Epipolar Geometry and Log–Polar Transform in Wide Baseline Stereo Matching

Kimmo Palander  
Dept. of Biomedical Engineering and Computational Science  
P.O. Box 9203, FI-02015 TKK, Finland  
kimmo.palander@gmail.com

Sami S. Brandt  
Machine Vision Group  
Information Processing Laboratory  
P.O. Box 4500, FI-90014  
University of Oulu, Finland  
Sami.Brandt@ee.oulu.fi

Abstract

This paper presents an interesting observation that epipolar geometry and log–polar transform can be naturally combined by setting the center of the log–polar transform into the epipoles. This choice preserves the linearity of the epipolar lines. Moreover, the setting is especially advantageous, when camera moves towards the optical axis of the camera as the log–polar transform compensates the large scale changes in the scene. This practically implies that conventional matching techniques can be used with wide baseline images. We discuss the approach with both calibrated and uncalibrated cameras and show some dense wide baseline reconstruction examples where the epipoles are close to the image centers.

1. Introduction

Stereo vision is one of the fundamental research areas in computer vision. The classic setting contains two images taken by two cameras close to each other [5, 11]. Recently, however, there has been lots of research with image reconstruction with widely separated views that is also known as the wide baseline stereo problem [13, 8].

Space-variant representations of images, such as log–polar transform, have been popular due to their data reduction property [3]. They are often motivated by the human retina system and applied in robotics, where real-time data processing is crucial [2]. Among others, log–polar transform is perhaps the most widely used since it is invariant to rotation and scaling and provides a wide field of view. In computing dense stereo correspondence, however, it has been relatively rarely used, possibly due to the complex epipolar geometry in log–polar domain. Previously, Grosso and Tistarelli computed sparse depth maps from pre-filtered log–polar images [4]. Bernardino and Santos-Victor produced dense depth maps from log–polar images by pre-computed look-up tables [1]. Quite recently, Schindler formulated the epipolar geometry on the log–polar plane [12]. None of the previous works did not, however, consider how the center of the coordinate system should be selected for the log–polar domain.

The principal idea of this paper is to show how the epipolar geometry in log–polar domain is remarkably simplified by a particular choice of the coordinate center. Previous works on log–polar epipolar geometry set the center of the transformation to the Cartesian image center [12]. However, this choice results in complex geometry and thus in complex computation of stereo correspondence. Since epipolar lines originate from the epipoles, we obtain a simple log–polar epipolar geometry by setting the centers in the log–polar transform to the corresponding epipoles. This choice leaves the

Figure 1. Epipolar lines (left) do not generally remain straight under log–polar transform (middle), unless the center of the transform is set to the epipole (right).
epipolar lines straight in both domains, see Fig. 1. If the epipoles are at infinity, this scheme is obviously not realizable. Furthermore, since robotic applications require high resolution at the region of interest, the epipoles should preferably be located in the images. As an application, we demonstrate how to compute dense depth maps from wide baseline stereo images where the camera approximately translates towards the scene.

2. Log–polar image transform

The log–polar mapping is a 2D conformal transformation \([x, y] \rightarrow [\rho, \alpha]\) around the chosen center \([x_c, y_c]\) or

\[
\begin{align*}
\rho &= \frac{1}{\lambda} \log \lambda \left( (x - x_c)^2 + (y - y_c)^2 \right) \\
\alpha &= \arctan \left( \frac{y - y_c}{x - x_c} \right)
\end{align*}
\]

(1)

where \(\lambda\) is the base of the logarithm. This mapping has a singularity at the center that is avoided in practise by defining the mapping only for radial values greater than \(r_{\text{min}}\). The inverse transform is

\[
\begin{align*}
x &= \lambda^\rho \cos(\alpha) + x_c \\
y &= \lambda^\rho \sin(\alpha) + y_c
\end{align*}
\]

(2)

Figure 2 illustrates the transformation.

Although there is even hardware available that produces true log–polar images [2], the common way to acquire log–polar images is to resample Cartesian images into log–polar domain using (1). Since the sampling is space-variant, we need to apply adaptive low-pass filtering in order to avoid aliasing artefacts. While some authors do this with mapping templates that model the human retinal system [3, 4], a fast implementation can be achieved by interpolation on image pyramids.

3. Log–polar stereo geometry

If a stereo image pair was acquired by forward (or backward) camera movement, both epipoles would have the same central location in the images. In such case, the epipoles are known as the focus of expansion (FOE) [5]. Hence, forward motion seems to be a plausible stereo scheme in log–polar domain. Furthermore, large movement results in a heavy scale change of the scene that suggests to take advantage of the scale invariance property of the log–polar transform.

We now investigate the depth parameter assuming that the camera moves towards its optical axis. We assume that camera parameters do not change between the views. However, we do not make any other restrictions about the internal parameters of the camera.

