САМАРСКИЙ УНИВЕРСИТЕТ
SAMARA UNIVERSITY

ITNT - 2022

VII The International Conference on Information Technology and Nanotechnology, 20-24 September, Samara, Russia

Coplanarity-based approach for camera motion estimation invariant to the scene depth

Ye. V. Goshin

Abstract

- In this paper, we propose a method for estimating the parameters of camera movement from images obtained from this camera. This method is equally effectively applicable to flat and three-dimensional scenes. The proposed method allows avoiding the restrictions imposed on the set of initial data when using the fundamental matrix and the projective transformation matrix.

Theory

Consider the following statement of the problem. Let the set of points on the previous and current frames be given: $\{\mathbf{p}_i\} = \{(x_i, y_i, 1)\}$ and $\{\mathbf{p}'_i\} = \{(x'_i, y'_i, 1)\}$, obtained as a result of projecting points $\{\mathbf{M}_i\} = \{(X_i, Y_i, Z_i)\}$ of three-dimensional space and transformed points $\{\mathbf{M}'_i\} = \{(X'_i, Y'_i, Z'_i)\} = \{\mathbf{R}(X_i, Y_i, Z_i) + \mathbf{t}\}$ of three-dimensional space, respectively, defined with some error. It is necessary to calculate an estimate of the motion parameters $\mathbf{R}$ and $\mathbf{t}$.

Let us introduce the notation

$$x_i^R = R_{11}x_i + R_{12}y_i + R_{13},$$
$$y_i^R = R_{21}x_i + R_{22}y_i + R_{23},$$
$$z_i^R = R_{31}x_i + R_{32}y_i + R_{33}.$$

In previous work [Goshin Y. V. Estimating intrinsic camera parameters using the sum of cosine distances //Journal of Physics: Conference Series. – IOP Publishing, 2018. – Vol. 1096. – N. 1. – P. 012092], it was shown that a straight line in the plane passing through the points $\left(\frac{x_i^R}{z_i^R}, \frac{y_i^R}{z_i^R}\right)$ and (x'_i, y'_i) also passes through the point $\left(\frac{t_x}{t_z}, \frac{t_y}{t_z}\right)$. This requirement can be written as a triple product:

$$\left(\left(\frac{x_i^R}{z_i^R}, \frac{y_i^R}{z_i^R}, 1\right) \times (x'_i, y'_i, 1)\right) \cdot \left(\frac{t_x}{t_z}, \frac{t_y}{t_z}, 1\right) = 0.$$

Since multiplication by a constant does not change the equation, the following expression also will be true:

$$\left((x_i^R, y_i^R, z_i^R) \times (x'_i, y'_i, 1)\right) \cdot (t_x, t_y, t_z) = 0. \qquad (*)$$

Note that the vector in expression (1) does not depend on either the initial points or the rotation parameters. Let us decompose this expression into parts depending on $\mathbf{R}$ and $\mathbf{t}$.

We define vectors $\mathbf{v}_i$ as:

$$\mathbf{v}_i = (x_i^R, y_i^R, z_i^R) \times (x'_i, y'_i, 1)$$

and a matrix composed of these row vectors as:

$$\boldsymbol{V}(R) = \begin{pmatrix} - \\ \mathbf{v}_i \\ - \end{pmatrix}.$$

The resulting matrix $\boldsymbol{V}(R)$ has dimension $n \times 3$, where n is the number of pairs of initial corresponding points. The argument R in the notation is used to explicitly show that this matrix depends only on the rotation parameters.

It is obvious that the product of each row of the matrix $\boldsymbol{V}(R)$ by the vector $\mathbf{t}$ will be equal to zero (according to (*)) and, therefore, $\mathbf{t}$ will be the null vector of the resulting matrix:

$$\boldsymbol{V}(R) \cdot \mathbf{t} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Acknowledgements

The research was carried out within the state assignment theme 0777-2020-0017.

Algorithm

In general, the camera movement is present, so the vector $\mathbf{t}$ is assumed non-zero. Since the matrix has a non-zero null vector, its rank must be equal to 2 or less. In other words, vectors of matrix $\mathbf{v}_i$ should be coplanar.

Obviously, in real conditions, in the presence of noise in the initial data, coplanarity is impossible. Therefore, as a criterion for the correspondence of the rotation parameter to the initial data, we could instead estimate vector multicollinearity (in its geometrical meaning rather than that in statistics).

