

Auto-calibración de un sistema binocular de visión activa

Antonio Sánchez
Dept. Ingeniería de Sistemas y Automática. Universidad Politécnica de Valencia
Camino de Vera. Apdo. 22012 E-46071 Valencia,
e-mail: asanchez@isa.upv.es

Ricardo Calatayud
C/ Francisco Baixauli, 6. Alfafar, Valencia,
e-mail: ricaga@inf.upv.es

Resumen

El objetivo principal de este artículo es implementar un proceso de calibración de un sistema binocular de visión activa que ofrezca las características de autonomía, eficiencia, versatilidad y precisión necesarias para poder realizar dicho proceso en línea. Este sistema se ha desarrollado con Visual C++ bajo Windows 98, para ello se han implementado métodos de detección de esquinas, correspondencias y estimación de la matriz fundamental.

Palabras Clave: Auto-calibración, visión activa, visión 3D, detección de esquinas, correspondencia.

1 INTRODUCCIÓN

La calibración de una cámara en el contexto de sistemas de visión tridimensionales, es un paso necesario para poder extraer información métrica a partir de imágenes 2D. El proceso de calibración permite determinar las características ópticas de la cámara (parámetros intrínsecos) así como su localización (parámetros extrínsecos).

Matemáticamente se puede representar la cámara mediante el modelo de orificio que modela la cámara por su plano retinal y su centro óptico. Dado un punto en el espacio 3D (w) y dada la proyección de dicho punto en el plano (m) la matriz que permite expresar la proyección es mT_w .

$$k \cdot m = {}^mT_w \cdot w \quad (1)$$

donde k es un factor de escala denominado profundidad proyectiva. Podemos descomponer la matriz mT_w utilizando una factorización QR, con lo que se obtiene:

$${}^mT_w = A \cdot [R|t] \quad (2)$$

donde A , que tiene la forma de una matriz triangular superior, depende únicamente de los parámetros intrínsecos de la cámara y R expresa los parámetros extrínsecos de la cámara (la posición y orientación).

En el caso de un sistema de visión activa, donde los parámetros de las cámaras cambian continuamente, no es posible realizar una calibración clásica [5] donde se necesita utilizar una plantilla de punto 3D conocidos. En este caso, utilizando el mismo modelo de cámara de orificio se puede:

1) calibrar las cámaras en todas las configuraciones de la cabeza, mediante un procedimiento fuera de línea, almacenando los resultados en una tabla que se accederá durante el funcionamiento del sistema según la configuración actual de la cabeza. Algunos trabajos puntuales realizados en esta línea son [6], [1].

2) calcular la estructura proyectiva en línea utilizando una calibración débil con la matriz fundamental, que aunque no contiene información métrica, contiene información como coplanaridad, colineaciones y relaciones de las relaciones de las distancias.

Nuestro sistema es un sistema binocular de visión activa con siete grados de movilidad mecánicos y tres ópticos por cámara (fig. 1).

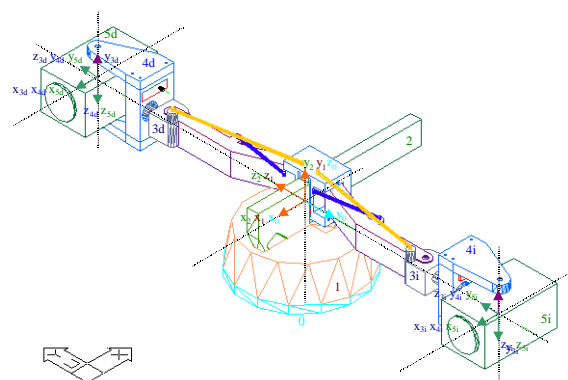


Figura 1: Sistema binocular de visión activa SiviS.

El procedimiento de calibración que se va a utilizar consiste en una primera fase de extracción de características de las dos imágenes, una fase de búsqueda de correspondencias entre las listas de características de cada imagen y finalmente el cálculo de la matriz fundamental a partir de los pares de características homólogas.

2 CARACTERÍSTICAS

En la fase de extracción de características se pretende encontrar características con la propiedad de unicidad en las imágenes para facilitar la fase posterior de correspondencias. En este sentido, los píxeles que más garantías de unicidad ofrecen son las esquinas de ambas imágenes.

2.1 DETECCIÓN DE ESQUINAS

Para la detección de esquinas se utiliza el algoritmo de SUSAN [4] acrónimo de Smallest Univalve Segment Assimilating Nucleus.

