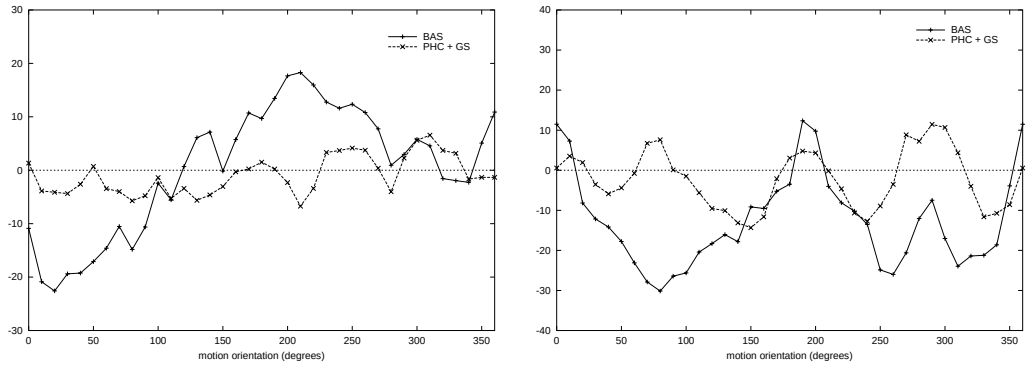
(b) with smoothing

Figure 2: Effect of a median filter ($w_M = 9$) for smoothing $\Delta^*\rho(\eta)$ on motion estimation error.

The estimation error depends on several issues. For instance, to observe how the error varies, we repeated the experiment varying the direction of motion. In figure 3 we plot the estimation errors, comparing the results of the basic method with those of the phase correlation and gaussian smoothing ($w_G = 0.3S$). As can be seen, when employing the phase correlation instead of the SSD, the error of motion estimation is significantly reduced (notice in figure 3(a) how big the maximum error in the basic method is, compared with that resulting from phase correlation). We have carried out some more experiments using a set of about ten images, obtaining results similar to those reported here.

5 Conclusions

Translational motion estimation in log-polar space, although more difficult than in cartesian coordinates, is necessary to keep the interesting advantages that this



(a) Error in orientation ($\theta_T - \hat{\theta}_T$)

(b) Error in magnitude ($\frac{\rho_T - \hat{\rho}_T}{\rho_T} \cdot 100$)

Figure 3: Estimation error for a fixed motion magnitude and varying motion direction (BAS = basic method, PHC = phase correlation, GS = gaussian smoothing).

domain offer for some other visual tasks. We have addressed this problem starting form a basic idea suggested in [4]. Our experiments, reported in this paper, show that this method, although theoretically valid, when tested with real images produces motion estimations that generally tend to be more accurate in direction than in module. We have proposed variants to the basic approach to make it more robust and accurate. The results obtained show they are valid for our purposes: active tracking with log-polar images.

References

[1] I. Ahrns and H. Neumann. Real-time monocular fixation control using the log-polar transformation and a confidence-based similarity measure. In A. K. Jain, S. Venkatesh, and B. C. Lovell, editors, *Intl. Conf. on Pattern Recognition (ICPR)*, pages 310–315, Brisbane, Australia, Aug. 1998.

[2] M. Bolduc and M. D. Levine. A review of biologically motivated space-variant data reduction models for robotic vision. *Computer Vision and Image Understanding (CVIU)*, 69(2):170–184, Feb. 1998.

[3] F. Jurie. A new log-polar mapping for space variant imaging. Application to face detection and tracking. *Pattern Recognition*, 32:865–875, 1999.

[4] F. Panerai, C. Capurro, and G. Sandini. Space variant vision for an active camera mount. Technical Report TR 1/95, LIRA, DIST, Univ. Genova, Italy, Feb. 1995.

[5] F. Pla and M. Bober. Estimating translation/deformation motion through phase correlation. In A. del Bimbo, editor, *Intl. Conf. on Image Analysis and Processing (ICIAP)*, pages 653–660, Florence, Italy, 1997. Springer-Verlag. Lecture Notes in Computer Science.

[6] M. Tistarelli and G. Sandini. Dynamic aspects in active vision. *Computer Vision, Graphics, and Image Processing (CVGIP): Image Understanding*, 56(1):108–129, July 1992.