Range Estimation from a Pair of Omnidirectional Images

Roland Bunschoten  
Ben Kröse

RWCP, Autonomous Learning Functions Lab, SNN  
Computer Science Institute, Faculty of Science  
University of Amsterdam, The Netherlands  
\{bunschot,kroese\}@science.uva.nl

Abstract

In this paper we address the problem of recovering range information from a pair of images obtained from camera viewpoints whose relative poses are known (stereo vision). The images we deal with are cylindrical re-projections of images captured by an omnidirectional vision sensor. Images obtained by such a device have a relatively low spatial resolution in comparison to standard camera images. We analyze the sensitivity of range estimation from image correspondences with respect to errors caused by discretization. The analysis reveals the significance of obtaining sub-pixel accuracy for range estimation. We present our method and experimental results. The results indicate that high accuracy range estimates can be obtained using our method.

1 Introduction

For global, goal-directed navigation a mobile robot needs an internal representation of its environment. Traditionally, the environment is modeled as an explicit geometric model. Recently, several appearance based modeling approaches have been proposed, using the exteroceptive sensor systems on the robot. In these approaches the relationship between the sensory observations (or features derived from them) and robot positions is modeled. Omnidirectional vision sensors have proven to be very useful for environment modeling, both in the geometric domain [12], as well as in the sensory domain [5, 3].

A drawback of appearance based environment representations is that many training images are needed to obtain an accurate model. One approach to this problem was presented in [1], where based on a number of measured range profiles synthetic profiles are generated. We are working on an extension of this method [6] which incorporates vision data. In this paper we describe an omnidirectional stereo vision algorithm which provides the range estimates required by the method.

The omnidirectional vision system on our robot is based on a hyperbolic mirror for the image formation. In that sense it is different from other systems which have been used for range estimation from omnidirectional vision. McMillan and Bishop [7] and Kang and Desikan [4] composite a panoramic omnidirectional image from perspective images captured by a rotating camera mounted on a tripod. The advantage of the mirror based system is the fast acquisition rate. A disadvantage is that the omnidirectional images acquired have a relatively low spatial resolution which may cause considerable errors in range estimation.

Stereo vision is a method to measure range from two (or more) images by the principle of triangulation. Establishing matches between pixels taken from the different images is the fundamental problem in stereo vision. Once the matches are established, range can be estimated using a triangulation technique.

This paper is organized as follows. In section 2 we describe our omnidirectional vision system. The derivation the epipolar geometry (which reduces the search space for possible matches), the image matching procedure and the range estimation procedure are described in sections 3–5. In section 6 an analysis of the sensitivity of the system to matching errors is presented. The analysis shows that obtaining matches at sub-pixel accuracy is desirable. Our method to obtain sub-pixel accuracy is also presented in section 6. In section 7 we show some experimental results obtained using synthetic, but realistic, image data generated by a ray-tracer program.
3 Epipolar Geometry for a Panoramic Vision Sensor

Stereo vision is the process of recovering range information from two (or more) images obtained from different but known relative camera poses. The range to a physical point can be computed by triangulation if projections of the point in at least two images is known (reconstruction problem). Establishing the matches between image elements in the different images is the fundamental problem in stereo vision (correspondence problem). If knowledge about the camera geometry and relative viewpoints is available, a powerful geometric constraint, known as the epipolar constraint, reduces the search space for possible matches from two dimensions (the entire image plane) to one dimension (the epipolar curve).

Two viewpoints $v_0$ and $v_1$ are related by a rotation, given by a $(3 \times 3)$ rotation matrix $R$ and a translation given by a $(3 \times 1)$ translation vector $t$. Let $p_1$ be a physical point in the first coordinate frame, then the representation of the point $p_2$ in the second frame is

$$ p_2 = R p_1 + t. \quad (1) $$

Let $x_1$ be the projection of $p_1$ onto a unit radius cylinder defined in the $v_1$ coordinate system. The plane spanned by $t$ and $x_1$ defined by its normal $n_1$ is called the epipolar plane. In the $v_2$ coordinate system the normal to the epipolar plane is computed as

$$ n_2 = R n_1 = R (t \times x_1). \quad (2) $$

This establishes the epipolar geometry: in order to find the $x_2$ corresponding to $x_1$ only those $x_2$’s satisfying $n_2 \cdot x_2 = 0$ need to be searched.

Expanding the dot product and expressing $x_2$ in cylindrical coordinates $(\phi_2, z_2)$ gives

$$ n_{2, x} \cos(\phi_2) + n_{2, y} \sin(\phi_2) + n_{2, z} z_2 = 0. \quad (3) $$

which can be rewritten as

$$ z(\phi) = -\frac{n_{2, x} \cos(\phi) + n_{2, y} \sin(\phi)}{n_z}. \quad (4) $$

Besides computational gain, an important advantage of using the epipolar constraint is that there is less chance to establish erroneous pixel correspondences. The domain of correspondence can be further restricted by assuming a minimal and a maximal distance to the objects in the environment.

