

UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR



**AUTOCALIBRACIÓN Y
SINCRONIZACIÓN DE MÚLTIPLES
CÁMARAS PTZ**

-PROYECTO FIN DE CARRERA-

Javier García Ocón
Mayo de 2007

AUTOCALIBRACIÓN Y SINCRONIZACIÓN DE MÚLTIPLES CÁMARAS PTZ

AUTOR: Javier García Ocón
TUTOR: Jesús Bescós Cano

Grupo de Tratamiento de Imágenes
Dpto. de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid

PROYECTO FIN DE CARRERA

Título: *Autocalibración y sincronización de múltiples cámaras PTZ.*

Autor: D. Javier García Ocón

Tutor: D. Jesús Bescós Cano

Tribunal:

Presidente: D. José María Martínez Sánchez

Vocal: D. Miguel Ángel García

Vocal secretario: D. Jesús Bescós Cano

Fecha de lectura:

Calificación:.

Palabras clave

Calibración, Autocalibración, Matriz de Proyección, Zhang, parámetros intrínsecos, parámetros extrínsecos.

Resumen

El contexto de este proyecto es el conjunto de técnicas orientadas a extraer cuadros o información clave de secuencias de vídeo obtenidas de cámaras fijas. En este caso se analizará la problemática asociada a tres cámaras PTZ (*Pan-Tilt-Zoom*) situadas en el vestíbulo de entrada del edificio A de la EPS, y visibles las tres entre sí.

Para poder guiar la selección de forma eficiente es necesario conocer realmente el significado de la información que muestran las imágenes y, en particular, poder realizar mediciones precisas de la geometría de la escena captada . Esto exige el desarrollo de técnicas de calibración o autocalibración. Así mismo, la obtención de información complementaria de varias cámaras permite aumentar la precisión y robustez de esta tarea. Este segundo aspecto exige la sincronización total o parcial de las cámaras.

El objetivo de este proyecto es analizar y aplicar distintas técnicas de calibración de cámaras fijas y autocalibración de cámaras PTZ para así poder realizar mediciones precisas del tamaño de los objetos presentes en la escena y de sus trayectorias, con aplicación en entornos de seguridad.

Abstract

The context of this project is the set of techniques oriented to extract frames or key information of video sequences obtained by means of fixed cameras. In this case, we will analyse the problem related with three PTZ (*Pan-Tilt-Zoom*) cameras placed in the hall of the EPS building A, all visible among them.

We will need to know really the meaning of the information provided by the images for being able to guide the selection in an efficient way and, specially, for being able to do accurate measures of the captured scene geometry. That needs the development of calibration or autocalibration techniques. Also, the acquisition of complementary information from several cameras increases the accuracy and strength of this task. It also needs the total or partial synchronization of the cameras.

The objective of this project is to analyse and apply several calibration techniques for fixed cameras and autocalibration for PTZ cameras. This way, we will obtain accurate measures of the scene objects size and trajectories, specially applied to security environments.

A mi tutor, Jesús Bescós, por su dedicación y ayuda en este proyecto. También en agradecimiento por la ayuda que, junto al profesor José María Martínez, nos han prestado durante toda la carrera a los alumnos.

A los compañeros del laboratorio, que me han prestado su ayuda. En especial a Álvaro (el hombre que sostenía tableros de ajedrez) y a Juan Carlos (el que lo grababa todo).

A Javi y a Kike, que dedicaron parte de sus horas felices a grabar tableros de ajedrez.

A José Ignacio Ronda, por su inestimable ayuda en estos terrenos de la calibración.

Por supuesto a mi tablero de ajedrez, compañero infatigable de calibraciones.

A mi madre, que me ha apoyado durante la carrera y, durante el último año, me ha preguntado todos los días, con infatigable constancia la misma pregunta: "...¿Y cuando lees el proyecto?" Obteniendo siempre la misma respuesta: "Esta tarde madre".

A mi padre, que me ha ayudado y enseñado muchas cosas de este mundo de los teleñecos desde el primer día de la carrera. En espera de crear el negocio con el que revolucionaremos el sector.

A Marta, que ha escuchado cada día con paciencia y sorprendente interés mis historias sobre "esas cosas raras de las cámaras", y me ha mostrado siempre su confianza en que llegaré a ser un gran teleñeco.

Índice de contenidos

0	INTRODUCCIÓN	10
0.1	MOTIVACIÓN DEL PROYECTO	10
0.1.1	Aplicaciones del proceso de calibración	10
0.1.2	Sensibilidad del proceso de calibración	10
0.2	OBJETIVOS DEL PROYECTO	11
0.3	ESTRUCTURA DE LA MEMORIA.....	12
1	REVISIÓN DE CONCEPTOS BÁSICOS	13
1.1	INTRODUCCIÓN A LA GEOMETRÍA PROYECTIVA.....	13
1.2	COORDENADAS HOMOGÉNEAS EN EL PLANO.....	14
1.2.1	Ecuación afín de una recta y coordenadas homogéneas de la recta	14
1.2.2	Coordenadas homogéneas del punto	14
1.2.3	El punto impropio	15
1.2.4	Dualidad punto recta	15
1.3	EL ESPACIO PROYECTIVO P^N	17
1.4	LA RECTA PROYECTIVA P^1	17
1.5	EL PLANO PROYECTIVO P^2	18
1.6	TRANSFORMACIONES PROYECTIVAS 2D: HOMOGRAFÍAS	19
1.7	TRANSFORMACIONES PROYECTIVAS 3D.....	21
2	CALIBRACIÓN DE CÁMARAS FIJAS	24
2.1	ESTUDIO DEL PROCESO DE CALIBRACIÓN DE UNA CÁMARA	24
2.1.1	Modelo pin-hole.....	24
2.1.2	La matriz de proyección	26
2.1.3	Corrección de la distorsión.....	30
2.2	MÉTODOS DE CALIBRACIÓN DE CÁMARA.....	32
2.2.1	Método de Faugeras [Faugeras, 1986].....	33
2.2.2	Método de Zhang [Zhang, 1998][Zhang, 2000]	36
2.2.3	Conclusiones sobre el estudio de los métodos de calibración	42
2.3	IMPLEMENTACIÓN DE DISTINTOS MÉTODOS DE CALIBRACIÓN DE CÁMARAS	43
2.3.1	Implementación del método de Faugeras.....	43
2.3.2	Implementación del método de Zhang.....	44
2.3.2.1	Programas de ejecución del método	44
2.3.2.2	Programas de interfaz de usuario.....	45
2.4	PRUEBAS SOBRE CÁMARAS FIJAS	47
2.4.1	Cámaras fijas	47
2.4.2	Patrón utilizado.....	48

2.4.3	Requisitos de las imágenes	48
2.4.4	Desarrollo de las pruebas	49
2.4.5	Resultados obtenidos de las pruebas	51
3	AUTOCALIBRACIÓN DE CÁMARAS PTZ.....	54
3.1	ESTUDIO DE DISTINTOS MÉTODOS DE AUTOCALIBRACIÓN DE CÁMARA	54
3.1.1	Introducción.....	54
3.1.2	Clasificación de los métodos de autocalibración	54
3.1.3	Método de Hartley [Hartley, 1994a].....	55
3.1.4	Métodos de autocalibración de cámaras con movimiento de rotación y con zoom.	58
3.1.5	Método lineal de Agapito.....	59
3.2	ESTUDIO DEL ESTADO DEL ARTE DEL PROCESO DE AUTOCALIBRACIÓN DE CÁMARAS PTZ.....	62
3.2.1	Autocalibración de un sistema “arm-eye”.....	
3.2.2	Calibración de una red de cámaras PTZ	58
4	DISEÑO Y DESARROLLO DE UNA APLICACIÓN SOBRE TRES CÁMARAS PTZ QUE GRABAN UNA MISMA ESCENA.....	72
4.1	INTRODUCCIÓN.....	72
4.2	CÁMARAS PTZ	72
4.2.1	Especificaciones de la cámara:.....	73
4.2.1.1	Factor de escala de las cámaras PTZ.....	73
4.2.1.2	Punto principal de las cámaras PTZ.....	74
4.2.1.3	Movimiento pan de la cámara.....	74
4.2.1.4	Movimiento tilt de la cámara	77
4.2.2	Posición de las tres cámaras	78
4.3	CÁLCULO DE LOS PARÁMETROS INTRÍNSECOS DE LAS CÁMARAS MEDIANTE EL MÉTODO DE CALIBRACIÓN DE ZHANG.....	79
4.3.1	Calibración con el nivel más bajo de zoom.....	80
4.3.2	Calibración con distintos niveles de zoom	80
4.4	CÁLCULO DE LOS PARÁMETROS EXTRÍNSECOS DE LAS CÁMARAS, [R T], CONOCIDA LA MATRIZ DE PARÁMETROS INTRÍNSECOS	83
4.4.1	Introducción.....	83
4.4.2	Desarrollo del método	84
4.4.3	Implementación del método	86
4.4.4	Pruebas sobre el método y conclusiones	86
4.4.4.1	Secuencia de imágenes grabadas con un barrido pan	87
4.4.4.2	Secuencia de imágenes grabadas con un barrido tilt.....	88
4.4.4.3	Análisis del error de cuantificación	90
4.5	DESARROLLO DE LA APLICACIÓN	92
4.5.1	Imágenes capturadas por las cámaras	93

