Research on Visual Information Analysis Method of Moving Objects in Video

Longwei Chen, Yun Zhang, Qing He and Yang Sun*

College of Statistics and Mathematics, Yunnan University of Finance and Economics
Kunming, Yunnan, 650221, P.R. China

*Corresponding author

Abstract. For analyzing the size and distance of objects in video, we build a physical distance measurement model based on the geometric transformation of pixel and space points by different shooting environments and shooting methods, and realize the measurement of the size and distance of the physical object in the picture and the distance between the shooting point and the physical object in the picture.

Keywords: Camera imaging principle; Camera coordinate system; Rotation matrix.

1. Introduction

Extracting the size, distance, speed and other information of objects from images or videos is one of the important contents of visual information analysis [1]. In most cases, researchers can only determine the focal length of a camera from one photo and then analyze other information. According to the principle of camera imaging, the perspective projection model under monocular vision is established [2]. This model describes the transformation relationship between image pixels and points in the world coordinate system [3]. By iterating, fitting and establishing the equation, the unknown parameters of the rotation matrix are solved, and then the information in the image is analyzed. In the analysis of video, static graph is used to analyze the distance information in motion.

2. Establish the Projection Conversion Model of Pixel Point and 3D Coordinate Point

The working principle of camera imaging is mainly related to the transformation relationship among the imaging physical plane, pixel plane, camera and actual coordinate system [4]. By studying the transformation relationship between several coordinate systems, the transformation relationship between the pixel plane and the actual coordinate system can be deduced [5]. The matrix transformation formula between the pixel and the world coordinate system is obtained by superposing the matrix transformation under several different coordinate systems.

\[
\begin{bmatrix}
 u \\
 v \\
 1 
\end{bmatrix} = \begin{bmatrix}
 f / d_x & 0 & u_0 & 0 \\
 0 & f / d_y & v_0 & 0 \\
 0 & 0 & 1 & 0 
\end{bmatrix} \begin{bmatrix}
 r_{11} & r_{12} & r_{13} & t_x \\
 r_{21} & r_{22} & r_{23} & t_y \\
 r_{31} & r_{32} & r_{33} & t_z \\
 0 & 0 & 0 & 1 
\end{bmatrix} \begin{bmatrix}
 x_W \\
 y_W \\
 z_W \\
 1 
\end{bmatrix}
\]

(1)

Here, the specific expression of the rotation matrix \( R_{3\times3} \) of the camera is as follows:
In formula (1), \( f \) is the camera focal length, \((u, v)\) is the pixel coordinate, \((u_0, v_0)\) is the origin of the pixel coordinate system, \((x_w, y_w, z_w)\) is the world coordinate, \(d_x\) and \(d_y\) are the length and width of the pixel block, and \(z_c\) is the coordinate in the camera physical coordinate system. There are 11 unknowns in formula (2).

3. Establishing the Distance Model of Image Interior Point
When analyzing the actual distance between two points in the image, the distance model between two points in the picture was established by using the distance formula between two points in the actual space.

\[
D = \sqrt{(x_{w1} - x_{w2})^2 + (y_{w1} - y_{w2})^2 + (z_{w1} - z_{w2})^2}.
\]

4. Design Algorithm
There are 11 unknowns in the transformation matrix of the model. We select the corresponding points for iteration and fitting to solve the unknown parameters. Then the actual distance problem is solved based on two distance models.

4.1. Formula Deduction
In equation (1), let \( F = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \), its inverse matrix is \( F^{-1} = \begin{bmatrix} 1/f & 0 & -u_0/f \\ 0 & 1/f & -v_0/f \\ 0 & 0 & 1 \end{bmatrix} \). Let \( f_x = d_x/f \), \( f_y = d_y/f \), and let \( z_c d_x/f = f_1 \), \( z_c d_y/f = f_2 \), so equation (1) becomes the following equation:

\[
\begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} + \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \begin{bmatrix} f_1 & 0 & -u_0f_1 \\ 0 & f_2 & -v_0f_2 \\ 0 & 0 & z_c \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}.
\]

4.2. Output Parameter Value by Nested Fitting Method of Ergodic Cycle
In equation (4), the unknown parameters in the rotation matrix can be obtained by performing the ergodic calculation (triple cycle). The unknown parameters in the rotation matrix are traversed by \( \alpha, \beta, \gamma \in [-\pi/2, \pi/2] \) (triple cycle), and the translation vectors \( T = [t_x, t_y, t_z] \) and \( f_1, f_2, f_3, z_c \) of other related unknown quantities are all linear at one time, which can be fitted by the least square method through the corresponding relationship between the world coordinates and the pixel points (\( n \geq 6 \)). Its least square fitting function is as follows. In equation (5),

\[
\min G = \sum_{i=1}^{n} ([u_i - u_0]f_1 - t_x - R_1 \cdot W_i)^2 + ([v_i - v_0]f_2 - t_y - R_2 \cdot W_i)^2 + [z_i - t_z - R_3 \cdot W_i]^2
\]

It’s \( R_1 = [r_{11}, r_{12}, r_{13}] \), \( R_2 = [r_{21}, r_{22}, r_{23}] \), \( R_3 = [r_{31}, r_{32}, r_{33}] \). Form the normal equations of function \( G \), The values of \( f_1 \) and \( t_x \) can be solved by
the values of $f_2$ and $t_y$ can be solved in the same manner.

