A new Method on Camera Ego-motion Estimation

Ding Yuan
School of Astronautics
Beijing University of Aeronautics and Astronautics
Beijing, China

Yalong Yu
School of Astronautics
Beijing University of Aeronautics and Astronautics
Beijing, China

Abstract—In this paper, we address the problem of ego-motion estimation for a monocular moving camera, which is under arbitrary translation and rotation. It has been one of the most important problems on the application of computer vision in the mobile robots. The problem is equivalent to determine the 3D motion parameters of a camera by observing an image sequence taken by it over time. A robotic system can be well guided if there is a valid way for it to obtain the information about its own motion, so that an accurate estimation of ego-motion is very useful for the robots’ navigation. The new method we propose here is uniquely based on the spatial-temporal image derivatives of an image sequence, that is, the normal flow, which is the projection of the optical flow on the direction of the image gradient. The computation of the normal flow, which is the image flow component that can be estimated based on the local measurements alone, does not require any special assumption about the scene structure. This method is less demanding than those methods based on the optical flow about the observed scenes, and does not need to add any special assumption to the observed scenes, so that it can be used wildly in the real world. First, we determine the range of each rotational parameter roughly and we are sure that the ground truth of each rotational parameter must be included in the corresponding range. Second, we search for the ground truth of each rotational parameter in the corresponding range. Once the ground truth of rotational parameters is determined, the location of the Focus of Expansion (FOE) is determined simultaneously, which gives the direction of the camera linear velocity. We have conducted many experiments with synthetic data and real images to verify the accurateness and robustness of the method we propose in this paper.

Keywords—ego-motion estimation; the normal flow; FOE

I. INTRODUCTION

We address the problem of ego-motion estimation for a monocular moving camera with arbitrary translation and rotation. The problem consists of determining the 3D motion parameters of a camera by observing an image sequence taken by it over time. This problem has attracted a lot of attention in the past many years. It is a key step for the navigation of robots where a robot must be able to estimate and control its motion parameters before any higher level tasks to be addressed.

The motion field observed in an image sequence can be caused by the motion of a camera, that is, the ego-motion and by the motion of objects included in the observed scene, however, we only take the ego-motion into consideration and leave out the motion of objects in this paper. Estimating the motion parameters of a camera is a difficult task because the image motion at each pixel depends, in addition to six parameters of the ego-motion, on the depth information of the corresponding scene point. In order to overcome this difficulty, additional constraints are usually needed to add to the motion model or to the environment model. However, we can only retrieve 5 unknown parameters in all, that is, the rotational parameters and the direction of the translation of the camera, because the depth information is unknown.

Up to now, there have been many methods for camera ego-motion estimation. They mainly fall into two major categories: indirect and direct methods.

For example, methods based on feature correspondences or the optical flow belong to the indirect; they first obtain the image motions by detecting the corresponding features or computing the optical flow in images, and then solve the equations relating these image motions to the 3D motion parameters. Generally speaking, however, the correspondence problem and the estimation of the optical flow are all ill-posed, it is usually necessary to add very restrictive assumptions to the observed scenes. These methods often require extensive amounts of computation, and they are also time-consuming.

Methods based on the normal flow belong to the direct; they estimate the ego-motion parameters directly from the normal flow, which is fully determined by the spatial-temporal derivatives of an image sequence, and is the image flow component that can be estimated based on the local measurements alone due to the “aperture problem.” They are less demanding about the observed scenes than the indirect ones, and do not need to add any assumption to the observed scenes, so that they are used widely. These methods estimate the translational and rotational parameters separately due to the bilinear nature of the image motion equations. They usually first determine the location of the FOE, and then estimate the rotational parameters based on the obtained information.

The approach we propose in this paper is also based on the normal flow. We take advantage of the normal flow as the only input to the system. But unlike those direct methods mentioned above, in the first step, we determine the range of each rotational parameter roughly and we are sure that the ground truth of each rotational parameter must be included in the corresponding range; in the second step, we search for the ground truth of rotational parameters in these three ranges. And once the ground truth of rotational parameters is obtained, the location of the FOE is determined in the meantime, which gives the direction of linear velocity of the camera.
II. PROBLEM STATEMENT

