Computation of 3D-Motion Parameters Using the Log-polar Transform

Konstantinos Daniilidis

Computer Science Institute, Christian-Albrechts University Kiel
Preusserstr. 1-9, 24105 Kiel, Germany
email: kd@informatik.uni-kiel.de

Abstract. Arificial vision systems for mobile robots necessitate sensors and representations that enable a real-time reactive behavior. The log-polar transform has been shown to be a variable resolution scheme that achieves a high compression of the non-foveal part of an image. Such space variant sensors must inevitably be active in order to utilize the high- and homogeneous resolution fovea. We study here the computation of the heading direction using a log-polar sensor able to fixate. The polar nature of the complex logarithmic mapping produces a computationally superior representation of the optical flow. Based on an insight for the translational case we present a new algorithm for computing the focus of expansion by applying fixation in case of general motion.

1 Introduction

In this paper, we study the properties of the motion field using a space variant resolution scheme. We propose algorithms for the recovery of 3D-motion which exploit the advantages of space variant sensing and fixation. Space variant sensing is one of the main features of an attentive vision system. Attention arises as a necessity in building robot systems that are able to react according to the visual stimuli in a dynamically changing environment. Even with the fastest image processing architectures it is not possible to process a uniformly sampled image with a wide field of view in a process time enabling immediate reaction. To overcome the reaction time constraints we need an image data reduction scheme which conserves high resolution in the center of the image - the fovea- and a gradually decreasing resolution area - the periphery. Since most visual motion tasks necessitate a wide field of view we have to elaborate algorithms for computing the optical flow and recovering the 3D-motion and structure using space-variant resolution. In this paper, we do not only elaborate motion algorithms regarding the logarithmic polar plane as a practical necessity. We show that the log-polar mapping has also computational advantages with respect to motion recovery tasks:

- We propose a 1D-search algorithm for the computation of the translation direction in case of pure translation.
- We show that fixation enables the computation of the translation direction in case of general motion.

Hlaváč, Šára (Eds.): CAIP '95 Proceedings, LNCS 970

Pursuit and optokinetic eye movements have long been considered by the psychophysicists as a means to overcome the limited field of view, to bound the binocular disparity, and to eliminate the motion blur [3]. Here, we prove that eye movements are not only an implication of the inhomogeneous retina but they support the recovery of 3D-motion.

We proceed with a survey of approaches to space variant motion sensing. We refer to non-motion related work on space-variant sensing in a companion paper in the same proceedings. Jain et al. [6] were the first who applied the complex logarithmic mapping in image sequence analysis. Given the motion information from other sensors and assuming pure translation they transformed the images into the log-polar plane using the focus of expansion as the center of the transformation. The depths of the viewed points were easily obtained by inspecting only the shifts along one coordinate of the log-polar plane. Tistarelli and Sandini [10] derived the motion equations for the log-polar plane and proposed a method for the computation of the time to collision. Weiman [11] proposed several tracking behaviors based on the log-polar mapping. The following references concern fixation in the cartesian plane. Aloimonos et al. [1] and Bandopadhay and Ballard [2] showed how the structure from motion problem is simplified in case of active tracking. Fermüller and Aloimonos [4] proved that normal flow measurements are sufficient for bounding the locus of the focus of expansion and the time to collision. Raviv and Herman [8] studied the geometric properties of fixation and derived the locus of 3D-points whose projections yield a zero-optical flow in case of motion with three degrees of freedom (road following). Taalebinezhaad [9] proposed a method that simulates tracking by a pixel shifting process without actively controlling the mechanical degrees of freedom of the camera. The translation direction is then obtained by minimizing the depth deviation from the fixation point using only normal flow measurements. A similar method to ours was proposed in [5] for the derotation of the cartesian motion field.