El algoritmo de SUSAN requiere el tratamiento con una imagen de grises. En primer lugar, será necesario calcular una tabla T que marcará el valor que se debe asignar a la diferencia de nivel de gris de dos píxeles. El cómputo de la tabla se realiza en base a la siguiente fórmula:

$$T[\textit{posición}] = 100.0 \bullet e^{-(\textit{diferencia}/\textit{umbral})^6} \quad (3)$$

donde diferencia tomará valores entre -256 y 257 , y umbral es una variable que determinará la máxima diferencia de nivel de gris entre dos píxeles de forma que ambos píxeles se consideren pertenecientes al mismo objeto.

Obtenida la tabla, se continua con la segunda etapa del algoritmo en la que se realiza un barrido sobre la imagen de izquierda a derecha y de arriba abajo, calculando el resultado de las diferencias de brillos de los píxeles según la máscara (fig. 2). La característica más destacable de esta máscara es ser invariable a la rotación.

```

000
00000
0000000
000X000
0000000
00000
000

```

Figura 2: Máscara de detección de esquinas.

El algoritmo de detección de esquinas acumulará los valores de la tabla precalculada para las diferencias de brillos entre el píxel central de la máscara (marcado con una aspa roja en la máscara) y cada uno de los píxeles tratados. Tras el cómputo de la suma de los valores de la tabla se comprobará si se ha superado un cierto umbral, en cuyo caso no existirá esquina. Asegurada la posible existencia de la esquina, se deberá superar todavía un segundo filtro que asegure que es un máximo local. En la figura 3 se aprecia el resultado de aplicar este preproceso.



Figura 3: Detección de esquinas.

3 CORRESPONDENCIAS

Una correspondencia es un par de características homólogas (m, m') que son proyección del mismo punto del espacio en tres dimensiones.

Una vez se dispone de la lista de esquinas en cada una de las imágenes, se recorren todas las esquinas detectadas en una imagen y se calcula, utilizando reconocimiento de patrones, el porcentaje de similitud de cada una de éstas con todas las esquinas de la otra imagen. Para ello se extraerá una pequeña subimagen en cada una de las esquinas a tratar. En el momento en el que se ofrezca una cota mínima de similitud, se aceptará la pareja de esquinas como par homólogo.

Cabe destacar que se está aplicando técnicas heurísticas para buscar la pareja de esquinas que se corresponden, por lo tanto se puede encontrar falsas correspondencias, lo cuál influirá negativamente en la precisión del proceso de autocalibración.

4 MATRIZ FUNDAMENTAL

El estudio de la matriz fundamental fue introducido por Faugeras [2] como una generalización de la matriz esencial [3] para cámaras no calibradas. La matriz fundamental contiene información de la geometría epipolar (fig. 4) relativa a las dos cámaras de un sistema binocular, concretamente determina la

posición de los epipolos (e, e') y la transformación que relaciona las rectas epipolares entre las imágenes. Esta matriz se calcula a partir de un determinado número de pares de puntos correspondientes entre las imágenes.

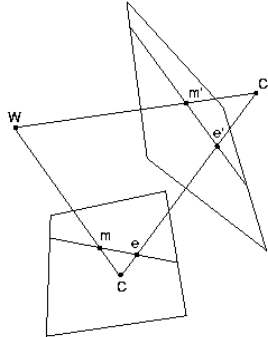


Figura 4: Restricción epipolar.

Dado un par de características homólogas (m, m') calculadas en la fase anterior, la matriz fundamental F cumple la siguiente ecuación.

$$m^T \cdot F \cdot m' = 0 \quad (4)$$

Por lo tanto, para estimar la matriz fundamental se debe resolver el siguiente sistema:

$$U_n \cdot f = 0 \quad (5)$$

donde $u_i = [u_i, v_i, 1]$ y $f = [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]$

La matriz F es una matriz 3×3 de rango 2, lo cual implica que de sus 9 parámetros sólo tiene 7 parámetros independientes entre sí.

Existen diferentes técnicas que permiten calcular la matriz fundamental [7]. Entre ellas, cabe destacar:

1.- Solución exacta con 7 pares de puntos: técnica basada en el hecho de conocer 7 pares de puntos obteniendo dos vectores f_1 y f_2 con los que se obtienen las matrices F_1 y F_2 y con las que se resolverá la siguiente ecuación:

$$\det[\mathbf{a} \cdot F_1 + (1 - \mathbf{a}) \cdot F_2] = 0 \quad (6)$$

de donde se obtendrá F como sigue:

$$F = \mathbf{a} \cdot F_1 + (1 - \mathbf{a}) \cdot F_2 \quad (7)$$

2.- Método analítico con 8 o más pares de puntos: técnica que utiliza una resolución mediante el método de mínimos cuadrados.