Throughout the paper, we consider a camera-centered coordinate system, as depicted in Fig. 1. The image formation model we use in this paper is the perspective projection. For a 3D point \( P(X,Y,Z) \) in the space, the projection of this point on the image plane is \( p(x,y) = (fX/Z, fY/Z) \), where \( f \) denotes the focal length of the camera. Now we consider a camera with linear velocity \( \mathbf{t} = [U, V, W]^T \) and angular velocity \( \mathbf{w} = [w_1, w_2, w_3]^T \), the optical flow vector at each image point in the image plane can be calculated by the following well-known equation:

\[
v(x) = \rho(x)(x - \Sigma) + \mathbf{B(x)}w
\]

where \( v(x) = [u(x), v(x)]^T \) is the optical flow observed at a given image point \( x = [x, y]^T \); \( \Sigma = [\rho U, \rho V, \rho W]^T \) is the Focus of Expansion(FOE) which corresponds to the projection of the linear velocity of the camera on the image plane; the function \( \rho(x) = W/Z \) denotes the inverse of the time-to-crash which expresses the time remaining till the object hit the infinitely large image plane, \( Z \) denotes the scene depth of each corresponding scene point ; and \( \mathbf{B(x)} \) is as follows:

\[
\mathbf{B(x)} = \begin{bmatrix}
x y/f & -((x^2 + f^2 + f)/x + f) \\
(y^2 + f^2 + f)/x & -xy/f & -x
\end{bmatrix}
\]

Due to the well-known constraint of the “aperture problem”, we can only observe the projection of the optical flow on the direction of the image gradient \( \mathbf{n(x)} = [n_x, n_y]^T \), namely, the normal flow. Thus, the value of the normal flow at each image point can be calculated by the function \( V_x = \mathbf{n(x)^T v(x)} \). Now, we expand this function as follows:

\[
V_x = V_x + V_\mathbf{r} = \rho(x)\mathbf{n(x)^T (x - \Sigma)} + \mathbf{n(x)^T B(x)}w
\]  

where \( V_x \) and \( V_\mathbf{r} \) are the translational and rotational component of the normal flow \( V_x \) respectively. However, computing the normal flow in images is as difficult as detecting edges. The difficulty in the computation of the normal flow is only due to the discrete aspect of digital images. Here, we assume that the image intensity remains constant, and the normal flow is computed from the spatial-temporal derivatives of the image intensity function by employing the motion constraint equation.

Our method proposed in this paper works in two successive steps: the first step is used to determine the range of each rotational parameter; and the second step is aimed to determine the ground truth of the rotational parameters and the FOE based on the information obtained in the first step. The rationale of our method will be represented in the following section at length.

III. EGO-MOTION ESTIMATION

According to (3), we know that the normal flow consists of the translational component and the rotational component. However, we cannot simultaneously estimate all of the motion parameters only by solving this equation, because the normal flow at each image point also has something to do with the corresponding scene depth which is unknown to us. This is undoubtedly a big difficulty in the ego-motion estimation. In order to overcome this difficulty, many researchers in this domain try their best to search for some special image points in the images where the translational components of the normal flows vanish, thus the normal flows at these points only contain rotational components so that these points can be utilized to estimate rotational parameters. These special image points can even be used to determine the location of the FOE by the equation \( \mathbf{n(x)^T (x - \Sigma)} = 0 \). These special image points, however, cannot be easily found in the images.

We have known that the lines perpendicular to the normal flow vectors at these special image points mentioned above must intersect at one point, which is happen to be the FOE, as depicted in Fig. 2. In our approach, we take advantage of this important property in the process of searching for the ground truth of each rotational parameter in the corresponding range. According to whether the intersections of the lines which are perpendicular to the normal flow vectors at the selected image points are converged at one point or not, we can determine the ground truth of rotational parameters.

A. Estimating the range of each rotational parameter

Assuming that each image point in the image is a possible FOE, we can use the equation \( \mathbf{n(x)^T (x - p\Sigma)} = 0 \), where \( p\Sigma \) denotes the possible FOE, to search for those image points at which the normal flows, we think, only contain rotational components. Thus we can use these image points to estimate the rotational parameters. In fact, the normal flows at these image points may also contain translational components. So the normal flows at these image points can be expressed as follows:

\[
V_x = \text{Err} + \mathbf{n(x)^T B(x)}w
\]
where $Err$ denotes the possible translation component. The $Err$ may be negative, zero or positive. Here, we use the least square method to estimate the rotational parameters. As far as we know the least square method has the property that the estimated result is equivariant with respect to the response variable [12]. Therefore, when $Err$ is positive, the estimated result is larger than the ground truth. On the contrary, the estimated result is less than the ground truth.

