CALIBRACION DE CAMARAS NO METRICAS

La orientación interior o determinación de los parámetros de una cámara es un problema clásico no sólo en la visión por computadora y la fotogrametría, sino también de las matemáticas. El conocimiento preciso de los parámetros de proyección de una imagen es un prerrequisito esencial para cualquier clase de medida geométrica cuantitativa en la visión por computadora. La calibración precisa de las cámaras es requerida para:

- La interpretación de imágenes 3D
- Reconstrucción de modelos tridimensionales de objetos o superficie del terreno (estereo-restitucion)
- Interacción espacial de robots con su entorno (coordinación ojo-mano)

Los parámetros verdaderos de proyección dependen de numerosos elementos técnicos y no son usualmente siempre provistos por los fabricantes de sistemas de captura y procesamiento de imágenes, además, para el caso de equipos compuestos de cámaras con teleobjetivos, los parámetros de proyección pueden ser variables.
Existen muchas aproximaciones para la calibración de cámaras, las mismas emplean estrategias diferentes basadas principalmente en el registro de una escena 3D. Estas metodologías pueden variar desde los métodos clásicos de proyección de rayos epipolares hasta el uso de una cantidad de marcas de un esquema especial calibrado 3D (patrón de calibración), donde la posición 3D de todos los puntos y el centro de la cámara (punto principal) es conocido. Otros acercamientos como el de Zhang's, utilizan múltiples vistas de un patrón de estructura 3D, pero de orientación y posición desconocidas en el espacio. Finalmente, existen los métodos de calibración que no emplean ninguna marca o plantilla ni suposiciones acerca de la estructura 3D de la escena, simplemente utilizan puntos de vista múltiples de estructuras arbitrarias rígidas, esto es designado comúnmente como "auto calibración". En la mayoría de estas metodologías los parámetros intrínsecos de la cámara y los parámetros extrínsecos, la estructura 3D que se observa, son recobrados conjuntamente. Basados en el modelo de imagen descripto a continuación, los siguientes parámetros pueden ser recobrados:

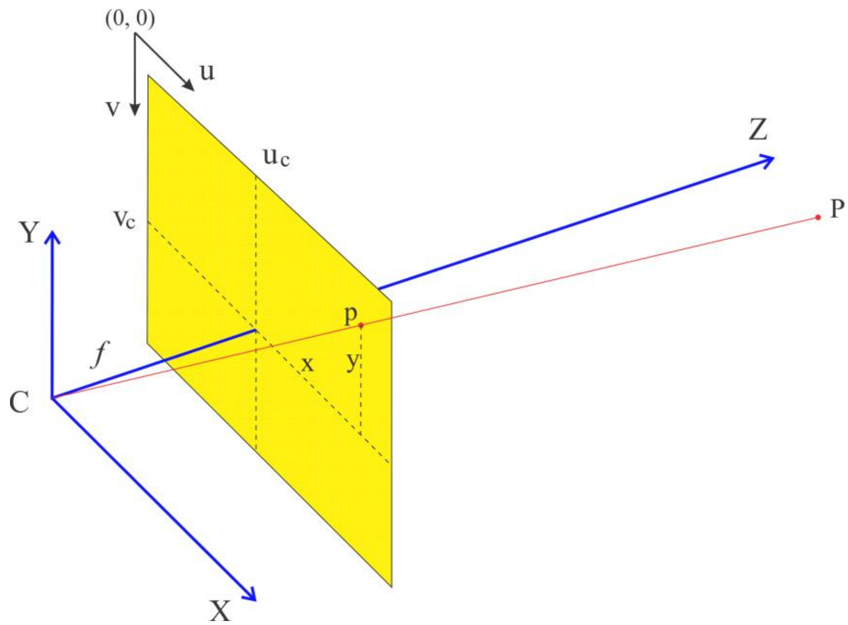
- Los parámetros intrínsecos de la cámara, las transformaciones interiores de la cámara, incluyendo distancia focal, la posición del punto principal, la escala del sensor y el desalineamiento (ortogonalidad del sensor).
- Los parámetros de la distorsión no lineal del lente (objetivo).
- Los parámetros de transformación externos (la rotación 3D y la traslación) para cada uno de los puntos de vista del patrón de referencia.