The relation $z_c - t_z = 1/n \sum_{i=1}^{n} R_i W_i$ can be obtained by fitting.

Because the physical size of the sensitive element is generally $dx \times dy = 5.2 \times 5.2 \mu m$ (micron), $f_1/f_2 = dx/dy$ can be seen from $z_c dx/f_1 = z_{c} dy/f_2$, so $dx \approx dy$. That is to make $|f_1 - f_2|$ as small as possible. Therefore, in the actual calculation, $C[f_1 - f_2]$ should be added to the objective function as the correction term, where $C$ is the weight coefficient. The following objective functions are obtained:

$$
\min G = \sum_{i=1}^{n} [(u_i - u_0) f_1 - t_x - R_i \cdot W_i]^2 + [(v_i - v_0) f_2 - t_y - R_2 \cdot W_i]^2 + [z_{c} - t_z - R_3 \cdot W_i]^2 + C[f_1 - f_2].
$$

(6)

$G = G(\alpha, \beta, \gamma), \quad \alpha, \beta, \gamma \in [-\pi/2, \pi/2]$, Find the minimum value through traversal, The optimal value is $(\alpha_0, \beta_0, \gamma_0)$, Fit the value of output parameter $f_1, f_2, t_x, t_y$, but can't get the value of $z_c, t_z, f$.

4.3. Establish Equations to Solve the Values of Unknown Parameters $z_c$ and $t_z$

According to the calculation, if the relation is fitted for the second time, the values of $z_c$ and $t_z$ cannot be fitted. It can be seen that the focal length of the lens, the size of the field of view and the distance between the lens and the object to be taken are calculated as $f = L w/W$ (In the formula, $f$ represents the focal length of the camera lens, $L$ represents the distance between the object to be photographed and the lens, $w$ represents the imaging width of the object in the image plane, and $W$ represents the actual width of the object).

According to the camera imaging diagram and the equations obtained by least square fitting, the following equations can be established:

$$
\begin{align*}
Z_c - t_z &= k_0 \\
\sqrt{OP} \cos \alpha \cos \beta &= Z_c \\
(t_x + w_x)^2 + (t_y + w_y)^2 + (z_{c}^2 - k_0^2) \cos^2 \alpha \cos^2 \beta &= z_{c}^2
\end{align*}
$$

(7)

In the formula, $k_0 = z_{c} - t_z = 1/n \sum_{i=1}^{n} R_i W_i$ , This formula is fitted by the above formula. From equation (10), the following equation is obtained:

$$(\cos^2 \alpha \cos^2 \beta - 1)t_z^2 + 2(w_x \cos^2 \alpha \cos^2 \beta - k_0)t_z + w_x^2 \cos^2 \alpha \cos^2 \beta + (t_x + w_x)^2 + (z_{c}^2 - k_0^2) - k_0^2 = 0$$

To solve the equation, the value of $t_z$ can be solved, and then the above equation can be solved.

5. Measuring the Actual Distance in Various Situations Based on the Model

5.1. Analyze the Actual Distance of Each Pixel in Static graph
5.1.1. Find the Distance between Two Cars in the Picture and the Distance between the Photographer and the Left Side of the Road. Set up the following world coordinate system in the given picture, select the following six points, as shown below; Use MATLAB to obtain the pixel coordinates of six points on the way, as shown in the following table.

We can know that the deflection angles between the camera coordinate system and the world coordinate system are \( \alpha = 1.5292 \), \( \beta = -0.0708 \), \( \gamma = -0.0708 \) respectively. The remaining parameters are \( f_1 = -1.7179 \), \( f_2 = -1.2024 \), \( f = 0.1888 \), \( x_c = 20.6000 \), \( y_c = -34.3554 \), \( z_c = -623.5860 \).

| points | \( u \) | \( v \) | \( x_w \) | \( y_w \) | \( z_w \) | points | \( u \) | \( v \) | \( x_w \) | \( y_w \) | \( z_w \) |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| A      | 466.5  | 638    | 470    | -180   | 70     | D      | 544.5  | 641    | 470    | 0      | 70     |
| B      | 463.5  | 686    | 470    | -180   | 0      | E      | 742.5  | 659    | 0      | 0      | 0      |
| C      | 547.5  | 695    | 470    | 0      | 0      | F      | 652.5  | 572    | 235    | -90    | 150    |

The unit of pixel block in the table is micron, and the unit of world coordinate is millimeter.