Thus, we can obtain the range of each rotational parameter roughly based on the estimated results of rotation parameters. We are sure that the ground truth of each rotational parameter must be included in the corresponding range.

\section*{B. Determining the ground truth of rotational parameters and the location of the FOE}

In this stage, we make use of the information obtained above and the property that the lines perpendicular to normal flow vectors at those special image points must intersect at the FOE to search for the ground truth of rotational parameters and determine the location of FOE.

Having known the range of each rotational parameter, we take any set of rotational parameters, which is formed by randomly choosing a value in each corresponding range, into consideration. For a given set of rotational parameters $\hat{\mathbf{w}}$, firstly, we recalculate the normal flow at each image point using the formula $\hat{\mathbf{V}}_n = \mathbf{n}(\mathbf{x})^\top \mathbf{B}(\mathbf{x}) \hat{\mathbf{w}}$. Secondly, we compare $\hat{\mathbf{V}}_n$ with $\mathbf{V}_n$ at each image point, and select out those image points at which $\hat{\mathbf{V}}_n$ are equal to $\mathbf{V}_n$ in the image. Thirdly, we calculate the coordinates of the intersection of any two of lines which are perpendicular to the normal flow vectors at those selected image points using the equation $\mathbf{n}(\mathbf{x})^\top (\mathbf{x} - \mathbf{x}_p) = 0$, where $\mathbf{x}_p$ represents the intersection point. In this step, if the directions of the image gradients at two image points are parallel to each other, the intersection of the lines perpendicular to normal flow vectors at these two image points will be at infinity. These intersection points at infinity will not be taken into consideration in the following step. Fourthly, we take advantage of some robust statistics methods, such as MSE, to analyze the distribution of these intersections. We calculate the x-MSE and the y-MSE separately, which are the MSE values of the x-coordinate values and the y-coordinate values of these intersections respectively. If the summary of the x-MSE and the y-MSE is larger than the threshold $\varepsilon$ we set, it means that these intersections are dispersed. On the contrary, if the summary is less than the threshold $\varepsilon$, it means that these intersections must be converged at a point, that is, the FOE.

If the values of the rotational parameters are equal to the ground truth, the image points selected in the second step will be those special image points at which the normal flow values only contain rotational components. Thus, the intersections of every two of the lines perpendicular to the normal flow vectors at these special image points must be converged at a point, namely, the FOE, except for those intersections at infinity. In other words, the summary of the x-MSE and the y-MSE is less than the threshold $\varepsilon$ we set. On the contrary, if the values of the rotational parameters deviate from the ground truth, the image points selected in the second step will be those image points where the normal flow values also contain translational components. Thus, the intersections of every two of the lines perpendicular to the normal flow vectors at these image points must be dispersed. In other words, the summary of the x-MSE and the y-MSE is larger than the threshold $\varepsilon$ we set.

Based on the fact stated above, we can determine whether the given values of the rotational parameters are equal to the ground truth or not. And once the values of the rotational parameters are determined, the location of the FOE can be obtained simultaneously.

\section*{IV. Experiments and Results}

The approach we propose in this paper has been tested in a series of experiments using both synthetic date and real images. As the range of each rotational parameter determined by assuming that image points every several pixels are possible FOEs is approximate to the range determined by supposing that each image point is a possible FOE, we use the former in the following experiments. Doing so not only can we save the time of computation, but also can show the experimental results clearly.

The parameters of the synthetic data are as follows: focal length in the X-direction is 1163 pixels, focal length in the Y-direction is 1316 pixels, the image size is $574 \times 652$, center of the image is (332,269) measured from the top left corner. The image gradient and the scene depth of each image point are generated randomly. The normal flows induced by the given motion of a camera in the image are calculated by (3).

In the first experiment, using synthetic data, we assume that the FOE is located at (-129,-146), the values of the rotational parameters are $\mathbf{w} = [0.010,0.0020,0.050]^\top$. In our algorithm, we firstly determine the range of each rotational parameter, and the results are shown in Fig. 3; secondly, we search for the ground truth of each rotational parameter in the corresponding range based on whether the intersections of every two of the lines perpendicular to the normal flow vectors at those selected image points are converged at a point, namely, the FOE or not. Once the values of the rotational parameters are determined,
the location of the FOE in the image plane is determined at the same time, the result is shown in Table I.