Modelo de proyección Perspectiva

Modelo Pinhole

El modelo de cámara adoptado es el modelo de imagen óptica estenopeica (pinhole) en condiciones ideales. Se describe a continuación, el proceso subyacente de proyección de puntos del espacio 3D a coordenadas 2D del plano del sensor (plano imagen).

Para el análisis se da por supuesto que el plano de imagen está situado delante del centro óptico, así la imagen no está invertida. El plano de imagen por tanto está situado a la distancia ***f*** del centro óptica, C = (0, 0, 0)^T, y es perpendicular al eje óptico. El centro óptica C es el origen del sistema de coordenadas 3D de la cámara.
El eje óptico se alinea con el eje Z del sistema de coordenadas y se intersecta con el plano imagen en (0, 0, *f*)^T.
Además, se da por supuesto que inicialmente el sistema de coordenadas de la cámara es idéntico al sistema de coordenadas de referencia (WCS), posteriormente se puede remover esta restricción y usar dos sistemas de coordenadas separados.



Un punto **P** en el espacio se sitúa sobre un rayo que atraviesa el centro óptico o centro de proyección C = (0, 0, 0) ^{T} y el punto correspondiente de la imagen **pi** = (x_i, y_i).
De triángulos similares, cada punto 3D, **P** = (X, Y, Z)^T, se producen las siguientes relaciones matriciales:

en notación vectorial

$$\begin{pmatrix} x_u \\ y_v \end{pmatrix} = \frac{f}{Z}\begin{pmatrix} \frac{X}{Y} \end{pmatrix} \quad (1)$$

Matriz de Proyección

La ecuación anterior (1) describe transformaciones no lineales en el dominio cartesiano de coordenadas. Usando coordenadas homogéneas, la transformación perspectiva puede ser escrita como una ecuación matricial lineal.

$$(x.y) \rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad (x, y, z) \rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Coordenadas imagen homogéneas Coordenadas escena homogéneas

La proyección perspectiva no es uno a uno, ejemplo:

Punto objeto (10, 6, 4), d=2

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/2 & 0 \end{bmatrix} \begin{bmatrix} 10 \\ 6 \\ 4 \\ 1 \end{bmatrix} = [10 \quad 6 \quad -2]$$

X' = -5; y' = -3

Punto objeto (25, 15, 10), d=2

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/2 & 0 \end{bmatrix} \begin{bmatrix} 25 \\ 15 \\ 10 \\ 1 \end{bmatrix} = [25 \quad 15 \quad -5]$$

X' = -5; y' = -3

Dado que la ecuación (1) es una transformación no lineal (división por Z), la transformación perspectiva en coordenadas homogéneas y notación matricial resulta:

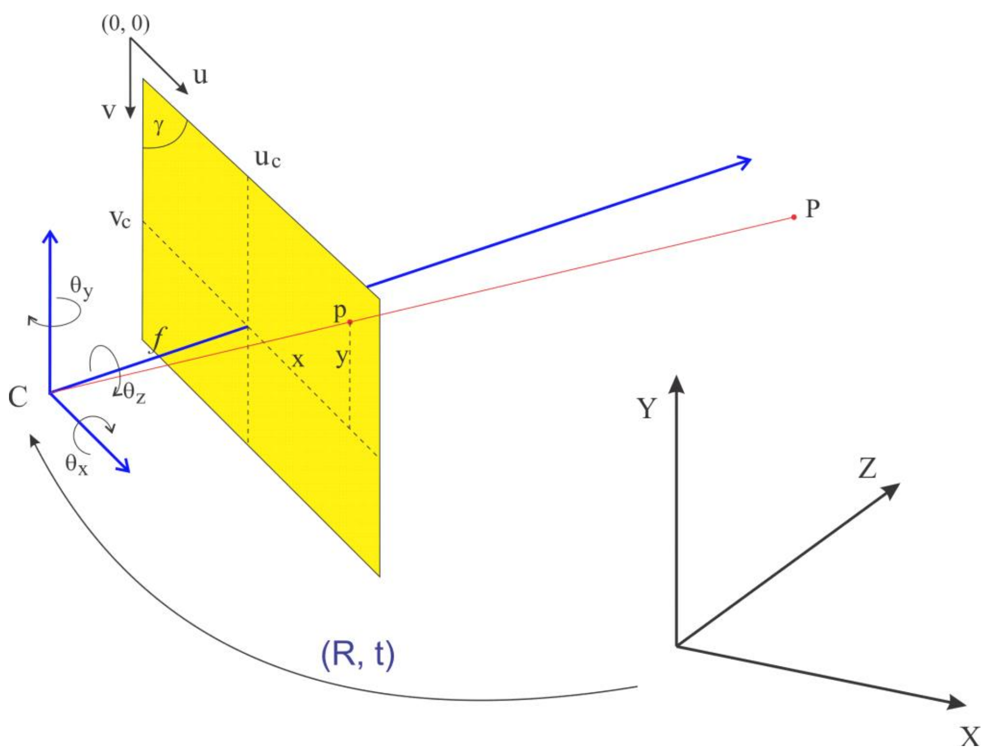
$$\begin{pmatrix} x_u \\ y_v \\ 1 \end{pmatrix} = \begin{pmatrix} f\,\frac{X}{Z} \\ f\,\frac{Y}{Z} \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} x_u \\ y_v \\ 1 \end{pmatrix} = \begin{pmatrix} f\,\frac{X}{Z} \\ f\,\frac{Y}{Z} \\ 1 \end{pmatrix} = \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Donde la matriz de proyección M_P puede ser descompuesta en dos matrices de la forma:

$$\text{Mp} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} x \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Esta matriz por lo tanto representa el modelo ideal de la cámara estenopeica (Pinhole), de distancia focal ***f***, y el modelo estándar (o canónico) de la matriz de proyección, que le corresponda a la geometría de una simple vista (eje óptico alineado con el eje Z).
Si la cámara tiene su propio sistema de coordenadas (foco canónico), y observa en forma bidimensional una escena tridimensional (3D), correspondiente a los puntos de un cuerpo rígido en el espacio, se debe tomar en consideración, además, que, por estar la cámara en el espacio, la transformación proyectiva será modificada por la rotación y traslación. De esta forma ahora un punto P' en un lugar 3D será modificado de su posición original por la transformación homogénea del punto, P' = hom (P).



Transformación de coordenadas espaciales (WCS) a la cámara: dado un punto (X, Y, Z) expresado en coordenadas 3D, su posición en el sistema de coordenadas 3D de la cámara será:

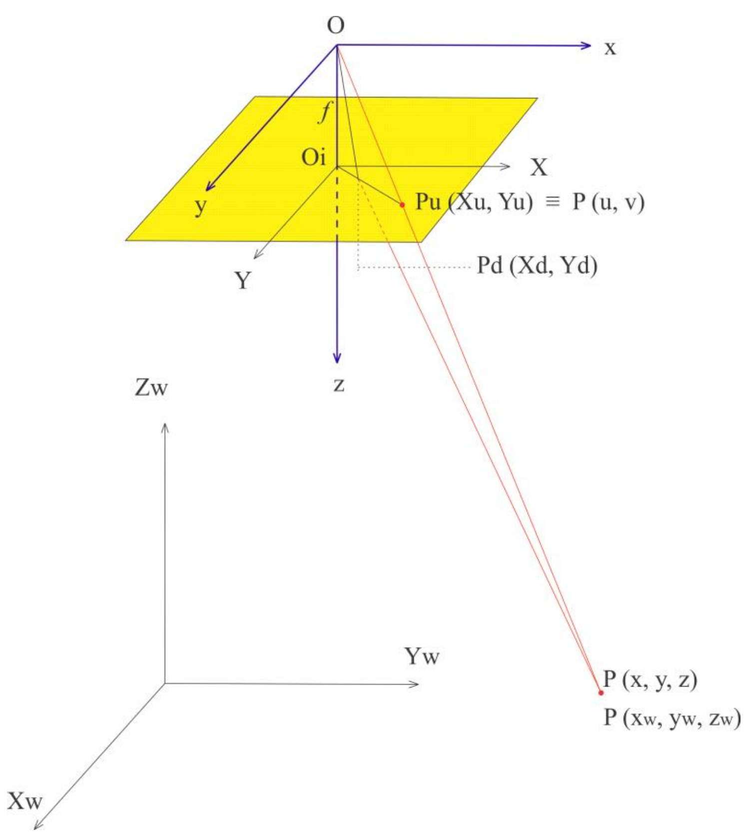
$$\begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Proyección sobre el plano imagen normalizado (ideal): la proyección perspectiva desde el punto 3D en coordenadas 3D de la cámara ***P'*** = (*X', Y', Z'*), sobre las coordenadas continuas normalizadas 2D del plano imagen de la cámara será:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{Z'}\begin{pmatrix} X' \\ Y' \end{pmatrix} = hom^{-1}\left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix}\right]$$