Finally, the distance between the two vehicles is 8.021075 meters.

5.1.2. Find the Distance and Height between the Photographer and the Sentry Box in the Picture. Six points are selected as the fitting data in the Figure 2. According to the algorithm, solve the equation, traverse the fitting to solve the value of each unknown parameter in the model, their values are respectively \( f_1 = -0.9126 \), \( f_2 = 1.6540 \), \( t_x = -1.2846 \), \( t_y = -92.6999 \), \( t_z = -1133.3 \),\( f = -0.2569 \), \( \alpha = -1.5708 \), \( \beta = -0.0708 \), \( \gamma = -1.0708 \), \( x_c = -5.6414 \), \( y_c = 1257.7 \), \( z_c = -1065.4 \). According to the model, the distance between the camera and the sentry box is 18 meters.

5.2. Analyze the Actual Distance Problem in Dynamic

5.2.1. Solving the Actual Distance of the Inner Point of the Image Based on the Specular Distance Model. Due to the specular reflection in the car rearview mirror, the distance between the light entering the camera is not equal to the distance between the reflector and the mobile phone object [6], so the distance model between the camera point and the pixel point is established based on the specular reflection as follows:
\[ D_j = D - 2L_{\text{back}} - L_{\text{vehicle}} \]  

In the static graph captured by MATLAB software, the following world coordinates are established and six points are selected. The pixel coordinates of these six points and the coordinates in the world coordinate system are shown in the following table.

On the basis of the designed algorithm, the value of each unknown parameter in the model can be obtained by solving the equation and traversing the fitting. They are respectively 
\[ f_1 = 0.5203, \quad f_2 = 0.6102, \quad t_x = -367.5048, \quad t_y = -55.1356, \]
\[ f = -0.3717, \quad \alpha = -1.5708, \quad \beta = -0.5708, \quad \gamma = -0.0708, \quad x_c = -53.9539, \quad y_c = 367.7437, \]
\[ z_c = -371.4707. \]

According to the model, the distance difference between the two vehicles can be calculated to be about 29.83 meters.

5.2.2. Analysis of the Actual Distance in the Train Video. The following static graph is selected, the world coordinate system is established and six points are selected. According to the designed algorithm steps, solve the equations set up according to the camera principle, traverse and fit based on the model, and solve the values of each parameter in the model, which are respectively 
\[ f_1 = 3.8008, \quad f_2 = 3.7080, \quad t_x = 408.1280, \quad t_y = 1.0145, \quad f = -0.1772, \]
\[ \alpha = -0.0708, \quad \beta = 0.9292, \quad \gamma = 0.9292, \quad x_c = -1250.9, \quad y_c = -207.6336, \quad z_c = -1295.1. \]

Based on the model and algorithm, the height of the bridge deck from the water surface is 5.306531 meters, the river width is 145.89 meters.

6. Conclusion

The projection model based on monocular vision pixels and three-dimensional space points established in this paper solves the problem of image information analysis in monocular imaging system. By iteratively fitting the values of each parameter, the distance between each point in the image and the distance between the camera and the target in the image are analyzed, and the problem of image distance under the condition of specular reflection and motion shooting is analyzed. The results are reliable and reflect the wide applicability of the model.

Acknowledgement

The authors gratefully acknowledges the support of the Yunnan Tobacco Company’s science and technology project “Research on Data Resource Mining Based on Multi-objective Decision Making”.

Reference

[1] LIU Xuejun, WANG Meizhen, ZHENG Yan, LU Yue. Geometric Measurement Base on Single Image: A Survey. J. Geomatics and Information Science of Wuhan University, 2011, 36(08):941-947.

[2] LAI Jiawei, HE Yuqing, LI Xiaopeng, LIU Mingqi, ZHAI Xiangyang. Design of manipulator target positioning system based on monocular vision. J. Optical Technique., 2019, 45(01): 6-11.

[3] LIU Jun, HOU Shihao, ZHANG Kai, YAN Xiaojuan. Vehicle distance measurement with implementation of vehicle attitude angle estimation and inverse perspective mapping based on monocular vision[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(13): 70-76.

[4] ZHAO Lianjun. Research on Mono-vision Pose Measurement based on Features of Target, D. Institute of Optics and Electronics Chinese Academy of Sciences, 2014.

[5] LU Jinwen. Research on Vehicle and Pedestrian Detection and Ranging Method Using Vision Fusion, D. South China University of Technology, 2018.

[6] YANG Xiujie, YAO Shujiao. Application of Car Rearview Mirror Imaging Based on OpenGL, J. Journal of Shenyang Ligong University. 2012, 31(01):40-43.