2 The log-polar motion field

We use (x, y) for the cartesian coordinates and (ρ, η) for the polar coordinates in the plane. We apply the log-polar mapping on the non-foveal part of a retinal image. Therefore, we define as the domain of the mapping the ring-shaped area $\rho_0 < \rho < \rho_{max}$ where ρ_0 and ρ_{max} are the radius of the fovea and the half-size of the retinal image, respectively. Furthermore, a hardware CCD-sensor with the log-polar property or a software implementation of the mapping needs a discretization of the log-polar plane. By assuming that N_r is the number of cells in the radial direction and N_a is the number of cells in the angular direction the mapping from the polar coordinates (ρ, η) to the log-polar coordinates (ξ, γ) reads as follows (see also [10])

$$\xi = \log_a(\frac{\rho}{\rho_0}) \qquad \gamma = \frac{N_a}{2\pi}\eta \quad \text{with} \quad a = e^{\frac{1}{N_r} ln(\frac{\rho_{max}}{\rho_0})}.$$

From now on we will use only η ranging from 0 to 2π.

We proceed with a brief summary of the cartesian motion equations. Let an object be moving with translational velocity $\boldsymbol{v} = (v_x, v_y, v_z)^T$ and angular velocity $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)^T$ relative to the camera. We denote by $\boldsymbol{X}$ the position of a point on the object with respect to the camera coordinate system, by $\hat{\boldsymbol{z}}$ the unit-vector in the z-axis taken as the optical axis, and by $\boldsymbol{x} = (x, y, 1)^T$ the projection of $\boldsymbol{X}$ on the image plane $Z = 1$. The motion field vector reads

$$\dot{\boldsymbol{x}} = \frac{1}{\hat{\boldsymbol{z}}^T \boldsymbol{X}} \hat{\boldsymbol{z}} \times (\boldsymbol{v} \times \boldsymbol{x}) + \hat{\boldsymbol{z}} \times (\boldsymbol{x} \times (\boldsymbol{x} \times \boldsymbol{\omega})). \tag{1}$$

In case of ego-motion of the camera in a stationary environment the above equation as well as all following equations have to be read with the opposite sign for $\boldsymbol{v}$ and $\boldsymbol{\omega}$. The 3D-motion estimation problem is known as the recovery of all the depths, the direction of translation $\boldsymbol{v}$, and the angular velocity from the motion field $\dot{\boldsymbol{x}}$. The magnitude of translation cannot be recovered due to the scale ambiguity that couples translation and depth.

We first compute the motion field vectors in the polar plane. The definition of the polar coordinates yields

$$\dot{\rho} = \dot{x} \cos\eta + \dot{y} \sin\eta \qquad \dot{\eta} = \frac{1}{\rho}(-\dot{x} \sin\eta + \dot{y} \cos\eta). \tag{2}$$

The radial component of the log-polar motion field can be easily obtained:

$$\dot{\xi} = \frac{1}{\ln a} \frac{\dot{\rho}}{\rho}. \tag{3}$$

In order to make the equations for the log-polar plane more readable we introduce the polar unit-vectors $\hat{\boldsymbol{\rho}} = (\cos\eta, \sin\eta)^T$ and $\hat{\boldsymbol{\eta}} = (-\sin\eta, \cos\eta)^T$. Furthermore we introduce the vectors $\boldsymbol{v}_{xy} = (v_x, v_y)^T$ and $\boldsymbol{\omega}_{xy} = (\omega_x, \omega_y)^T$ to describe motion parallel to the image plane. We will carry out the computations for the polar motion field $(\dot{\rho}, \dot{\eta})$. The log-polar motion field is different only in the radial component $\dot{\xi}$ which can be computed from $\dot{\rho}$ straightforward (3). This will also enable us to find out that most of the advantages are due to the polar nature of log-polar plane.

Using the cartesian motion field (1) and the transformation rules (2) we obtain the following expressions which relate the polar motion field to the 3D geometry (depths Z) and motion $(\boldsymbol{v}, \boldsymbol{\omega})$ of the scene:

$$\begin{aligned} \dot{\rho} &= -\frac{\rho v_z}{Z} + \frac{\boldsymbol{v}_{xy}^T \hat{\boldsymbol{\rho}}}{Z} + (1 + \rho^2) \boldsymbol{\omega}_{xy}^T \hat{\boldsymbol{\eta}} \\ \dot{\eta} &= \frac{\boldsymbol{v}_{xy}^T \hat{\boldsymbol{\eta}}}{\rho Z} - \frac{\boldsymbol{\omega}_{xy}^T \hat{\boldsymbol{\rho}}}{\rho} + \omega_z. \end{aligned} \tag{4}$$