4.5.2	Conclusiones sobre la aplicación.....	96
5	CONCLUSIONES Y TRABAJOS FUTUROS	54
5.1	CONCLUSIONES SOBRE EL ESTUDIO DE LOS MÉTODOS DE CALIBRACIÓN	54
5.1.1	Conclusiones sobre la autonomía del método de Zhang.....	54
5.1.2	Conclusiones sobre la eficiencia del método de Zhang.....	54
5.1.3	Conclusiones sobre la versatilidad del método de Zhang	55
5.1.4	Conclusiones sobre la precisión del método de Zhang	58
5.2	CONCLUSIONES SOBRE EL ESTUDIO DE LOS MÉTODOS DE AUTOCALIBRACIÓN	54
5.2.1	Conclusiones sobre el estudio de método de Hartley.....	54
5.2.2	Conclusiones sobre el estudio de método de Agapito	54
5.2.3	Conclusiones sobre el estudio de método que calcula los parámetros intrínsecos de una red de cámaras PTZ.....	55
5.3	CONCLUSIONES SOBRE EL ESTUDIO DE LA CALIBRACIÓN DE VARIAS CÁMARAS PTZ QUE GRABAN UNA MISMA ESCENA.....	54
5.3.1	Conclusiones sobre la autonomía del método propuesto.....	54
5.3.2	Conclusiones sobre la eficiencia del método propuesto.....	54
5.3.3	Conclusiones sobre la versatilidad del método propuesto.....	55
5.3.4	Conclusiones sobre la precisión del método propuesto	58

Índice de figuras

Figura 1-1:	Incidencia de líneas en un punto.....	10
Figura 1-2:	Cada punto de la recta proyectiva es una dirección de R^2	11
Figura 1-3:	Proyección $(X, Y, Z) \rightarrow (x, y)$	13
Figura 1-4:	Transformación proyectiva euclídea.....	14
Figura 1-5:	Transformación 3D Euclídea	16
Figura 1-6:	Rotación de los ejes Z, Y y X.....	17
Figura 2-1:	Modelo de cámara pin-hole	19
Figura 2-2:	Relación entre el punto ideal y el punto real	24
Figura 2-3:	Distorsión radial	25
Figura 2-4:	Distorsión tangencial	25
Figura 2-5:	Plantilla plana	31

Figura 2-6:	Puntos que hay que marcar y orden en el que se marquen en el caso de considerar 9 puntos	40
Figura 2-7:	Cámara fija SONY-DFW-X7	41
Figura 2-8:	Relación entre píxeles totales y píxeles activos de la cámara	41
Figura 2-9:	Ejemplo de una imagen del patrón	42
Figura 2-10:	Imágenes del patrón	43
Figura 2-11:	Puntos que se deben marcar cuando se consideran 9 puntos	44
Figura 2-12:	Puntos que se deben marcar cuando se consideran 25 puntos	45
Figura 3-1:	Proyección del mismo punto M en dos planos imagen diferentes	50
Figura 3-2:	Cámara PTZ	56
Figura 3-3:	Correspondencia entre puntos de distintas imágenes cuando hay un movimiento de rotación y de zoom	60
Figura 3-4:	Imágenes capturadas por la cámara cuando se efectúa un movimiento de rotación	61
Figura 3-5:	Figura extraída de [<i>Towards Calibrating a PTZ Camera Network</i> (Sinha y Pollefeys)]	62
Figura 4-1:	Cámara SONY SNC RZ50	64
Figura 4-2:	Plano imagen. Valores del parámetro “AreaZoom”	67
Figura 4-3:	Vista superior de la cámara. Rango de los ángulos horizontales que puede tomar la cámara	68
Figura 4-4:	Vista lateral de la cámara. Rango de los ángulos horizontales que puede tomar la cámara	69
Figura 4-5:	Imágenes utilizadas para la calibración con zoom x1	71
Figura 4-6:	Puntos considerados de la DINA-4	78
Figura 4-7:	Secuencia de imágenes grabadas mediante un barrido pan	79
Figura 4-8:	Valor obtenido del vector de translación T (m)	80
Figura 4-9:	Secuencia de imágenes grabadas mediante un barrido tilt	81
Figura 4-10:	Valor obtenido del vector de translación T (m)	82
Figura 4-11:	Esquema de la posición de las tres cámaras respecto al sistema de referencia	84
Figura 4-12:	Imágenes de la escena grabadas por PTZ 100 (barrido pan)	85
Figura 4-13:	Imágenes de la escena grabadas por PTZ 101 (barrido pan)	86
Figura 4-14:	Imágenes de la escena grabadas por PTZ 102 (barrido tilt)	87

Índice de tablas

Tabla 1-1:	Matrices de rotación de los ejes Z, Y, X	17
Tabla 2-1:	Parámetros de la cámara calculados según el método de Faugeras	30
Tabla 2-2:	Parámetros intrínsecos calculados mediante el método de Zhang	35
Tabla 4-1:	Valores de la distancia focal y el factor de escala para cada nivel de zoom	66
Tabla 4-2:	Valores obtenidos con nivel de zoom x1 (f = 3,5 mm.)	72
Tabla 4-3:	Código asociado a cada zoom	73
Tabla 4-3:	Valores obtenidos con nivel de zoom x2 (f = 7 mm.).....	73
Tabla 4-3:	Valores obtenidos con nivel de zoom x3 (f = 10,5 mm.)	74
Tabla 4-3:	Valores obtenidos con nivel de zoom x4 (f = 14 mm.).....	74

0. Introducción.

0.1. Motivación del proyecto.

0.1.1. Aplicaciones del proceso de calibración.

La calibración de una cámara es el proceso que permite la obtención de los parámetros que definen las condiciones de formación de la imagen dentro del campo de la visión artificial.

Estos parámetros definen la geometría interna y la óptica de la cámara (parámetros intrínsecos), así como su posición y orientación (parámetros extrínsecos respecto a un objeto de referencia o patrón de calibración).

La calibración es, por tanto, un procedimiento que trata de conocer cómo una cámara proyecta un objeto 3D en el plano imagen para así poder extraer información métrica a partir de las imágenes.

El uso de cámaras calibradas permite resolver aplicaciones relacionadas con la obtención de la posición 3D de los objetos en el espacio a partir de sus imágenes o para la reconstrucción tridimensional del entorno captado por las cámaras.

Esto permite realizar tareas como la obtención de mapas del entorno de la cámara, el seguimiento de un objeto específico para aplicaciones de seguridad o la obtención de la posición de la cámara respecto a objetos que la rodeen. También puede facilitar la navegación por su entorno de un robot móvil, permitiendo evitar obstáculos, dirigirse a objetos determinados o facilitar la definición de la trayectoria más adecuada para alcanzar su destino.

0.1.2. Sensibilidad del proceso de calibración.

Existen múltiples factores que influyen sobre los resultados en el proceso de calibración de una cámara. Esto provoca que los parámetros que se obtienen tengan una alta sensibilidad a las condiciones en las que se obtienen las imágenes que se utilizan para calibrar las cámaras.

Así mismo el proceso de calibración resulta muy complejo debido al elevado número de parámetros que es necesario obtener.

Los métodos de calibración acotan en parte esta complejidad empleando modelos de cámara que son, en realidad, modelos ideales o simplificados de sus equivalentes físicos. Con estas simplificaciones se consiguen resultados bastante aceptables (aunque no exactos), pero no se parametrizan muchos factores.

El modelo de cámara simplificado que utilizan la mayoría de los métodos de calibración es el modelo de agujero o *pinhole*. Esta simplificación no modela aspectos ópticos como

la distancia de enfoque, la profundidad de campo, la apertura o la posible desalineación entre el plano imagen y la lente. Incluso un efecto tan importante como la distorsión de las imágenes se modela, en la mayoría de los casos, de forma muy simplificada, y no en todos los casos se parametriza.

Otros factores importantes que afectan a la sensibilidad de los resultados son las posibles fuentes de ruido que influyen en el proceso de formación de la imagen como los errores de cuantificación que introduce la cámara, o la precisión en la ubicación real de los elementos (puntos del patrón) utilizados para realizar la calibración.

0.2. Objetivos del proyecto.

El objetivo principal de este proyecto es desarrollar un método que permita conocer la posición de las cámaras que graban una escena con respecto a un sistema de referencia común. A partir del conocimiento de esta posición se podrá realizar la operación inversa. Es decir, la obtención de la posición 3D de los objetos en el espacio a partir de sus imágenes.

Este método se aplicará sobre tres cámaras PTZ (*Pan-Tilt-Zoom*) situadas en el vestíbulo de entrada del edificio A de la Escuela Politécnica Superior, y visibles las tres entre sí.

Debido a la especial complejidad formal de este proyecto, a lo largo del mismo existen una serie de objetivos parciales que se centran en el estudio de diferentes procesos de calibración y autocalibración, y en el análisis de las posibilidades reales en el caso concreto de las cámaras PTZ.