Modelo completo de proyección (transformación perspectiva)

Dado un punto P en el espacio y considerando un sistema de coordenadas con origen en O, WCS ($O_w - X_w\, Y_w\, Z_w$) coincidente con el eje Z correspondiente al eje de la cámara (Tsai); el punto P en el sistema resulta implícito: P = (x_w, y_w, x_w), para el cual corresponden las coordenadas imagen P_u = (X_u, Y_u) o habitualmente (u, v). Hay ademas un factor de distorsión radial (tangencial) de la lente (P_d) no considerado, en realidad este factor de distorsión corresponde al conjunto de lentes que forman el objetivo de la cámara. Basado en el modelo de imagen descripto, el siguiente grafico representa un punto en el espacio 3D y como puede ser obtenido en el plano de imagen:



La transformación completa resulta:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R \quad t] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & u_c \\ 0 & f_y & v_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

Donde λ refiere a un factor de escala ($\neq 0$); $\boldsymbol{K}$ es llamada matriz de los parámetros intrínsecos de la cámara; ($\boldsymbol{u_c}, \boldsymbol{v_c}$) es el punto principal; $\boldsymbol{s}$ es el factor de distorsión; $\boldsymbol{f_x}$ $\boldsymbol{f_y}$ representan la distancia focal en la dirección del eje $\boldsymbol{u}$ y la dirección del eje $\boldsymbol{v}$ respectivamente; $\boldsymbol{r_{ij}}$ es la matriz de rotación (3 x 3) y $(\boldsymbol{t_1}, \boldsymbol{t_2}, \boldsymbol{t_3})^{\boldsymbol{T}}$ es el vector traslación.
Si el factor de distorsión es 0, la matriz $\boldsymbol{K}$ es igual a:

$$\begin{bmatrix} f_x & 0 & u_c \\ 0 & f_y & v_c \\ 0 & 0 & 1 \end{bmatrix}$$

Para hacer una transformación perspectiva de la imagen útil como un modelo para cámaras reales, se necesita definir además, el mapa de coordenadas continuas x, y en el plano imagen, definidas por la matriz de pixeles respecto a la posiblemente diferente escala del sensor en la dirección x e y ($\boldsymbol{S_x}, \boldsymbol{S_y}$), la posición del centro de imagen o punto principal ($\boldsymbol{uc}$, $\boldsymbol{vc}$) con relación al sistema de coordenadas de la imagen ($\boldsymbol{0}, \boldsymbol{0}$), y la distorsión diagonal $\boldsymbol{s}$ del plano imagen (el cual usualmente es insignificante o cero).

$$s = u \, cotg \, \gamma$$

Teniendo en cuenta estos adicionales, los parámetros intrínsecos de la cámara y los parámetros extrínsecos, se obtiene la perspectiva completa de transformación de la imagen:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R \quad t] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} f_x s_x & s & u_c \\ 0 & f_y s_y & v_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (2)$$