Hence, the log-component of the motion field reads

$$\ln a \, \dot{\xi} = -\frac{v_z}{Z} + \frac{\boldsymbol{v}_{xy}^T \hat{\boldsymbol{\rho}}}{\rho Z} + (\frac{1}{\rho} + \rho) \, \boldsymbol{\omega}_{xy}^T \hat{\boldsymbol{\eta}}. \tag{5}$$

We see that the motion field is the sum of the translational part including the depth information and a rotational part like in the cartesian formulation. However, we note that the influence of the motion-components parallel to the optical axis are decoupled. The translation v_z appears only in $\dot{\rho}$ and the rotation ω_z only in $\dot{\eta}$. This was expected due to the properties of the complex logarithmic mapping regarding two-dimensional expansions and rotations, respectively. However, as it is already known the main problem in recovery of 3D-motion is to totally decouple the rotational from the translational effects.

2.1 Pure translation

If there is only translation parallel to the optical axis the polar (and the log-polar) motion field has only one component. The magnitude variation of the log-component $\dot{\xi}$ in this case depends only on the depths of the projected 3D structures. Closer inspection of the equations of the polar translational motion field

$$\dot{\rho} = -\frac{\rho v_z}{Z} \qquad \dot{\eta} = \frac{\boldsymbol{v}_{xy}^T \hat{\boldsymbol{\eta}}}{\rho Z} \tag{6}$$

yields following facts:

1. The angular component reads $\dot{\eta} = \frac{\boldsymbol{v}_{xy}^T \hat{\boldsymbol{\eta}}}{\rho Z}$. There are two lines in the motion field $\eta = \phi$ and $\eta = \phi + \pi$ (or $\phi - \pi$ depending on which of both is in $[0, 2\pi)$) where $\dot{\eta}$ vanishes for every ρ. The angle ϕ gives the direction of (v_x, v_y) or the line where the focus of expansion lies. Assuming that there is no direction η where the integral $\int_{\rho_0}^{\rho_{max}} \frac{1}{\rho^2 Z^2} d\rho$ vanishes the desired directions ϕ and $\phi \pm \pi$ are given by the global minima of $\int \dot{\eta}^2 d\rho$ in the presence of noisy measurements. The above sufficient condition is satisfied if the camera does not gaze on a point of infinite depth. For example, the condition is not satisfied if the origin of the cartesian image lies above the horizon's projection. If the integral $\int \dot{\eta}^2 d\rho$ vanishes everywhere we imply the existence of a pure v_z-translation.

2. Along the lines $\eta = \phi + \pi/2$ and $\eta = \phi - \pi/2$ or $(\phi + 3\pi/2)$ the flow reads

$$\dot{\rho}_{\phi+\pi/2} = -\rho \frac{v_z}{Z} \qquad \dot{\eta}_{\phi+\pi/2} = \pm \frac{\sqrt{v_x^2 + v_y^2}}{\rho Z}.$$

By dividing we obtain

$$\frac{\dot{\rho}_{\phi+\pi/2}}{\rho^2 \dot{\eta}_{\phi+\pi/2}} = \pm \frac{v_z}{\sqrt{v_x^2 + v_y^2}} \tag{7}$$

which is the tangent of the polar angle of the translation direction.

Hence, it is possible by an explicit search for the global minima along the angle coordinate - which is feasible in the polar plane due to the low resolution - to obtain the full translation direction.

2.2 Fixation in the log-polar plane

We assume that the camera mount has two controllable degrees of freedom which enable a rotation about an axis parallel to the image plane through the optical center. We denote by $(\Omega_x, \Omega_y, 0)$ the resulting additional angular velocity. The motion field arising from the relative motion of the camera to the environment (ego and/or object motion) follows from (1)

$$\dot{x} = \frac{v_x - xv_z}{Z} - xy(\omega_x - \Omega_x) + (1 + x^2)(\omega_y - \Omega_y) - y\omega_z$$
$$\dot{y} = \frac{v_y - yv_z}{Z} - (1 + y^2)(\omega_x - \Omega_x) + xy(\omega_y - \Omega_y) + x\omega_z. \quad (8)$$