The epipolar geometry can also be derived for the hyperboloid omnidirectional images [11]. An advantage of the panoramic representation of omnidirectional images is that a single type of epipolar curve
is obtained. In the hyperboloid domain, the epipolar curves are general conics of various types, each of which requires a different parameterization [10].

The epipolar geometry for cylindrical projection surfaces is illustrated in Figure 2. In the figure, $d_{\min}$ and $d_{\infty}$ represent vectors in the direction of $x_1$ whose length is indicated by the subscript.

4 Matching

Matching in stereo vision involves comparison of local image regions (or intrinsic local characteristics, i.e., features) and determining which regions are most similar according to some matching criterion. Matching methods can be categorized as local or global. Local matching methods attempt to match each pixel independently. Global methods attempt to match jointly across epipolar lines or jointly across the entire image domain. Global matching methods consider physical constraints such as surface continuity which helps to resolve local ambiguities. The optimization process performed by global methods is however computationally more expensive.

A further distinction between matching methods is whether they compare discrete abstract image features or correlate small image windows. Features are usually chosen to be insensitive to viewpoint and illumination changes. In contrast with window based correlation methods, feature based matching methods generally produce sparse range results. Window based correlation methods implicitly assume that the disparity is similar for each pixel in the window under consideration. This assumption is violated for large windows.

Furthermore, using larger windows increases the computational cost. Small windows, on the other hand, have a lower signal-to-noise ratio and are more susceptible to perceptual aliasing. The interested reader is referred to the book of Faugeras [2] for an extensive treatment on stereo vision and related topics.

The matching performed by our stereo algorithm can be characterized as local window-based correlation matching. Recall that our omnidirectional camera is mounted on top of our mobile robot. The camera motion is thus restricted to translation parallel to the ground plane (which we assume to be planar) and rotation about a vertical axis. Although a pair of omnidirectional images can be used directly to establish image matches we use panoramic images. This enables us to use standard correlation techniques. Whereas a robot rotation causes a shift and rotation of the image content in the omnidirectional image, it causes a translation only in the panoramic image.

In order to obtain some illumination invariance we take the Laplacian of Gaussian transform of the panoramic intensity images. The Laplacian measures intensity edge steepness and can be computed by convolving an intensity image with the partial derivatives of a Gaussian at a particular scale [8].

Matching is done as follows. For a given point $(\phi_1, z_1)$ in image 1, the epipolar geometry relationships are used to find the interval $[\phi_2 - d_{\min}, \phi_2 + d_{\infty}]$ in $v_2$. The epipolar curve is sampled resulting in a set of points $(\phi_2, z_2)$. The best match is found by computing the correlation values for each point and selecting the one which matches best. The similarity measure we choose is the Euclidean norm or Sum of Absolute Differences (SAD).

Window based correlation is a relatively expensive operation. In order to avoid unnecessary computations the epipolar curve is only scanned when all sampled points on the epipolar curve reside inside the $v_2$ image domain. This can be confirmed by checking the end-points of the search interval and by checking whether the epipolar curve has an extremum in the interval and, if it does, whether the extremum is inside the image boundaries.

5 Range Estimation

Range can be estimated from the angular difference under which $p$ is perceived from different viewpoints. The cosine rule is used to compute the angle between

Figure 2: Epipolar geometry for cylindrical projection surfaces.
the vector $\mathbf{z}_1$ and translation vector $\mathbf{t}$

$$\theta_1 = \arccos\left(\frac{\mathbf{z}_1 \cdot \mathbf{t}}{\|\mathbf{z}_1\| \|\mathbf{t}\|}\right).$$

(5)

Similarly, the angle between $\mathbf{z}_2$ and the rotated translation vector $\mathbf{Rt}$ is computed as

$$\theta_2 = \arccos\left(\frac{\mathbf{z}_2 \cdot \mathbf{Rt}}{\|\mathbf{z}_2\| \|\mathbf{Rt}\|}\right).$$

(6)

Rewriting the sine rule gives us the lengths $d_1$ and $d_2$ which correspond to the range from viewpoint $v_1$ and $v_2$ to $\mathbf{p}$ respectively.

$$d_1 = \frac{\sin \theta_2}{\sin(\theta_2 - \theta_1)} \|\mathbf{t}\|,$$

(7)

$$d_2 = \frac{\sin \theta_1}{\sin(\theta_2 - \theta_1)} \|\mathbf{t}\|.$$  

(8)

The three dimensional point $\mathbf{p}$ can be reconstructed using the relationship

$$\mathbf{p}_i = \frac{\mathbf{z}_i}{\|\mathbf{z}_i\|} d_i$$

(9)

where $i$ indexes the viewpoint in whose coordinate system $\mathbf{p}$ is expressed.

Note that the theory presented in this section is general in the sense that angular difference is measured in the epipolar plane, i.e. the transformation between two viewpoints can be general. In our implementation we follow a slightly different scheme which takes advantage of the constraints on viewpoint transformations discussed in the previous section.