Desde un principio se ha planteado el proyecto como la apertura de una línea de investigación propia y se ha tratado por tanto de conseguir que esta línea sea auto contenida. En concreto se van a contemplar los siguientes objetivos parciales:

- El desarrollo e implementación en software propio del método de calibración de Zhang. Con esto se busca independizar nuestro estudio de otros estudios existentes que tratan con sus circunstancias especiales.
- El análisis de la fiabilidad y sensibilidad del método de calibración de Zhang implementado. El desarrollo del método que permita conocer la posición de las cámaras que graban una escena con respecto a un sistema de referencia común se va a realizar suponiendo conocidos los parámetros intrínsecos de las cámaras. Para ello es necesario calibrarlas una vez y esto se va a realizar mediante el método de Zhang. Por tanto es muy importante conocer qué grado de fiabilidad tienen los resultados obtenidos.
- El análisis de las posibilidades reales de autocalibración en el contexto de las cámaras PTZ.

0.3 Estructura de la memoria.

La estructura de este PFC se organiza como sigue:

En el capítulo 1 se realiza una revisión de los conceptos básicos de la geometría proyectiva relacionados con el proceso de captación de imágenes. En este capítulo se introduce la notación que se va a seguir en el resto de la memoria.

En el capítulo 2 se presenta el modelo de cámara *pinhole* del que parten los métodos que se van a estudiar para realizar la calibración. A continuación se describen los algoritmos de los métodos de Faugeras y Zhang.

El método de Faugeras se plantea como un método lineal y muy simple cuyo resultado es utilizado como punto de partida para la realización de otros métodos iterativos más complejos y completos como el de Zhang.

Por su parte, el método de Zhang se presenta como un método útil de calibración de los parámetros intrínsecos de una cámara debido a su flexibilidad.

Este método será utilizado para calibrar los parámetros intrínsecos de las cámaras PTZ, permitiendo utilizar los parámetros obtenidos en el desarrollo de la aplicación que trata de conocer la posición de las cámaras que graban una escena con respecto a un sistema de referencia común. Por ello es necesario implementar y analizar la fiabilidad del método de Zhang. El capítulo continúa con una descripción de la implementación en software propio del método de Zhang y con la presentación de los resultados obtenidos en unas pruebas realizadas sobre cámaras fijas. Con estos resultados se analizará la sensibilidad e inestabilidad del proceso de calibración.

En el capítulo 3 se realiza un estudio de distintos métodos de autocalibración de cámara y del estado del arte del proceso de autocalibración de cámaras PTZ. En este capítulo se analizarán los problemas que presentan las cámaras PTZ a la hora de autocalibrar y como se van a limitar las posibilidades reales de autocalibración.

En el capítulo 4 se realiza la aplicación que permite conocer la posición de las cámaras que graban una escena con respecto a un sistema de referencia común.

Para el desarrollo de esta aplicación es necesario conocer los parámetros intrínsecos de las cámaras. La primera parte del capítulo describe el contexto de las cámaras PTZ y los resultados obtenidos en la calibración de los parámetros intrínsecos mediante el método de Zhang.

La segunda parte del capítulo describe el método desarrollado para averiguar la posición de las cámaras y los resultados de las pruebas realizadas del método.

En el capítulo 5 se presentan las conclusiones del proyecto y las posibles líneas de mejora del mismo.

1. Revisión de conceptos básicos.

En este capítulo se analizan los conceptos de geometría proyectiva que están asociados al proceso de captación de una imagen. El proceso de proyección central (modelo *pinhole* que se introducirá en el siguiente capítulo) con el que se modela el proceso de captación de una imagen es básicamente proyectivo.

Se partirá de una introducción general sobre la geometría proyectiva que permita definir los conceptos que se van a manejar durante el trabajo.

1.1. Introducción a la geometría proyectiva.

La geometría proyectiva suministra un modelo lineal del proceso de captación de imágenes (si no hay distorsiones)⁽¹⁾ pues estudia la relación entre figuras geométricas y su proyección. El ejemplo común usado consiste en figuras en 3D con proyecciones en un plano 2D.

Desde el punto de vista sintético, la geometría proyectiva es una geometría que parte de los siguientes principios:

- Dos puntos definen una recta.
- Todo par de rectas se cortan en un punto (cuando dos rectas son paralelas decimos que se cortan en un punto del infinito).

La geometría proyectiva puede entenderse, informalmente, como la imagen que se obtiene en nuestro ojo del espacio que vemos cuando nos colocamos en un punto, mirando desde ese punto. Esto es, cualquier línea que incide en nuestro "ojo" nos parece ser sólo un punto, ya que el ojo no puede "ver" los puntos que hay detrás (ver Fig. 1.1.).

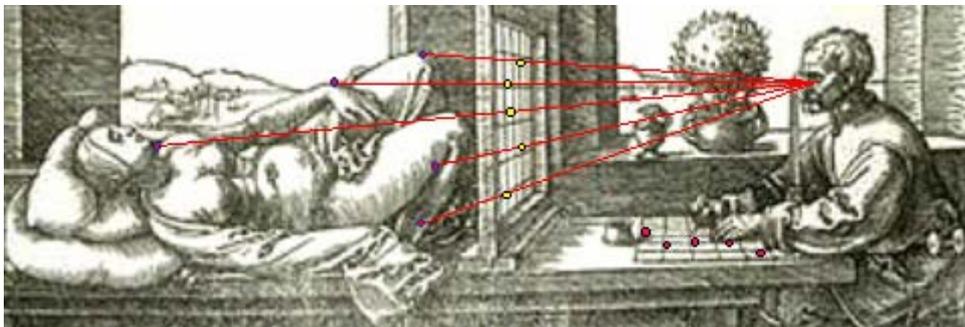


Fig. 1.1. Incidencia de líneas en un punto.

(1). En realidad las lentes provocan distorsiones, lo que equivale a alinealidades, que la geometría proyectiva no puede modelar.

1.2. Coordenadas homogéneas en el plano.

1.2.1. Ecuación afín de una recta y coordenadas homogéneas de la recta.

En el sistema familiar de la geometría analítica euclidiana, un punto \mathbf{m} se representa como un par de coordenadas (x, y) en R^2 . Comúnmente, R^2 se identifica con un plano. Se puede considerar entonces R^2 como un espacio vectorial en el que (x, y) es un vector. Se asocia así, un punto a un vector.

Una línea recta en el plano R^2 está representada por su ecuación afín:

$$ax + by + c = 0$$

De esta manera, una línea recta puede ser representada por un vector: $\mathbf{v} = [a, b, c]^T$. El vector \mathbf{v} representa las coordenadas homogéneas de la recta.

La correspondencia entre líneas rectas y vectores no es uno a uno, ya que $(a, b, c)^T$ y $[ka, kb, kc]^T$ representan exactamente la misma línea recta para $k \neq 0$, sin embargo son vectores distintos. Estos vectores son considerados como equivalentes y se definen como *vectores homogéneos*.

En geometría proyectiva se introducen coordenadas que permiten representaciones múltiples, tanto de puntos como de rectas, con las cuales es posible estudiar nociones como “punto impropio” y “recta proyectiva” de forma analítica.

1.2.2. Coordenadas homogéneas del punto.

Un punto de coordenadas euclidianas $\mathbf{m} = [x, y]^T$ en un plano R^2 tiene una representación homogénea en R^3 dada por $\mathbf{M} = [X \ Y \ Z]^T$. Siendo $[X \ Y \ Z]$ las coordenadas homogéneas del punto euclidiano $[x, y]$.

Las coordenadas homogéneas $[X \ Y \ Z]$ son números reales y la coordenada homogénea Z es no nula. La relación entre las coordenadas homogéneas y euclidianas de un punto es:

$$x = X/Z \quad \text{e} \quad y = Y/Z$$

Una forma sencilla para pasar de coordenadas cartesianas, $[x, y]$, a coordenadas homogéneas consiste en añadir $z = 1$ como coordenada Z : $[x, y, 1]$.

La ecuación de la recta afín en coordenadas homogéneas es:

$$ax + by + c = 0 \Leftrightarrow a \frac{X}{Z} + b \frac{Y}{Z} + c = 0 \Leftrightarrow aX + bY + cZ = 0$$

En este trabajo se van a representar las coordenadas homogéneas con letras mayúsculas, mientras que las euclidianas se representarán con letras minúsculas.

A cada punto representado en coordenadas homogéneas, $\mathbf{M} = [X \ Y \ Z]^T$, le corresponde un único punto de coordenadas euclídeas, $\mathbf{m} = [x, y]^T$; pero a este punto \mathbf{m} le corresponden infinitos puntos homogéneos, ya que tanto las coordenadas (X, Y, Z) como las coordenadas $[kX, kY, kZ]$ representan dicho punto, siempre que k sea un número real diferente de cero.

1.2.3. El punto impropio.

En el apartado anterior no se ha considerado el punto expresado en coordenadas homogéneas $[X, Y, 0]$. A continuación se va a estudiar qué representa dicho punto.