Cuando un objeto se mueve en el espacio 3D sin cambiar su tamaño o forma, está sujeto

a una transformación rígida, cambiando orientación y posición. Puede observarse en esta formula que hay 5 parámetros intrínsecos y 6 parámetros extrínsecos, si la distorsión $\boldsymbol{s}$ es cero se reduce a cuatro los parámetros intrínsecos.
La matriz $\boldsymbol{R}$ describe la rotación arbitraria 3D compuesta de tres individuales rotaciones sobre cada uno de los ejes X, Y, Z, llamadas R_x, R_y, R_z respectivamente:

$$R = R_x . R_y . R_z = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{pmatrix}$$

$$R_y = \begin{pmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{pmatrix}$$

$$R_z = \begin{pmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$R = \begin{bmatrix} \cos\theta_y \cos\theta_z & -\cos\theta_y \sin\theta_z & \sin\theta_y \\ \sin\theta_x \sin\theta_y \cos\theta_z + \cos\theta_x \sin\theta_z & \cos\theta_x \cos\theta_z - \sin\theta_x \sin\theta_y \sin\theta_z & -\sin\theta_x \cos\theta_y \\ -\cos\theta_x \sin\theta_y \cos\theta_z + \sin\theta_x \sin\theta_z & \sin\theta_x \cos\theta_z + \cos\theta_x \sin\theta_y \sin\theta_z & \cos\theta_x \cos\theta_y \end{bmatrix}$$

Hay muchos métodos para resolver la ecuación (2), uno de los más usados es el método de calibración de cámaras de Zhang's el cual emplea unas pocas (al menos dos) vistas de un patrón planar de calibración llamado modelo o "target" del cual su trazado y las dimensiones métricas son precisamente conocidas. Este modelo es simplemente una plantilla cuadriculada (tablero de ajedrez) bidimensional; las imágenes se toman desde diferentes puntos de vista ya sea moviendo el modelo o la cámara. De cada imagen se miden en pixeles los puntos que registra el sensor, asumiendo que cada uno de estos puntos se corresponde con el modelo de calibración. Luego para cada punto observado se asocia una homografía.

Desde las imágenes homografícas Hi, se pueden recobrar los cinco parámetros intrínsecos (***f_x, f_y, s, u_c, v_c***) de la cámara que son estimados usando una solución de la forma cerrada (lineal), ignorando cualquier distorsión del lente en este punto. Tres o más imágenes permite obtener una solución única (hasta un factor de escala indeterminado). Si asumimos que el plano del sensor está sin desalineamiento (***s*** = 0, lo cual es una suposición razonable), luego dos imágenes resultan suficientes, más puntos de vista (imágenes), generalmente conducen a resultados más precisos. Entre los varios métodos para estimar homografías la transformación lineal directa es la más usada:

Coordenadas homogéneas

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \left[\begin{pmatrix} f_x & 0 & u_c \\ 0 & f_y & v_c \\ 0 & 0 & 1 \end{pmatrix} . \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} . \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \right]$$

Si se considera que la ecuación anterior es una transformación proyectiva tridimensional, la tercera componente del vector que define la imagen del punto en el sistema interno es la distancia focal de la cámara, ***z*** = -***f***, dividiendo las dos primeras ecuaciones por la tercera resulta en coordenadas cartesianas un par de ecuaciones no lineales:

$$u = r_{11}X_w + r_{12}Y_w + r_{13}Z_w + T_x$$

$$v = r_{21}X_w + r_{22}Y_w + r_{23}Z_w + T_y$$

$$-f = r_{31}X_w + r_{32}Y_w + r_{33}Z_w + T_z$$

Para ***i*** puntos imagen:

$$u_i - u_c = f_x \frac{r_{11}X_i + r_{12}Y_i + r_{13}Z_i + T_x}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_z}$$

$$v_i - v_c = f_y \frac{r_{21}X_i + r_{22}Y_i + r_{23}Z_i + T_y}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_z}$$

Resolver estas ecuaciones requiere linealizarlas para obtener los parámetros intrínsecos por diversos métodos matemáticos complejos de estimación. Si se considera que el modelo de calibración es una plantilla bidimensional colocada frente a la cámara la componente $\boldsymbol{Z}$ es igual a cero por lo tanto estas ecuaciones resultan:

$$u_i - u_c = f_x \frac{r_{11}X_i + r_{12}Y_i + T_x}{r_{31}X_i + r_{32}Y_i + T_z}$$

$$v_i - v_c = f_y \frac{r_{21}X_i + r_{22}Y_i + T_y}{r_{31}X_i + r_{32}Y_i + T_z}$$