3.- Método analítico con rango 2: impone el rango 2 durante la minimización, aunque la solución obtenida depende de los valores de f_{32} , y f_{33} .

4.- M-estimators: intenta utilizar una función diferente a r_i^2 para conseguir que la solución no sea linealmente dependiente de los datos. Se basa en el uso de pesos en los datos según su proximidad a la solución. De esta forma si un dato se aleja de la recta de la solución su contribución a la solución final será menor que la de un punto que esté en el conjunto de puntos válidos. De esta forma se pueden detectar los puntos que no estén cerca de la recta y catalogarlos como falsas correspondencias

5.- Least Median of Squares (LmedS): estima los parámetros resolviendo un problema de minimización no lineal:

$$\min_i \text{mediana } r_i^2 \quad (8)$$

La técnica seleccionada para calcular F es M-estimators. Esta técnica intenta minimizar la siguiente ecuación:

$$\min \sum_i \rho(r_i) \quad (9)$$

donde ρ es simétrica y definida positiva, con un único mínimo en 0.

Es posible implementar un método iterativo de mínimos cuadrados con pesos asociados.

Se intentará resolver la siguiente ecuación:

$$\sum_i \mathbf{y}(r_i) \frac{\partial r_i}{\partial f_j} = 0 \quad (10)$$

donde la ecuación derivada se denominada función de influencia.

$$\mathbf{y}(x) = d\mathbf{r}(x)/dx \quad (11)$$

La función de pesos que se utiliza se define como:

$$w(x) = \frac{\mathbf{y}(x)}{x} \quad (12)$$

y el sistema final queda:

$$\min \sum_i w(r_i^{(k-1)}) r_i^2 \quad (13)$$

el superíndice marca el número de iteración y la función de los pesos w , se debería calcular después de cada iteración.

Para la reestimación de los residuos se utilizará la función “Tukey” descrita a continuación:

$$r_i = \begin{cases} \frac{c^2}{6} \left(1 - \left[1 - \left(\frac{r_i}{c \cdot s} \right)^2 \right]^3 \right) & \text{si } |r_i| \leq c \cdot s \\ c^2/6 & \text{sino} \end{cases} \quad (14)$$

que se corresponde con la siguiente función de pesos:

$$w_i = \begin{cases} \left[1 - \left(\frac{r_i}{c} \right)^2 \right]^2 & \text{si } |r_i| \leq c \cdot s \\ 0 & \text{sino} \end{cases} \quad (15)$$

donde el valor de σ para se calcula como sigue:

$$s = 1.4826 \left[1 + \frac{5}{(n-f)} \right] \cdot \text{mediand}_i |r_i| \quad (16)$$

5 RESULTADOS

La fase experimental del trabajo que se presenta en este artículo se ha realizado en el laboratorio de Robótica del departamento de Sistemas y Automática de la Universidad Politécnica de Valencia.

En este laboratorio se dispone de un sistema de visión binocular de visión activa compuesto por dos cámaras Sony EVI-401DR junto con la tarjeta Matrox Meteor-II/MultiChannel con el que se han adquirido las siguientes imágenes (fig. 5).

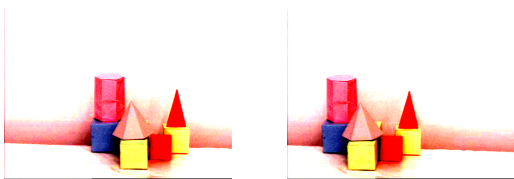


Figura 5: Imágenes originales.

El primer preproceso a realizar es la extracción de esquinas. En la figura 6 se muestran las esquinas, marcadas por una pequeña cruz verde, sobre una porción de la imagen derecha.

Posteriormente se generan las correspondencias entre las listas de esquinas de cada imagen. El proceso de correspondencias es crítico, ya que es costoso y además puede generar falsas correspondencias. En los experimentos realizados se ha restringido mucho

la aceptación de correspondencias para maximizar la fiabilidad.

Y finalmente, se estima la matriz fundamental utilizando el método M-estimators. Este método es robusto ante errores de localización de los pares de correspondencias.