Si se consideran las ecuaciones afines de dos rectas del plano:

$$aX + bY + cZ = 0, \quad a'X + b'Y + c'Z = 0$$

Se puede obtener el punto de corte de estas dos rectas (expresado en coordenadas homogéneas) mediante el producto vectorial de los vectores constituidos por las coordenadas homogéneas de la rectas:

$$[X, Y, Z] \sim [a, b, c] \times [a', b', c']$$

En la geometría euclidiana dos rectas paralelas no se cortan en ningún punto. Sin embargo si en coordenadas homogéneas tratamos hallar el punto de corte de dos rectas paralelas:

$$aX + bY + c = 0, \quad a'X + b'Y + c' = 0$$

$$[X, Y, Z] \sim [a, b, c] \times [a', b', c'] \sim [b, -a, 0]$$

Se obtiene el vector director de las rectas paralelas. Por lo que se puede entender el punto $[X, Y, 0]$ como el “punto” en que se cortan todas las rectas con vector director $\mathbf{v} = (x, Y)$. Este punto recibe el nombre de punto impropio o punto del infinito.

Los puntos del infinito forman la recta del infinito, $Z = 0$.

1.2.4. Dualidad punto recta.

Lo que se consigue mediante las coordenadas homogéneas es que los puntos y las rectas sean algebraicamente lo mismo. Así, la representación en coordenadas homogéneas de una recta del plano está formada por un vector de tres elementos, y la representación en coordenadas homogéneas de un punto del plano también está formada por un vector de tres elementos.

1.3. El espacio proyectivo P^n .

El espacio proyectivo es el espacio de las coordenadas homogéneas que se han presentado en el apartado anterior.

En este apartado se va a estudiar cómo la geometría proyectiva introduce las coordenadas que permiten representaciones múltiples, tanto de puntos como de rectas.

El espacio proyectivo de dimensión n , P^n está formado por todos los elementos del espacio euclídeo R^{n+1} excepto el elemento nulo, tal que dados $x, y \in P^n$, $x = [x_1, \dots, x_{n+1}]^T$ e $y = [y_1, \dots, y_{n+1}]^T$ con algún $x_i \neq 0$ e $y_i \neq 0$, entonces $x = y$ si existe un $\lambda \neq 0$ tal que $x = \lambda \cdot y$.

$$P^n = R^{n+1} - \{0\}$$

Para concretar la definición anterior, a continuación se van a estudiar como ejemplos los espacios proyectivos P^1 y P^2 .

1.4. La recta proyectiva P^1 .

La recta proyectiva es el espacio proyectivo de dimensión uno:

$$P^1 = R^2 - \{0\} \mid \text{si } x = [x_1, x_2]^T \text{ e } y = [y_1, y_2]^T \text{ con algún } x_i \neq 0 \text{ e } y_i \neq 0, \quad x, y \in P^1, \\ \text{entonces } x = y \text{ si existe un } \lambda \neq 0 \text{ tal que } x = \lambda \cdot y.$$

Consideremos un punto perteneciente a la recta proyectiva: $x \in P^1$ cuyas coordenadas homogéneas son $x = [x_1, x_2]^T$, como el punto x representado en la figura 1.2.. Si $x_2 \neq 0$, se pueden expresar sus coordenadas euclídeas como $x = x_2 \cdot [x_1/x_2, 1]^T$.

Una recta afín es aquella que pasa por el origen de coordenadas. Si dos puntos del espacio R^2 pertenecen a la misma recta afín, entonces pertenecen al mismo elemento del espacio proyectivo P^1 .

Es decir, los elementos de P^1 representan el conjunto de todas las rectas afines del espacio euclídeo R^2 (como las representadas en la figura 1.2). Por tanto, cada elemento de la recta proyectiva P^1 representa una dirección (que pasa por el origen) de R^2 .

Faltaría por considerar el punto $[x_1, 0]^T$ que no pertenece a la recta afín. Este punto es el definido como “punto impropio” o punto en el infinito.

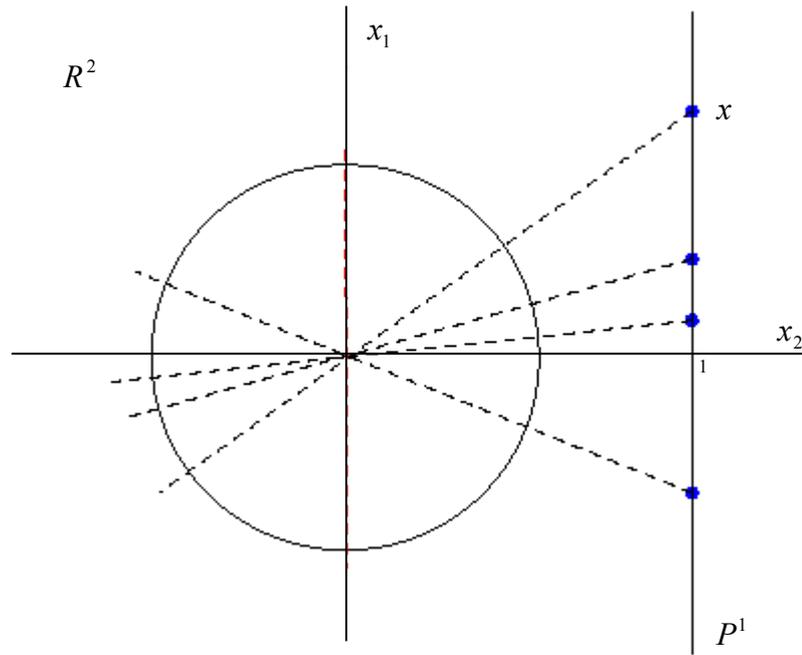


Fig. 1.2. El espacio proyectivo, P^1 .

1.5. El plano proyectivo P^2 .

El plano proyectivo P^2 en el que están representados todos los puntos no homogéneos $x = X/Z$ e $y = Y/Z$, puede ser interpretado utilizando la representación de la Fig. 1.3:

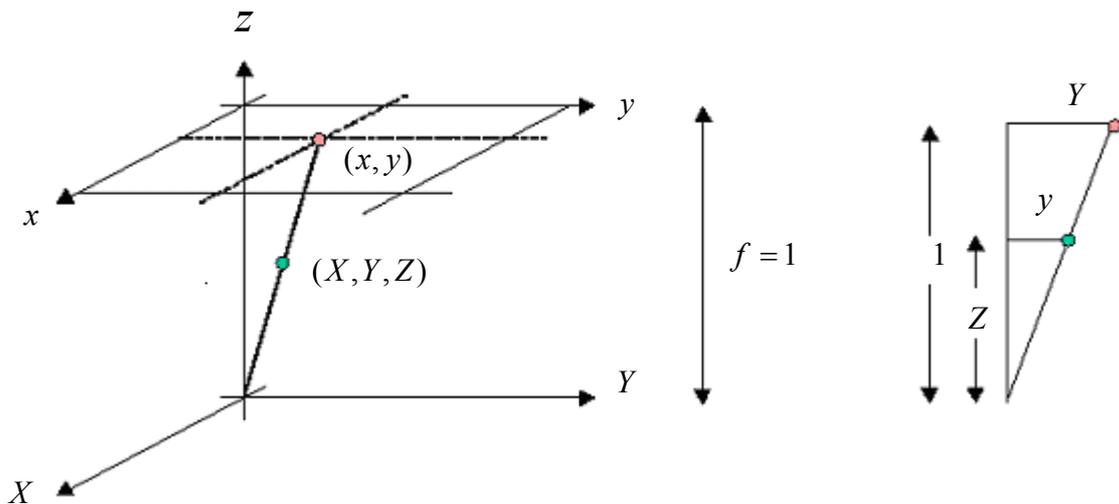


Fig. 1.3. Proyección $(X, Y, Z) \rightarrow (x, y)$

En este esquema la proyección de (X, Y, Z) en un plano (x, y) paralelo al plano $(X, Y, 0)$ ubicado en $Z = 1$, está dada por el punto (x, y) el cual puede ser calculado aplicando el Teorema de Tales como $\frac{x}{X} = \frac{y}{Y} = \frac{1}{Z}$. Se obtiene entonces $x = X/Z$ e $y = Y/Z$.

Se puede observar que cualquier punto 3D sobre la línea de proyección $O(XYZ)$ produce el mismo punto proyectado en el plano (x, y) . Esto es lo mismo que decir que la proyección de $k \cdot (X, Y, Z)$ es igual a la proyección de (X, Y, Z) , para $k \neq 0$, y que esta proyección está dada por $x = X/Z$ e $y = Y/Z$. En otras palabras, la clase de vectores homogéneos $k \cdot (X, Y, Z)$ representan el mismo punto en P^2 .

1.6. Transformaciones proyectivas 2D: homografías.

La geometría proyectiva 2D es el estudio de las propiedades del plano proyectivo P^2 que son invariantes bajo un grupo de transformaciones conocidas como *proyectividades* u *homografías*.

Una homografía es una transformación biyectiva del espacio proyectivo que viene dada por $h : P^2 \rightarrow P^2$ de manera tal que una línea recta es transformada como una línea recta. La proyectividad está definida como:

$$h(\mathbf{m}) = \mathbf{m}' = \mathbf{H} \cdot \mathbf{m}, \quad \mathbf{m}, \mathbf{m}' \in P^2$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \mapsto \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

donde \mathbf{H} es una matriz de 3×3 no singular. Se dice entonces que \mathbf{m}' es la transformación lineal \mathbf{H} de \mathbf{m} . Esta transformación es biunívoca entre dos planos 2D, cuyos puntos son representados homogéneamente por \mathbf{m} y \mathbf{m}' . Es decir, un punto en un plano 2D tiene una única correspondencia en un punto de otro plano 2D, y viceversa.