We propose to use any of the known criteria of multicollinearity as a criteria for the initial problem solution:

1. The minimum eigenvalue of the matrix $\boldsymbol{V}(R)$.
2. Condition number of the $\boldsymbol{V}(R)$.
3. Matrix determinant of the Gram matrix $\boldsymbol{V}^T(R)\boldsymbol{V}(R)$.

After determining the rotation parameters by optimizing the selected criterion $J(\mathbf{R})$

$$\hat{R} = \underset{R}{\operatorname{argmax}}\, J(\mathbf{R})$$

the value of the translation vector $\mathbf{t}$, up to the scale parameter, can be determined by solving an overdetermined homogeneous system of nonlinear equations, excluding the trivial solution:

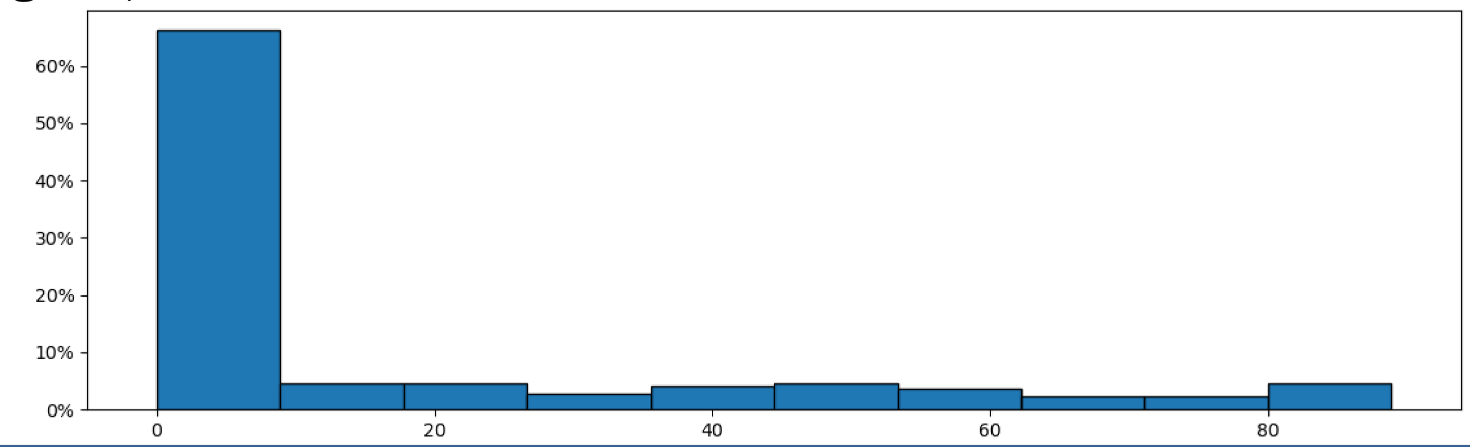
$$\boldsymbol{V}(\hat{\mathbf{R}}) \cdot \hat{\mathbf{t}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

It can be shown that for the points not located on a plane, the only value that satisfies the criterion is $\hat{\mathbf{R}} = \mathbf{R}$ and $\hat{\mathbf{t}} = \lambda\mathbf{t}$.

Experimental results

The proposed approach was tested on Blackbird dataset [Antonini, Amado & Guerra, Winter & Murali, Varun & Sayre-McCord, Thomas & Karaman, Sertac. (2020). The Blackbird UAV dataset. The International Journal of Robotics Research. 39. 027836492090833. 10.1177/0278364920908331].

It is shown that traditional approach using fundamental matrix estimation and decomposition gives about 95% wrong translation estimations (more than 45 degrees error). The proposed approach gives acceptable results for more than 75% of cases. In the following figure the distribution of error (in degrees) is shown.



Conclusion

The proposed approach for estimating the camera motion shows much more stable results for difficult scenes when number of corresponding points is low and most of these points lie on the same plane or in the same part of image.

FOR MORE INFORMATION:

Ye. V. Goshin [1] goshine@ssau.ru

[1] Samara National Research University,
Moskovskoye Shosse, 34, Samara, 443086, Russia

САМАРСКИЙ УНИВЕРСИТЕТ
SAMARA UNIVERSITY