Without loss of generality, we fix the world origin to the first camera center and fix the orientation of the world coordinate frame to that of the first camera. Then the projection matrices corresponding to the two views are

\[
P_1 = K[I|0], \quad P_2 = KR[I|t_C],
\]

(3)

where \(K\) is the calibration matrix of the camera, \(R\) describes the rotation of the second view and \(t_C\) denotes the translation of the camera center. The camera calibration matrix \(K\) is defined as

\[
K = \begin{pmatrix} 
\alpha_x & s & p_x \\
0 & \alpha_y & p_y \\
0 & 0 & 1
\end{pmatrix},
\]

(4)

where \(\alpha_x\) and \(\alpha_y\) express the focal length of the camera in the \(x\) and \(y\) directions respectively, \(s\) is the skew and \((p_x, p_y)\) is the principal point [5].

In our application, we set the origin of the image plane to the principal point so that \((p_x, p_y) = (0,0)\). Furthermore, in the case of pure translation, \(R = I\) and the movement \(t_C = (0,0, t_z)\). While these conditions are seldom true in real world applications, equivalent setting can be however achieved if the cameras are calibrated by applying the homography to the images that transforms the optical axis perpendicular to the transformed image plane. In the uncalibrated case, the best homography which makes the epipolar lines pairwise identical can be alternatively computed, as will be shown in Section 4.

Let \(M = (X, Y, Z, 1)^T\) represent a point in space where \(Z\) is the depth parameter, see Fig. 3. The point is projected onto the images by

\[
P_1 M \approx \begin{pmatrix} 
\alpha_x X + s Y \\
\alpha_y Y \\
Z
\end{pmatrix}, \quad P_2 M \approx \begin{pmatrix} 
\alpha_x X + s Y \\
\alpha_y Y \\
Z + t_z
\end{pmatrix}. 
\]

(5)
Figure 3. Geometrical problem setting for computing depth value $Z$.

The Euclidean distances from the origin are then

$$
\begin{align*}
&d_1 = \frac{1}{Z} \sqrt{(\alpha_x X + s Y)^2 + (\alpha_y Y)^2} \\
&d_2 = \frac{1}{Z - t Z} \sqrt{(\alpha_x X + s Y)^2 + (\alpha_y Y)^2}
\end{align*}
$$

(6)

and the relation of the distances is

$$
\frac{d_1}{d_2} = 1 - \frac{t Z}{Z}.
$$

(7)

Modifying the left hand side of (7) into logarithmic form we get

$$
1 - \lambda \rho_1 - \rho_2 = \frac{t Z}{Z'},
$$

(8)

where $\rho_i = \log_\lambda (d_i)$. For small scale changes, $\rho_1 - \rho_2$ is close to zero, so we can take the first order approximation of the left hand side of (8) and thus we have approximately

$$
\rho_2 - \rho_1 \propto \frac{1}{Z}.
$$

(9)

This means that, in principle, we can apply any conventional stereo algorithm on log–polar images, even for wide baseline schemes. The disparities in log–polar domain are then related to depth according to (8) (or approximately to (9)). The scale changes, which are problematic in classical stereo, are thus compensated by the log–polar transform.

4. Preprocessing for the uncalibrated case

In an uncalibrated case, the orientations of the optical axes are not known and the perspectivities, which make the optical axes parallel to the translation of the camera, can not be constructed. However, we may always correct the images so that the epipoles have the same position in the images and that the corresponding epipolar lines are the same line. The correction can be understood to have a similar role as the rectification methods [10, 9].

Assuming the fundamental matrix $F$ of the stereo pair is known, we can make the corresponding epipolar lines identical by applying a compatible homography to one of the images [9]. Homographies compatible with the fundamental matrix are of the form

$$
H = [e']_x F - e' v^T,
$$

(10)

where $v$ is a random vector so that $\det(H) \neq 0$ [5]. The homography can be solved by requiring $H x = x'$, where $x$ and $x'$ are corresponding points in the images. Assuming Gaussian noise in the point locations we compute the homography by simply solving

$$
\min_v \sum_i \|H x - x'\|^2.
$$

(11)

5. Experiments

To experiment the principle proposed in this paper, we computed dense reconstructions for uncalibrated scenes with wide baselines, where the camera moved towards the scene between the views. Since the proposed principle is directly usable for any dense stereo reconstruction algorithm, we experimented correlation and graph cut methods, using the code provided by Kolmogorov [6].

Two pairs of stereo images of the size $1024 \times 768$ pixels of indoor and outdoor scenes were acquired by a digital hand-held camera for which the fundamental matrix as well as the compatible homography were estimated by using the SIFT features [7]. In order to discard outliers in the feature correspondences, we also used a robust LMedS estimator [14]. Finally the log–polar images were sampled around the estimated epipoles into the size of $640 \times 640$ pixels.

The results are shown in Figs. 4 and 5. In Fig. 4, due to oversampling close to the center of the log–polar transformation, the computed depth values are somewhat ambiguous. Although the result may seem awkward by the first sight in the log–polar domain, the inverse transformed Cartesian depth maps show that the results are reasonable. Fig. 5 shows an outdoor scene that appears to be more challenging to the stereo algorithms due to large scale variation in the images. Also in this case the reconstruction results are promising.

6. Conclusions

We have proposed to use the log–polar transform with epipolar geometry by setting the center of the
transform into the epipoles. This choice is practically useful when the epipoles are located in the images or when the camera moves towards the scene. Since scale changes become translations on the log–polar plane, conventional stereo algorithms can be applied with wide baseline stereo pairs in the log–polar domain. The experiments verified the applicability of our approach and showed promising results.

References