Es importante destacar que, debido a la naturaleza de las coordenadas homogéneas, dos matrices proporcionales definen la misma homografía ya que los puntos $[X, Y, Z]$ y $[kX, kY, kZ]$ son el mismo punto proyectivo.

Como ejemplo de transformación proyectiva 2D, se va a estudiar la transformación isométrica, que es una transformación euclídea. Esta transformación define una rotación seguida de una translación. En el capítulo 3 se verá que el modelo pinhole incluye esta transformación. La razón de estudiarla en el espacio 2D es que facilitará su entendimiento en el espacio 3D.

En la transformación isométrica se conserva la distancia Euclídea, es decir la distancia entre dos puntos es igual a la distancia entre los puntos transformados. La transformación proyectiva isométrica es ilustrada en la Fig. 1.4.:

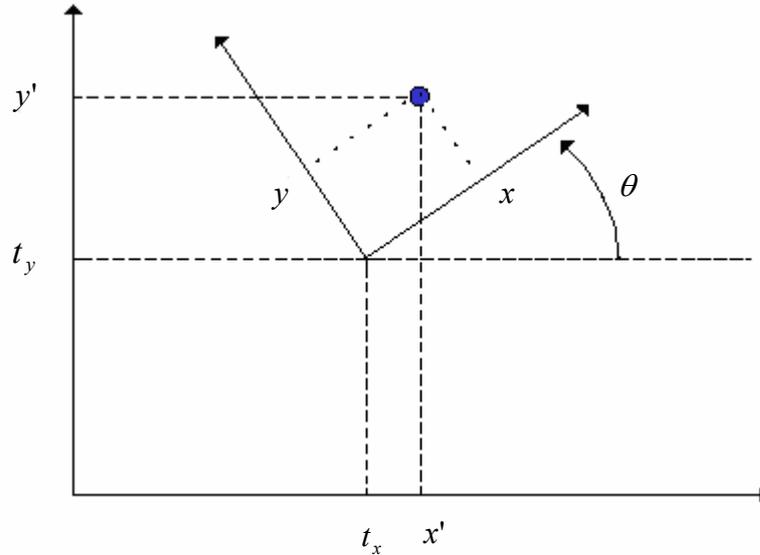


Fig. 1.4. Transformación proyectiva euclídea.

En coordenadas euclídeas, esta figura corresponde a la transformación de coordenadas $(x, y) \leftrightarrow (x', y')$, donde θ es el ángulo de la rotación de los ejes y (t_x, t_y) es el desplazamiento del origen. Esta transformación puede ser escrita en coordenadas cartesianas como:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\text{sen}(\theta) \\ \text{sen}(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} = \mathbf{R} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \mathbf{t} \quad (1.1.)$$

o bien en coordenadas homogéneas, como la proyectividad:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\text{sen}(\theta) & t_x \\ \text{sen}(\theta) & \cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1.2.)$$

con $\mathbf{0}^T = [0 \ 0]$. La transformación inversa $(x', y') \leftrightarrow (x, y)$ se obtiene de:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = [\mathbf{R}']^{-1} \cdot \left(\begin{bmatrix} x \\ y \end{bmatrix} - \mathbf{t}' \right) \quad (1.3.)$$

Como la matriz \mathbf{R} es ortonormal (por ser una matriz de rotación), es decir que $\mathbf{R} \cdot [\mathbf{R}]^T = \mathbf{I}_{2 \times 2}$ se sabe entonces que la inversa de \mathbf{R} es su transpuesta. Definiendo $\mathbf{R}' = [\mathbf{R}]^T = [\mathbf{R}]^{-1}$ y $\mathbf{t}' = \mathbf{R}' \cdot \mathbf{t}$ se obtiene:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}' & \mathbf{t}' \\ \mathbf{0}^T & 1 \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \quad (1.4.)$$

1.7. Transformaciones proyectivas 3D.

Un punto \mathbf{M} que se encuentra en el espacio 3D se representa en coordenadas homogéneas como un vector de cuatro elementos. Si el punto 3D tiene coordenadas (no homogéneas) $[x, y, z]$ se expresará entonces como $\mathbf{M} = [X \ Y \ Z \ T]$ donde $x = X/T$, $y = Y/T$ y $z = Z/T$. Una forma sencilla de pasar de coordenadas no homogéneas a homogéneas es agregando un uno al final del vector, es decir $\mathbf{M} = [X \ Y \ Z \ 1]$.

La única transformación proyectiva 3D que se va a estudiar es la transformación 3D Euclídea, que representa los cambios de coordenadas que pueden sufrir los objetos rígidos al pasar de un sistema de coordenadas a otro. En el capítulo 3 se verá que el modelo de cámara *pinhole* incluye esta transformación.

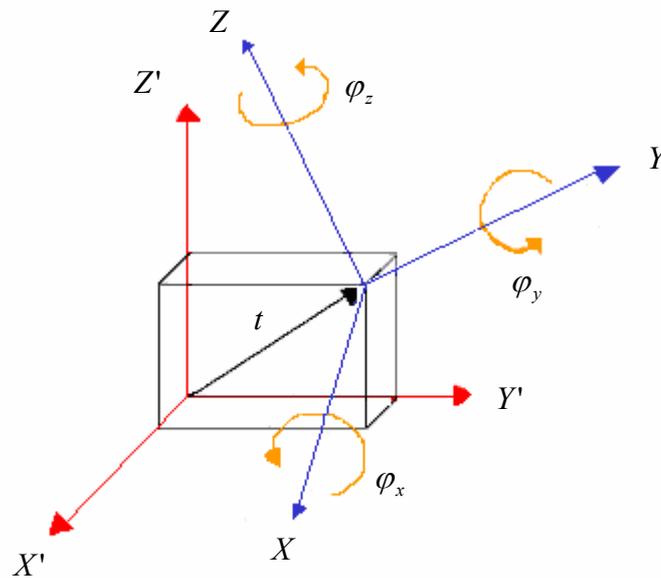


Fig. 1.5. Transformación 3D Euclídea.

Dado un sistema de coordenadas 3D $(x \ y \ z)$ que ha sufrido una rotación y una traslación como se aprecia en la Figura 1.5., el espacio 3D en el nuevo sistema de coordenadas $(x' \ y' \ z')$ queda definido por una transformación 3D Euclídea definida por:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathbf{R} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \mathbf{t} \quad (1.6.)$$

donde \mathbf{R} es una matriz 3×3 ortonormal, por ser una matriz de rotación, y \mathbf{t} es un vector 3×1 que definen la rotación y traslación del sistema de coordenadas respectivamente.

Esta transformación se define en coordenadas homogéneas de la siguiente forma:

$$\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1.7.)$$

A continuación se definirá la matriz ortonormal \mathbf{R} presente en la transformación Euclídea $\mathbb{R}^3 \rightarrow \mathbb{R}^3$. Una rotación de los ejes de coordenadas puede ser descompuesto en rotaciones de cada uno de los ejes tal como se muestra en la figura 1.6.:

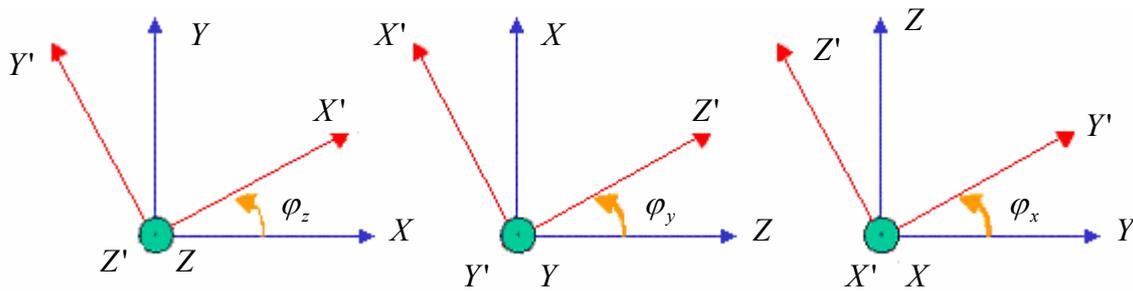


Fig. 1.6. Rotación de los ejes Z, Y y X

Las transformaciones de cada una de estas rotaciones están dadas por \mathbf{R}_Z , \mathbf{R}_Y y \mathbf{R}_X en la tabla 1.1.:

Rotación	Matriz de rotación
Eje Z	$\mathbf{R}_z = \begin{bmatrix} \cos(\varphi_z) & \text{sen}(\varphi_z) & 0 \\ -\text{sen}(\varphi_z) & \cos(\varphi_z) & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Eje Y	$\mathbf{R}_y = \begin{bmatrix} \cos(\varphi_y) & 0 & -\text{sen}(\varphi_y) \\ 0 & 1 & 0 \\ \text{sen}(\varphi_y) & 0 & \cos(\varphi_y) \end{bmatrix}$
Eje X	$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\varphi_x) & \text{sen}(\varphi_x) \\ 0 & -\text{sen}(\varphi_x) & \cos(\varphi_x) \end{bmatrix}$

Tabla 1.1. Matriz de rotación de los ejes Z, Y y X.