In the second experiment, using synthetic data, we assume that the FOE is also located at the point of (-129,-146), the values of the rotational parameters are \( w^T = [1.5, 0.20, 10] \). In the first step, we approximately estimate the range of each rotational parameter, and the results of this step are shown in Fig. 4; in the second step, we search for the ground truth of the rotational parameters and determine the location of the FOE in the image plane, and the result of this step is shown in Table I.

![Figure 3](image)

**Figure 3.** The ranges of rotational parameters in the first experiment.

![Figure 4](image)

**Figure 4.** The ranges of rotational parameters in the second experiment.

In the third experiment, with real images shown in Fig. 5, the parameters of the real images are as follows: focal length in the X-direction is 1163 pixels, focal length in the Y-direction is 1316 pixels, the image size is \( 574 \times 652 \), center of the image is (332, 269) measured from the top left corner, the ground true of the FOE is (-129,-146) (measured from the image center, the X-direction points to the right side and the Y-direction points down), \( w^T = [0.00012, 0.00025, 0.0010] \) [7]. The ranges of rotational parameters are shown in Fig. 6. The estimated result of motion parameters is shown in Table I.

![Figure 6](image)
TABLE I. THE RESULTS OF EXPERIMENTS

<table>
<thead>
<tr>
<th>Exp.</th>
<th>Ground truth</th>
<th>Estimated result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FOE</td>
<td>Rotational parameters</td>
</tr>
<tr>
<td>1</td>
<td>(-129.000,-146.000)</td>
<td>[0.010;0.0020;0.050]</td>
</tr>
<tr>
<td>2</td>
<td>(-129.000,-146.000)</td>
<td>[1.5;0.20;10]</td>
</tr>
<tr>
<td>3</td>
<td>(-129.000,-146.000)</td>
<td>[0.00012;0.00025;0.0010]</td>
</tr>
</tbody>
</table>

All translational values are in pixel, and rotational values are in rad/frame.

Figure 5. Sample images of the real image sequence.

Figure 6. The ranges of rotational parameters in the third experiment.

In the fourth experiment, with the real image sequence shown in Fig. 7, the parameters of the real images are as follows: the focal length is 820 pixels, and the image size is $481 \times 640$, the center of the image is at (305, 240) measured from the top left corner, the ground truth of motion parameters is computed from the database provided in [13]. The estimated result of motion parameters by using the approach mentioned in [2] is $\text{FOE} = (-151.138, -131.705)$ which is measured from the image center with the X-direction pointing to the right and the Y-direction pointing down, $w = [-0.00042, 0.00031, 0.000076]^T$; the ranges of the rotational parameters are shown in Fig. 8, and the estimated result obtained by making use of our approach is $\text{FOE} = (-137.505, -114.996)$, $w = [-0.00025, 0.00175, -0.00025]^T$.

Figure 7. Sample images of the real image sequence.
V. CONCLUSION

In this paper, we address the problem of ego-motion estimation for a monocular moving camera, which is under arbitrary translation and rotation. In comparison with indirect methods that rely on the optical flow, our approach is uniquely based on the spatial-temporal image derivatives, that is, the normal flow, and we do not need to make any assumption about the observed scenes. Comparing with other direct methods that rely on the normal flow, our approach can determine the ground truth of the rotational parameters and the location of FOE in the image plane simultaneously, and the determination of rotational parameters is based on whether the intersections of every two of the straight lines perpendicular to the normal flow vectors at those selected image points are converged at a point or not.

Our approach we propose in this paper works in two successive steps. First, we roughly determine the range of each rotational parameter, and are sure that the ground truth of each rotational parameter must be included in the corresponding range. Second, we search for the ground truth of each rotational parameter in the corresponding range. Once the ground truth of the rotational parameters is found, the location of the FOE in the image plane is determined simultaneously, which gives the direction of the camera linear velocity.

Learning from the results of the experiments above, we are sure that our approach is of high accuracy and robustness. According to the Table I, we learn that the estimated result of motion parameters is very close to the ground truth. Though the estimated result deviates from the ground truth in some extent, the errors can be accepted.

ACKNOWLEDGMENT

The work described in this paper was supported by National Natural Science Foundation of China (No.61005031).

REFERENCES