We define fixation as keeping the same point in the image center along time. The fixated point may be a point on a moving object in case of a stationary or moving observer or a stationary point in the case of a moving observer. Fixation is achieved by closed loop control where controllables are the pan and the tilt angle of the camera. The formally defined fixation criterion we use is the vanishing of the motion field vector at the central point of the image: $\dot{\boldsymbol{x}}|_{\boldsymbol{x}=0} = \boldsymbol{0}$. This is equivalent to

$$\frac{v_x}{Z_o} + \omega_y - \Omega_y = 0 \quad \text{and} \quad \frac{v_y}{Z_o} - (\omega_x - \Omega_x) = 0, \quad (9)$$

where Z_o is the depth of the point projected on the image center. What kind of control strategy is applied to achieve the fixation criterion is not the subject of this paper. We will assume that the system is in steady state mode and we will study the effects of the satisfied fixation criterion in the peripheral motion-field of the log-polar plane. We turn now to the motion field on the polar transform of the periphery which reads in case of fixation after rewriting using (9) and introducing $v'_z = v_z/Z_o$ and $\boldsymbol{v}'_{xy} = \boldsymbol{v}_{xy}/Z_o$

$$\dot{\rho} = -\frac{Z_o}{Z}\rho v'_z + (\frac{Z_o}{Z} - (1 + \rho^2))\boldsymbol{v}'^T_{xy}\hat{\rho}$$
$$\dot{\eta} = \frac{1}{\rho}(\frac{Z_o}{Z} - 1)\boldsymbol{v}'^T_{xy}\hat{\eta} + \omega_z. \quad (10)$$

To reduce the number of symbols we will use again non-primed symbols for the translation. We first note that we obtain as independent unknowns the depths $\frac{Z_o}{Z}$ relative to the depth of the fixation point which results in an object centered scene representation. Second, the other unknowns are reduced from initially five in the non-fixation case - three for rotation and two for the translation direction - to four: the three scaled translation components, and one angular velocity component. This dimension reduction was already proved in the cartesian plane in [1; 2].

We next propose an algorithm to recover the motion parameters in case of fixation from the motion field in the log-polar plane. Equivalent to the case of pure translation without fixation our method is based on the observation that the angular component $\dot{\eta}$ equals ω_z everywhere along the line $\eta = \phi$ where $\tan\phi = v_y/v_x$. However, this is not a necessary condition. The angular component $\dot{\eta}$ equals ω_z in the additional case that $Z = Z_0$ along a radial line, or more general $\dot{\eta}$ is constant if the relative depth is linearly varying: $\frac{Z_o}{Z} - 1 = k\rho$. This may happen if the environment is planar. For a particular η_0 the radial component for $\eta = \eta_0$ reads in this case

$$\dot{\rho} = -(1 + k\rho)v_z + (k\rho - \rho^2)\boldsymbol{v}^T_{xy}\hat{\rho}_{\eta_0}.$$

We can exclude this case if we test $\dot{\rho}$ subject to quadratic variation variation with respect to ρ. The algorithm we propose comprises following steps:

1. We build the average $\overline{\dot{\eta}}$ over ρ for every η and the variance $\int_{\rho}(\dot{\eta}-\overline{\dot{\eta}})^2 d\rho$. The variance vanishes at $\eta = \phi$ and $\eta = \phi \pm \pi$ and at the angles where $Z = Z_0$ on an entire radial line which is excluded as described above.

We carry out an one-dimensional search for the global minima of the above integral. The global minima give the direction of the line containing the FOE in the cartesian plane as well as ω_z (equal the average $\overline{\dot{\eta}}$ at $\eta = \phi$). We exclude additional global minima due to linear depth variation as described above.