A manera de ejemplo si el único movimiento existente es la rotación del eje X, la ecuación que transforma las coordenadas sería:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\varphi_x) & \text{sen}(\varphi_x) \\ 0 & -\text{sen}(\varphi_x) & \cos(\varphi_x) \end{bmatrix} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (1.8.)$$

La rotación total se puede definir entonces como primero una rotación del eje Z, luego del eje Y, y luego del eje X, eso se puede expresar matemáticamente como una multiplicación de las tres matrices de rotación en el siguiente orden:

$$\mathbf{R} = \mathbf{R}_x \cdot \mathbf{R}_y \cdot \mathbf{R}_z = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \quad (1.9.)$$

Análogamente al caso 2D, la utilidad de definir esta transformación en coordenadas homogéneas frente a las coordenadas cartesianas es que nos permite realizar una transformación que consta de dos pasos (multiplicar por la matriz de rotación \mathbf{R} y sumar el vector de translación \mathbf{t}) en un solo paso (multiplicar por la matriz $\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} = [\mathbf{R} \ \mathbf{t}]$).

2. Calibración de cámaras fijas.

2.1. Estudio del proceso de calibración de una cámara.

Se denomina calibración al proceso de obtención de los valores de los parámetros del modelo de una cámara. Estos parámetros relacionan una escena del espacio tridimensional con la imagen bidimensional captada por la cámara.

2.1.1. Modelo pin-hole.

La gran mayoría de los procedimientos de calibración se basan en el modelo de cámara *pin-hole*. Este modelo está basado en la geometría proyectiva: la proyección de un punto de la escena se obtiene de la intersección de una línea que pasa por este punto y el centro de proyección (foco) con el plano imagen.

El modelo consiste en un centro óptico C , en donde convergen todos los rayos de la proyección, y un plano de imagen en el cual la imagen es proyectada. El plano de imagen está ubicado a una distancia focal f del centro óptico y perpendicular al eje óptico Z_c . (ver Fig 2.1).

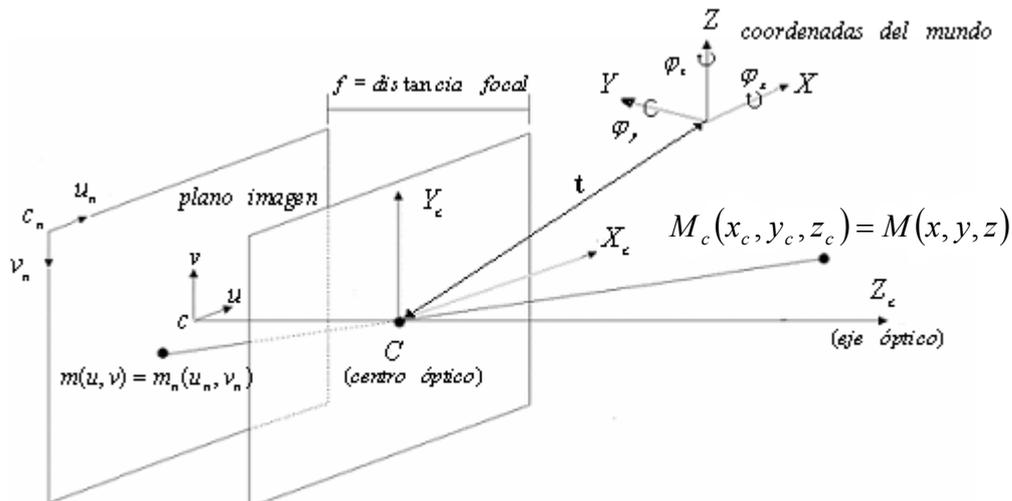


Fig. 2.1. Modelo de cámara *pin-hole*.

Como se ve en la Figura 2.1., el modelo *pinhole* describe la proyección de un punto \mathbf{M} de la escena en un punto \mathbf{m} de la imagen. Para modelar la proyección es necesario realizar varias transformaciones y referirse a varios sistemas de coordenadas distintos. A continuación se describen los sistemas de coordenadas que aparecen en el modelo de proyección.

- *Sistema de coordenadas del mundo:* son las coordenadas que describen la posición del punto 3D \mathbf{M} respecto de la escena. La elección de estas coordenadas es arbitraria. Como se ve en la Figura 2.1., los ejes de este sistema de coordenadas son XYZ.
- *Sistema de coordenadas de la cámara:* son las coordenadas que describen la posición del punto 3D \mathbf{M}_c respecto de la cámara. Como se ve en la Figura 2.1., los ejes de este sistema de coordenadas son $X_c Y_c Z_c$. Y su origen es el centro óptico de la cámara. El punto \mathbf{M} y el punto \mathbf{M}_c son el mismo punto como se indica en la Figura 2.1., nos referimos a él con distinta notación únicamente para indicar respecto a que sistema de coordenadas lo estamos refiriendo.
- *Sistema de coordenadas de la imagen:* son las coordenadas que describen la posición del punto 2D \mathbf{m} respecto del plano imagen. Normalmente este sistema de coordenadas tiene su origen en el centro del plano imagen (c en la Figura 2.1.). Como se ve en esta Figura, los ejes de este sistema de coordenadas son uv .
- *Sistema de coordenadas normalizadas de la imagen:* son las coordenadas que describen la posición del punto 2D m_c respecto del plano imagen, situando el origen de estas coordenadas en la esquina superior izquierda del plano imagen (c en la Figura 2.1.). Como se ve en esta Figura, los ejes de este sistema de coordenadas son $u_n v_n$. El punto \mathbf{m} y el punto m_c también son el mismo punto como se indica en la Figura 2.1., de nuevo nos referimos a él con distinta notación únicamente para indicar respecto a que sistema de coordenadas lo estamos refiriendo.

El modelo de proyección depende de una serie de parámetros de la cámara. Estos parámetros se clasifican en extrínsecos e intrínsecos:

- **Parámetros extrínsecos:** son los que relacionan la posición relativa entre un objeto de la escena que se está captando y la cámara fija.
 - **Traducción:** $\mathbf{t} = [t_x, t_y, t_z]^T$. (ver Fig. 2.1.). es la distancia entre el centro óptico de la cámara y el centro de coordenadas del mundo.
 - **Rotación:** son los ángulos rotados sobre cada uno de los ejes, $[\varphi_x, \varphi_y, \varphi_z]$ (ver Fig. 2.1.).
- **Parámetros intrínsecos:** son los parámetros propios de la geometría interna y de la óptica de la cámara. Estos parámetros son:
 - **Centro del eje óptico ($\mathbf{c} = [u_0, v_0]$):** también llamado punto principal. Define el punto donde el eje óptico (z_c) atraviesa el plano imagen. Las coordenadas de este punto vienen dadas en píxeles.
 - **Factores de escalado (k_u, k_v):** indican la proporción de tamaño de un objeto visto en la realidad respecto a su proyección en el plano imagen. La

proporción puede ser distinta en cada eje. Este parámetro se descompone a su vez en:

- Factores de conversión píxel-milímetros (d_u, d_v): indican el número de píxeles por milímetro que usa la cámara. Esta relación se obtiene dividiendo la dimensión en píxeles de la imagen por el tamaño en mm. del CCD.
- Distancia focal (f): distancia entre el centro óptico y el centro del plano imagen, viene dada en mm.
- Factor de proporción (s): indica la relación de tamaño entre la dimensión horizontal y vertical de un píxel.

La relación entre ellos (cuando no existe distorsión) viene dada por:

$$k_u = s \cdot d_u \cdot f \quad k_v = d_v \cdot f$$

2.1.2. La matriz de proyección.

Si consideramos un punto 3D en coordenadas cartesianas y lo expresamos en coordenadas homogéneas, dicho punto tiene 4 coordenadas: $\mathbf{M} = [X, Y, Z, T]$. De la misma forma, un punto 2D en coordenadas cartesianas tiene 3 coordenadas si lo expresamos en coordenadas homogéneas: $\mathbf{m} = [U, V, S]$.

Supongamos que la cámara capta una escena y la plasma en el plano imagen. La relación entre un punto \mathbf{M} de la escena 3D y su correspondiente punto \mathbf{m} de la imagen, viene determinada por la forma:

$$\mathbf{m} \overset{\circ}{=} \mathbf{P} \cdot \mathbf{M}$$

El símbolo " $\overset{\circ}{=}$ " indica que los elementos de ambos lados son equivalentes excepto por un factor de proporcionalidad, es decir, $\lambda \mathbf{m} = \mathbf{P} \cdot \mathbf{M}$, siendo λ el citado factor, o factor de escala.

La matriz \mathbf{P} se denomina matriz de proyección y depende de los parámetros extrínsecos e intrínsecos.

Para determinar la forma de la matriz de proyección se estudia cómo se relacionan las coordenadas del punto \mathbf{M} , expresadas en base a las coordenadas "del mundo" (ejes XYZ de la figura 2.1.), con las coordenadas del punto \mathbf{m} , expresadas en referencia a las coordenadas normalizadas de la imagen (ejes $u_n v_n$ de la citada figura).

1. Proyección 3D-2D.

Conversión de las coordenadas de la cámara (x_c , y_c , z_c), a las del plano de la imagen (u , v).

