[0001]

The following section outlines a solution to the PnP problem.

[0002]

In order to solve the PnP problem, information about a view angle and a distortion of a camera which captured image data is required. In more detail, data called "camera matrix" and "distortion coefficient vector" are required.

[0003]

The camera matrix is a three-dimensional square matrix (matrix A below) including focal length fx and fy each expressed in a pixel unit, and a principal point (cx ,cy) which is usually an image center.

$$A = \begin{bmatrix} fx & 0 & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{bmatrix}$$

[0004]

In the above matrix A, a camera view angle is represented by fx and fy. For example, consider a case of a typical camera (pinhole camera model) where an image width is 2000 pixels and a horizontal angle of view is 60 degrees. In this case, "cx = 2000/2 = 1000" and "fx = 1/tan(60/2)*(2000/2) = around 1732".

[0005]

The camera matrix is a matrix necessary to find r11 to r33 and t1 to t3 that minimize an error in equation representing perspective projection transformation shown below in order to solve the PnP problem.

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

where (X,Y.Z) is a point coordinate in the 3D world coordinate system, (u, v) is a projection point coordinate in the 2D normalized image

coordinate system, r11 to r33 are elements of a rotation matrix, t1 to t3 are elements of a translation vector, and s is a constant factor.

[0006]

R11 to r33 and t1 to t3 obtained as a solution of the above equation (equation representing perspective projection transformation), correspond to an orientation of a camera (rotation matrix) and a position of the camera (translation vector).