2. At $\eta = \phi + \frac{\pi}{2}$ and $\eta = \phi - \frac{\pi}{2}$ (or $\phi + \frac{3\pi}{2}$) the motion field reads:

$$\dot{\rho} = -\frac{Z_o}{Z}\rho v_z \qquad \dot{\eta} = \frac{1}{\rho}(\frac{Z_o}{Z} - 1)\sqrt{v_x^2 + v_y^2} + \omega_z.$$

We build the following averages

$$\overline{\rho\dot{\eta}} = \overline{\rho}\omega_z + (\overline{Z_o/Z} - 1)\sqrt{v_x^2 + v_y^2} \quad \text{and} \quad \overline{\dot{\rho}/\rho} = -\overline{Z_o/Z}v_z.$$

3. Building for every point the deviation from the average we obtain

$$\frac{\dot{\rho}}{\rho} - \overline{\dot{\rho}/\rho} = -v_z(\frac{Z_o}{Z} - \overline{Z_o/Z})$$
$$\rho\dot{\eta} - \overline{\rho\dot{\eta}} - (\rho - \overline{\rho})\omega_z = \sqrt{v_x^2 + v_y^2}(\frac{Z_o}{Z} - \overline{Z_o/Z}).$$

After testing the vanishing of the lhs of the latter equation indicating a frontal translation we obtain

$$\frac{v_z}{\sqrt{v_x^2 + v_y^2}} = -\frac{\frac{\dot{\rho}}{\rho} - \overline{\dot{\rho}/\rho}}{\rho\dot{\eta} - \overline{\rho\dot{\eta}} - (\rho - \overline{\rho})\omega_z} \tag{11}$$

for every point along the lines $\eta = \phi + \frac{\pi}{2}$ and $\eta = \phi - \frac{\pi}{2}$ (or $\phi + \frac{3\pi}{2}$). Since ω_z is given by the second step we are able to recover the angle between translation and the optical axis. Its tangent is equal to the inverse of the above expression.

We elaborated a method to recover the direction of translation in the case of fixation from the polar motion field based on a feasible 1D search over the angle range. We emphasize that neither we did make any use of the amount of rotation required to fixate (Ω_x, Ω_y) nor did we constrain the motion degrees of freedom. The equations for the logarithmic field are obtained straightforward by substituting (3) in the above expressions. This substitution does not introduce any computational advantages besides complexity reduction, hence, the potential of the method relies only on the polar nature of the log-polar mapping.

3 Experiments on the translation computation

In this section, we present results on the computation of the translation direction in the case of pure translation. We tested the method proposed in section 2 in a log-polar sequence obtained from the cartesian real world sequence "Marbled Block" [1] [7]. One image of the sequence and its log-polar transform are shown in Fig. 1. The log-polar image is drawn such that the η-axis is the horizontal axis and the ξ-axis is the vertical axis pointing downwards. To interpret the log-polar images we note that the angle η is measured beginning counterclockwise from the y-axis which is pointing downwards. So moving horizontally in the log-polar plane we first see the transformed lower right quadrant, then the transformed upper right quadrant and so on. The compression rate obtained by the log-polar transformation is 1:25.

The optical flow is computed using the spatiotemporal derivatives and the assumption that the flow is locally affine in the cartesian domain. Based on the computed log-polar flow of the "Marbled Block" sequence Fig. 1 (bottom-left) we apply the method in section 2 to estimate the direction of translation. We show in Fig. 1 (bottom right) the computed average angular component $\dot{\eta}$ as a function of the angle. We obtain two global minima - differing by 180 degrees as expected - giving as result an angle of 253 degrees for the direction of the line containing the FOE. By averaging the expression $\frac{\dot{\rho}_{\phi+\pi/2}}{\rho^2\dot{\eta}_{\phi+\pi/2}}$ over the diameter line corresponding to $253 - 90 = 163$ degrees we obtain as estimated cartesian FOE-position $(-247, 102)$. Using the intrinsic calibration parameters of the sequence we find out that the angle error between the veridical and the estimated translation direction is 5 degrees which is an acceptable estimation error given the fact that the computational effort decreased to its 4%.

We introduced two new methods for computing the focus of expansion by exploiting the structure of the flow patterns in the log-polar motion field. Both are based on a 1D global minimum search which is inexpensive due to the low angular resolution. The second method necessitates the pursuit movement of the camera proving, thus, that a full attentional mechanism based on both fixation and space-variant sensing enables motion estimation in case of general motion. We implemented the first method in a translating real word sequence obtaining an estimation error of 5° in the translation direction. We plan to implement the second method in the near future using the controllable degrees of freedom of an active binocular head. The mathematical advantages of our approach are due to the polar representation in the log-polar image whereas the complexity advantages are due to the logarithmic nature of the log-polar image.