Basándose en la geometría de la proyección, podemos relacionar estas coordenadas mediante la distancia focal (f):

$$\frac{f}{z_c} = \frac{u}{x_c} = \frac{v}{y_c}$$

A partir de esta relación se construye la matriz de proyección perspectiva:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{f}{z_c} \cdot \begin{bmatrix} x_c \\ y_c \end{bmatrix}$$

En coordenadas homogéneas:

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

2. Transformación Cámara-Imagen.

Paso de las coordenadas del plano de la imagen (u , v) a las normalizadas de la imagen (u_n , v_n).

Se trata de un cambio de coordenada del sistema (c , u , v) al (c_n , u_n , v_n). Para ello, hay que sumar vectorialmente a m la posición de c respecto a c_n como se indica en la figura 2.2.

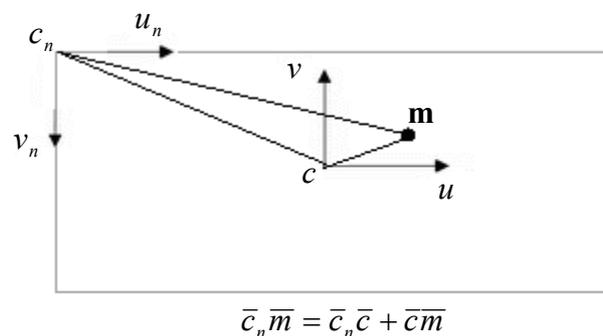


Fig 2.2. Cambio de coordenada del sistema (c , u , v) al (c_n , u_n , v_n).

Las coordenadas además, dejan de expresarse en milímetros y pasan a expresarse en pixels mediante los factores de conversión.

$$u_n = d_u \cdot u + u_0 \qquad v_n = d_v \cdot v + v_0$$

Con estos dos primeros pasos queda constituida la matriz de parámetros intrínsecos (en adelante \mathbf{A}) que relaciona el punto 3D \mathbf{M}_c (referido a las coordenadas de la cámara) con el punto \mathbf{m}_n (referido a las coordenadas normalizadas de la imagen). En coordenadas homogéneas:

$$\begin{bmatrix} U_n \\ V_n \\ S_n \end{bmatrix} = \begin{bmatrix} f \cdot d_u & 0 & u_0 & 0 \\ 0 & f \cdot d_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (2.1.)$$

$$\mathbf{m}_n = \mathbf{A} \cdot \mathbf{M}_c$$

Según se ha visto, el sistema de coordenadas 3D de la cámara no suele coincidir con el sistema de coordenadas de la escena que suele referirse al objeto que se está estudiando (“coordenadas del mundo”). Por ello, es necesario un nuevo cambio de coordenadas que relacione las coordenadas del mundo con las coordenadas de la cámara. Este cambio de coordenadas corresponde a una transformación 3D Euclídea que se ha estudiado en el apartado 1.7 de la memoria.

3. Transformación Escena-Cámara

Paso de las coordenadas del mundo (x, y, z) a las de la cámara (x_c, y_c, z_c) .

Es un cambio de coordenadas definido por una matriz de rotación y translación que se denomina matriz extrínseca.

Primero se realiza una rotación alrededor de los tres ejes:

$$\begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (2.2.)$$

$$\mathbf{M}_r = \mathbf{R} \cdot \mathbf{M}$$

Seguida de una translación:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (2.3)$$

$$\mathbf{M}_c = \mathbf{M}_r + \mathbf{t}$$

Combinando las ecuaciones 2.2. y 2.3. se obtiene la ecuación que realiza el paso de coordenadas del mundo a las coordenadas de la cámara:

$$\mathbf{M}_c = \mathbf{R} \cdot \mathbf{M} + \mathbf{t}$$

Donde \mathbf{R} es la matriz de rotación y \mathbf{t} es el vector de translación.

Según se ha visto, esta expresión matricial se puede escribir en coordenadas homogéneas mediante la matriz de parámetros extrínsecos (en adelante $[\mathbf{R} \ \mathbf{t}]$):

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.4)$$

En conclusión, la expresión general de la matriz de proyección que relaciona un punto \mathbf{M} de la escena 3D y su correspondiente punto \mathbf{m} de la imagen es el resultado de multiplicar la matriz de parámetros intrínsecos de la ecuación 2.1 (\mathbf{A}) y la matriz de parámetros extrínsecos de la ecuación 2.4. $[\mathbf{R} \ \mathbf{t}]$:

$$\mathbf{A} = \begin{bmatrix} f \cdot d_u & 0 & u_0 & 0 \\ 0 & f \cdot d_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad [\mathbf{R} \ \mathbf{t}] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_x \\ r_{31} & r_{32} & r_{33} & t_x \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} f \cdot d_u & 0 & u_0 & 0 \\ 0 & f \cdot d_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_x \\ r_{31} & r_{32} & r_{33} & t_x \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.5)$$

Por lo tanto, la expresión general que relaciona un punto \mathbf{M} de la escena con su correspondiente punto \mathbf{m} de la imagen es:

$$\lambda \cdot \begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = \begin{bmatrix} f \cdot d_u & 0 & u_0 & 0 \\ 0 & f \cdot d_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_x \\ r_{31} & r_{32} & r_{33} & t_x \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.6)$$

$$\lambda \cdot \mathbf{m} = \mathbf{P} \cdot \mathbf{M}$$

El modelo que se ha introducido no tiene en cuenta los efectos de distorsión de la lente, por lo que a continuación se va a explicar el problema de la distorsión y su efecto en el modelo.

2.1.3. Corrección de la distorsión introducida por las lentes.

El modelo *pinhole* no modela correctamente las lentes reales porque el centro óptico de estas lentes no es único.

La distorsión es la distancia existente entre la posición ideal dada por el sistema *pinhole* y la posición real que ocupa cada píxel en la imagen.

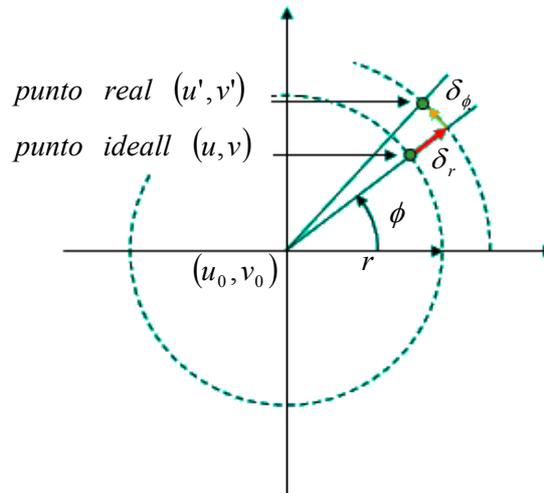


Fig. 2.3. Relación entre el punto ideal y el punto real.

Podemos expresar la relación entre el punto real y el punto ideal con las siguientes ecuaciones:

$$u' = u + \delta_u(u, v) \quad v' = v + \delta_v(u, v)$$

Donde δ_u representa la distorsión de la coordenada u del punto real respecto al punto ideal, y δ_v representa la distorsión de la coordenada v del punto real respecto al punto ideal.

Existen dos tipos principales de distorsión:

- **Distorsión radial** (es más apreciable cuanto más alejado se está del punto principal). Se produce en cámaras de gran angular donde la distancia focal es muy corta. La Fig. 2.4 muestra el efecto de esta distorsión.

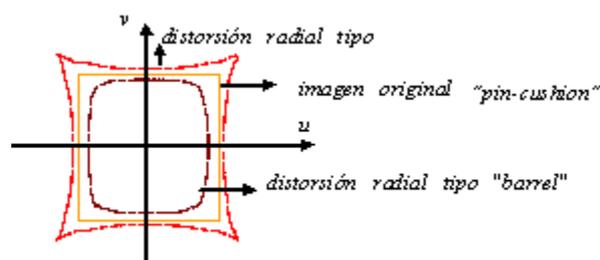


Fig. 2.4. Distorsión radial.

Un modelo utilizado frecuentemente para las ecuaciones de corrección son:

$$r^2 = \left(\frac{u-u_0}{f_x} \right)^2 + \left(\frac{v-v_0}{f_y} \right)^2$$

$$\delta_r = a_1 \cdot r^2 + a_2 \cdot r^4 + a_3 \cdot r^6 + a_4 \cdot r^8 + a_5 \cdot r^{10}$$

Donde δ_r es la distancia radial entre el punto real y el punto ideal como se muestra en la figura 2.3; los parámetros a_i son los términos del polinomio de corrección.

- **Distorsión tangencial:** debido a la no perpendicularidad entre el eje óptico de la lente fija y el plano imagen. La Fig. 2.5 muestra el efecto de esta distorsión

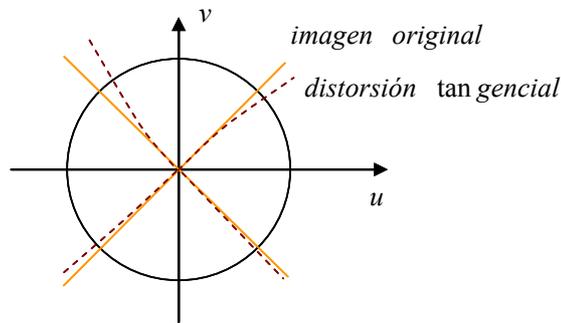


Fig. 2.5. Distorsión tangencial.