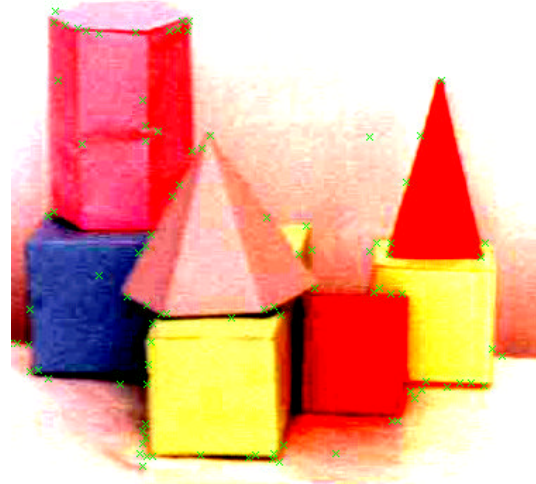


Figura 6: Extracción de esquinas.

6 CONCLUSIONES

Para determinar la bondad de un método de calibración hay que tener en cuenta una serie de criterios como: *autonomía*, el proceso de calibración no tiene que requerir intervención del usuario; *eficiencia*, debe tener un coste computacional aceptable; *versatilidad*, debe ser independiente del contexto y finalmente *precisión*, debe minimizar los errores de estimación de los parámetros.

El proceso de calibración propuesto es totalmente autónomo, es decir, este proceso no requiere la intervención por parte de un usuario para calcular la matriz fundamental del sistema de visión binocular.

El coste temporal de este proceso para un par de imágenes de 800x600 píxeles es de aproximadamente 3 segundos.

Es un proceso versátil ya que utiliza las esquinas como características de correspondencia. Las esquinas son características presentes en muchos tipos de escenas. Estas características se encuentran con gran facilidad en entornos artificiales, pero también pueden encontrarse sin problemas en entornos naturales. Por lo tanto, la cantidad de escenas que permiten calibrar el sistema binocular es muy amplio debido a que la cantidad de pares de esquinas necesaria es reducida, (sólo se necesitan

siete pares para calcular la matriz fundamental). Generalmente se dispondrá de más de siete pares para estimar la matriz recurriéndose a métodos de optimización para resolver el sistema.

Las características de autonomía, eficiencia y versatilidad permiten la calibración en línea del sistema binocular.

Por último, la característica más importante de precisión se conseguirá si se dispone de un método de optimización robusto ante posibles errores de localización de las esquinas.

Los métodos de optimización de mínimos cuadrados no son robustos ante errores de localización. Con estos métodos un “outlayer” puede producir una estimación errónea de la matriz fundamental. Por lo tanto, estos métodos no obtienen buenos resultados de precisión.

Para evitar falsas correspondencias es necesario disponer de un proceso de extracción de características estable y de un sistema de correspondencias fiable. El proceso de detección de esquinas es estable, sin embargo el proceso de correspondencias no es cien por cien fiable, por lo tanto el método de cálculo de la matriz fundamental deberá filtrar los posibles “outlayers”.

Para obtener mejores precisiones se ha utilizado un método de mínimos cuadrados iterativo con pesos denominado “M-estimators”. Este método intenta eliminar los “outlayers” consiguiéndose mejores precisiones.

Otro problema que se presenta a la hora de resolver el sistema de ecuaciones para calcular la matriz fundamental es que el sistema esté mal condicionado. Para intentar evitar este problema se puede utilizar otros métodos más robustos como Householder para resolver el sistema de ecuaciones.

7 REFERENCIAS

- [1] Chen Y., Shih S., Hung Y. and Fuh C. (2000) Camera calibration with a motorized zoom lens, 15th International Conference on Pattern Recognition. IEEE Computer Society, 4: pp. 495-498.
- [2] Faugueras O., Luong Q.-T. and Maybank S. (1992) Camera self calibration : theory and experiments, Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita, Italy, 588: pp 321-334.
- [3] Longuet-Higgins H. (1981) A computer algorithm for reconstructing a scene from two projections, Nature, 293: pp. 133-135.
- [4] Smith S. M., Brady J.M. (1995) SUSAN A new approach to low level image processing.
- [5] Tsai, (1986) An efficient and accurate camera calibration technique for 3D machine vision, Proceedings IEEE Computer Vision and Pattern Recognition.
- [6] Weckesser P. and Wallner F. (1994) Calibrating the active stereo vision system Kastor for real-time robot navigation, Close Range Techniques and Machine Vision. ISPRS Commission V, pp. 430-436.
- [7] Zhang Z. (1998) Determining the epipolar geometry and its uncertainty: A review, International Journal of Computer Vision.