6 Sensitivity and sub-pixel accuracy

Sensitivity with respect to errors in viewing angles and norm of the translation vector can be investigated by examining equations (7) and (8).

Differentiating 7 with respect to with respect to $\|\mathbf{t}\|$ (known as the baseline length in stereo) gives

$$\frac{\partial d_1}{\partial \|\mathbf{t}\|} = \frac{\sin \theta_2}{\sin(\theta_2 - \theta_1)} = \frac{d_1}{\|\mathbf{t}\|}.$$  

(10)

This shows that an error in $\|\mathbf{t}\|$ causes an error in the depth estimate which is proportional to the depth and inverse proportional to the baseline length.

Differentiating with respect to $\theta_2$ gives

$$\frac{\partial d_1}{\partial \theta_2} = d_1 \left(\frac{\cos \theta_2}{\sin \theta_2} - \frac{\cos(\theta_2 - \theta_1)}{\sin(\theta_2 - \theta_1)}\right).$$

(11)

For a point observed under an angle $\theta_1 = \pi/2$ relative to the translation vector this reduces to:

$$\frac{\partial d_1}{\partial \theta_2} = \frac{d_1^2}{\|\mathbf{t}\|} - \|\mathbf{t}\|.$$  

(12)

In a typical situation where the robot translates $0.5$ m and where the distance to the wall is $3$ m, an error of $1$ degree in the estimation of the disparity results in an error in the depth of about $0.3$ m. Note that the error can be reduced by extending the baseline. However, for a larger translation between viewpoints matching is more difficult due to image deformations and occlusions.

The above analysis shows that the disparity needs to be computed with high accuracy. Given the discrete representation of the panoramic images errors will be made. To find the angle $\theta_2$ for which the matching is best we sample the epipolar curve at a finite number of points. For each sampled point we select the nearest pixel and compute the correlation between the windows.

In finding the best match, there are two sources of error in the scheme described. First, correlation values are computed at nearest discrete pixel positions, not at the continuous point on the epipolar curve under investigation. One solution to this problem is to use some kind of interpolation. However, this would increase the computational demands. A second error source is that the epipolar curve is only sampled at a finite number of points, i.e. the angular resolution is limited. One solution would be to sample the epipolar curve more densely. However, without interpolation this would result in the same SAD value for multiple points along the epipolar line. With interpolation, denser sampling of the epipolar curve would increase the computational demands even further.

Our solution to these problems is to interpolate between correlation values along the epipolar curve by fitting a polynomial of degree 2 to the correlation measurements in a least-squares sense. In the fitting pro-

Figure 3: SAD values along the epipolar curve sampled at $\phi_2$ and the fitted parabola.
procedure, only neighboring points of the sample point which has the best match at discrete resolution contribute. By setting the derivatives of the polynomial with respect to \( \phi_2 \) to zero, a new minimum for \( \phi_2 \) can be computed. Via application of the epipolar curve equation this gives rise to a new point on the epipolar curve from which range can be estimated using the theory of the previous section. Figure 3 shows the correlation values as a function of \( \phi_2 \) and a fitted parabola. Three neighbors around the minimum on both sides contribute to the estimated coefficients.

7 Experiments

We have tested our system on synthetic, but realistic, image data. A ray tracer program is used to render images obtained by an omnidirectional vision sensor in a virtual environment. Figure 4 gives an overview of the virtual environment and the different representations of an image acquired in the environment. The environment consists of a box (6000 \( \times \) 6000 \( \times \) 6000 mm) with textured sides and bottom.

Range is computed from the panoramic images (720 \( \times \) 200 pixels) derived from omnidirectional images (640 \( \times \) 480 pixels) acquired at locations [0, 0, 1000] and [500, 0, 1000]. Figure 5 shows the range estimations obtained for a single row of pixels taken from the panoramic image obtained at [0, 0, 1000] (results for other rows give similar range estimates). Figure 5(a) shows the range profile estimated at pixel resolution. We have verified that the discontinuities in the curve are caused by discretization. Figure 5(b) shows the profile estimated by our sub-pixel method.

8 Discussion and Conclusion

We have presented our stereo vision method for an omnidirectional vision sensor consisting of a hyperboloid mirror and a perspective camera. Such omnidirectional sensors have a limited resolution. Discretization of the epipolar curve and discrete matching can introduce considerable error and bias in the range estimate. Our sub-pixel method compensates for this.

Our results indicate that range can be estimated reliably by our method. We are currently testing the system on data obtained using a real omnidirectional sensor. This poses some practical challenges, e.g. motion between viewpoints has to be estimated as in e.g. [7, 11], camera lens distortions need to be taken into account, etc.

Figure 4: The virtual environment and various representations of an image acquired in the environment.
Our goal is to warp panoramic images to images obtained at novel viewpoints [7, 4] and to use this ability in the context robot map learning in a manner similar to [1, 6].

Acknowledgements

The authors would like to thank Tomáš Svoboda for kindly providing the POV-Ray source code implementing the hyperboloid mirror he used in his PhD work [9]. This work is supported by the Real World Computing Program.

References