References

1. J.(Y.) Aloimonos, I. Weiss, and A. Bandyopadhyay. Active Vision. *International Journal of Computer Vision*, 1:333–356, 1988.
2. A. Bandopadhay and D.H. Ballard. Egomotion perception using visual tracking. *Computational Intelligence*, 7:39–47, 1990.
3. R.H.S. Carpenter. *Movements of the Eyes.* Pion Press, London, 1988.
4. C. Fermuller and Y. Aloimonos. The role of fixation in visual motion anaylsis.

[1] Created by Michael Otte at University of Karlsruhe and FhG-IITB, Germany.

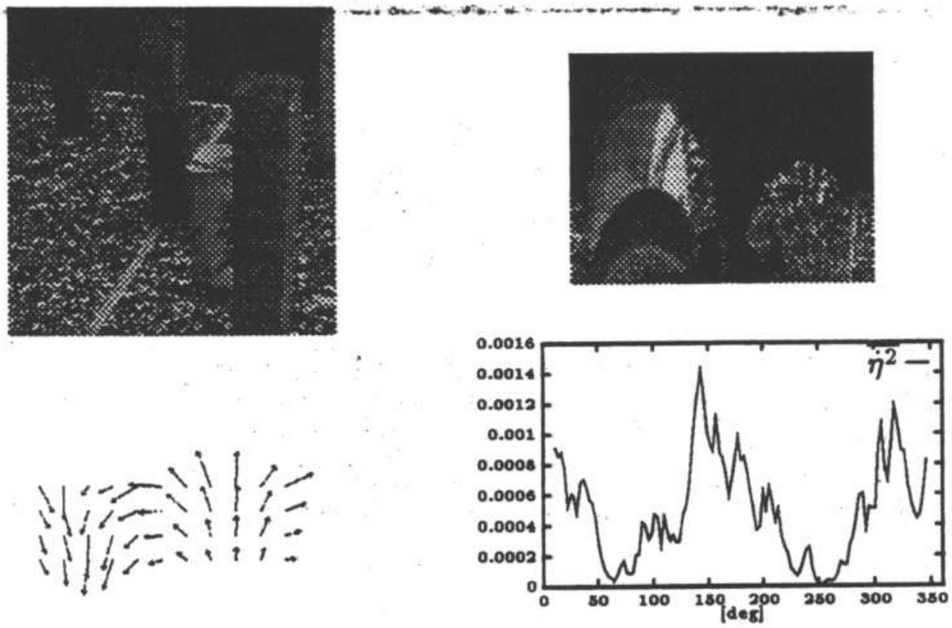

Fig. 1. The original cartesian image (left above), the log-polar transformed image magnified for illustration (right above), the computed optical flow in the log-polar domain (left below), and the average of the squared angular components along the radius as a function of the angle η for the Marbled Block sequence (right below).

International Journal of Computer Vision, 11:165–186, 1993.

5. R. Guissin and S. Ullman. Direct computation of the focus of expansion from velocity field measurements. In *Proc. IEEE Workshop on Visual Motion*, pp. 146–155, Princeton, NJ, Oct. 7-9, 1991.
6. R. Jain, S.L. Bartlett, and N. O'Brien. Motion Stereo Using Ego-Motion Complex Logarithmic Mapping. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9:356–369, 1987.
7. M. Otte and H.-H. Nagel. Optical flow estimation: advances and comparisons. In *Proc. Third European Conference on Computer Vision*, pp. 51–60. Stockholm, Sweden, May 2-6, 1994.
8. D. Raviv and M. Herman. Visual servoing from 2-D image cues. In Y. Aloimonos, editor, *Active Perception*, pp. 191–226. Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.
9. M.A. Taalebinezhaad. Direct recovery of motion and shape in the general case by fixation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14:847–853, 1992.
10. 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. Pattern Analysis and Machine Intelligence*, 15:401–410, 1993.
11. C.F.R. Weiman and R.D. Juday. Tracking algorithms using log-polar mapped image coordinates. In *Proc. SPIE Conf. on Intelligent Robots and computer Vision VIII*, Vol. 1192, Philadelphia, Nov., 1989.