Las ecuaciones de corrección son diferentes para cada coordenada:

$$\delta_{\phi_u} = p_1 \cdot (r^2 + 2 \cdot u^2) + 2 \cdot p_2 \cdot u \cdot v$$

$$\delta_{\phi_v} = p_2 \cdot (r^2 + 2 \cdot v^2) + 2 \cdot p_1 \cdot u \cdot v$$

Siendo δ_{ϕ_u} y δ_{ϕ_v} las componentes u y v de la distancia tangencial δ_{ϕ} entre el punto real y el punto ideal como se muestra en la figura; p_1 y p_2 son dos parámetros de segundo orden con los que se modela la distorsión

Para realizar la corrección, se añaden estos términos a las coordenadas ideales del modelo *pinhole*. La distorsión de cada coordenada se expresa de la siguiente forma:

$$\begin{aligned} \delta_u(u, v) &= \delta_r(r) \cdot \cos(\phi) - \delta_{\phi}(\phi) \cdot \text{sen}(\phi) \\ \delta_v(u, v) &= \delta_r(r) \cdot \text{sen}(\phi) + \delta_{\phi}(\phi) \cdot \cos(\phi) \end{aligned} \quad (2.7.)$$

donde r es la distancia entre el punto proyectado en la imagen y centro de la imagen, y ϕ es la distancia angular entre el punto proyectado en la imagen y el eje horizontal u de la imagen.

2.2. Métodos de calibración de cámara.

Como se ha señalado en el apartado anterior, el proceso de calibración de una cámara consiste en estimar sus parámetros extrínsecos e intrínsecos. Los métodos de calibración “clásicos” parten de una escena conocida: se sitúan una serie de puntos 3D cuyas posiciones respecto al sistema de coordenadas del mundo 3D son conocidas. Esto requiere una preparación exhaustiva de la escena.

Frente a los métodos clásicos se sitúan los métodos de autocalibración, que se caracterizan por no necesitar el conocimiento de ningún dato de la escena y que serán desarrollados en el siguiente capítulo.

La clasificación de los métodos de calibración se realiza en función de los siguientes criterios:

Clasificación en cuanto a la computación:

- **Lineal:** aquellos que usan técnicas de resolución de sistemas de ecuaciones lineales. Tienen las ventajas de ser muy simples de implementar y muy rápidos. Ejemplos de estos métodos son [Faugeras, 1986] y [Hall et al., 1982].
- **No lineal:** se basan en la utilización de métodos iterativos. Suelen necesitar una aproximación inicial que se obtiene mediante algún método lineal. Son más lentos pero permiten resolver sistemas de cámara más complejos. Ejemplos de estos métodos son [Tsai, 1987] y [Zhang, 1998].

Clasificación en cuanto a la obtención de los parámetros:

- **Explícita:** se obtienen directamente los valores de cada uno de los parámetros que forman el modelo de la cámara. Ejemplos de estos métodos son [Batista et al., 1998] y [Tsai, 1987].
- **Implícita:** se obtiene primero la matriz de proyección y a partir de ella se calculan los parámetros. Ejemplos de estos métodos son [Faugeras, 1986] y [Zhang, 1998].

En este apartado se van a estudiar dos métodos de calibración. El primer método que se va a estudiar es el propuesto por Faugeras. Este método está basado en una transformación lineal directa y propone una solución homogénea utilizando una restricción en cuanto al vector de rotación del eje z. Se ha escogido este método por ser un buen ejemplo para comenzar en el estudio de la calibración de cámaras.

El segundo que se va a estudiar es el propuesto por Zhang. La razón de elegir este método es que supone un paso muy importante en una de las líneas básicas del estudio de la calibración de cámaras: la simplificación del trabajo previo necesario para preparar la escena. Este es uno de los objetivos del proyecto por lo que se ha considerado fundamental su estudio. Esta simplificación se basa en el hecho de utilizar un patrón de referencia del que se toman varias imágenes para no tener que preparar y medir los puntos de la escena. El método de Zhang se considera un paso intermedio entre la calibración y la autocalibración.

2.2.1. Método de Faugeras [Faugeras, 1986]

El proceso de calibración se realiza en dos pasos. Primero se estima la matriz de proyección \mathbf{P} , y después se estiman los parámetros intrínsecos y extrínsecos de la cámara a partir de la matriz de proyección estimada en la fase anterior.

Este método resuelve la calibración mediante la descomposición en valores singulares de una matriz que forma el sistema de ecuaciones que proyecta un punto del espacio en el plano imagen de la cámara.

Lo que se obtiene es una matriz de proyección a partir de la cual se extraen todos los parámetros, tanto los extrínsecos (traslación y rotación del patrón respecto a la cámara) como los intrínsecos (centro del eje óptico, factores y factores de escalado).

El método de Faugeras no tiene en cuenta la distorsión introducida por las lentes, por lo que no la corrige.

Para realizar el método de Faugeras únicamente es necesario una imagen.

1. Estimación de la matriz de proyección \mathbf{P} .

Como ya se ha indicado, la matriz de proyección \mathbf{P} representa la transformación entre los puntos i 3D de la escena y la posición de los puntos 2D correspondientes en la imagen (ecuación 2.6).

$$\lambda_i \cdot \mathbf{m}_i = \mathbf{P} \cdot \mathbf{M}_i \quad (2.8)$$

Se parte de n correspondencias $\mathbf{m}_i \leftrightarrow \mathbf{M}_i$, ($i = 1, \dots, n$), donde \mathbf{M}_i es un punto de la escena de coordenadas conocidas y \mathbf{m}_i es su correspondiente punto de la imagen, de coordenadas también conocidas..

El objetivo es calcular la matriz \mathbf{P} a partir de n ecuaciones (2.8) asociadas a las n correspondencias.

Cada correspondencia genera dos ecuaciones:

$$u_i = \frac{p_{11} \cdot X_i + p_{12} \cdot Y_i + p_{13} \cdot Z_i + p_{14}}{p_{31} \cdot X_i + p_{32} \cdot Y_i + p_{33} \cdot Z_i + p_{34}} \quad (2.9)$$

$$v_i = \frac{p_{21} \cdot X_i + p_{22} \cdot Y_i + p_{23} \cdot Z_i + p_{24}}{p_{31} \cdot X_i + p_{32} \cdot Y_i + p_{33} \cdot Z_i + p_{34}} \quad (2.10)$$

donde (u_i, v_i) son las coordenadas del punto de la imagen \mathbf{m}_i , (X_i, Y_i, Z_i) son las coordenadas del punto de la escena \mathbf{M}_i , y p_{jk} son los elementos de la matriz \mathbf{P} .

Si en las ecuaciones (2.9) y (2.10) se pasa el denominador multiplicando al otro lado de la ecuación, se obtienen dos ecuaciones lineales con los elementos de la matriz \mathbf{P} como incógnitas.

$$u_i \cdot (p_{31} \cdot X_i + p_{32} \cdot Y_i + p_{33} \cdot Z_i + p_{34}) = p_{11} \cdot X_i + p_{12} \cdot Y_i + p_{13} \cdot Z_i + p_{14} \quad (2.11)$$

$$v_i \cdot (p_{31} \cdot X_i + p_{32} \cdot Y_i + p_{33} \cdot Z_i + p_{34}) = p_{21} \cdot X_i + p_{22} \cdot Y_i + p_{23} \cdot Z_i + p_{24} \quad (2.12)$$

Estas ecuaciones se reordenan como:

$$\begin{bmatrix} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -u_i \cdot X_i & -u_i \cdot Y_i & -u_i \cdot Z_i & -u_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -v_i \cdot X_i & -v_i \cdot Y_i & -v_i \cdot Z_i & -v_i \end{bmatrix} \cdot \mathbf{p} = 0 \quad (2.13)$$

donde \mathbf{p} es un vector que contiene los 12 elementos de la matriz \mathbf{P} :

$$\mathbf{p} = [p_{11} \quad p_{12} \quad p_{13} \quad p_{14} \quad p_{21} \quad p_{22} \quad p_{23} \quad p_{24} \quad p_{31} \quad p_{32} \quad p_{33} \quad p_{34}]^T$$

Se consideran $2n$ ecuaciones lineales semejantes a las de la ecuación (2.13) a partir de n correspondencias entre puntos ($n \geq 6$) para formar un sistema de $2n$ ecuaciones lineales. Este sistema se denota como:

$$\mathbf{L} \cdot \mathbf{p} = \mathbf{0} \quad (2.14)$$

donde \mathbf{L} es una matriz de dimensión $2n \times 12$.

En general este sistema no tiene una solución exacta, pero se puede obtener una solución lineal mediante el autovector correspondiente al autovalor más pequeño que minimiza $\mathbf{L}^T \cdot \mathbf{L}$. O, de forma equivalente, con el vector correspondiente al valor singular más pequeño de la SVD (*Single Value Descomposition* o descomposición en valores singulares) de \mathbf{L} . La herramienta SVD se define en el anexo A de esta memoria.

Al calcular explícitamente los coeficientes de \mathbf{P} en función de los parámetros que componen la matriz de proyección, se encuentra la particularidad de que $[r_{31}, r_{32}, r_{33}] = [p_{31}, p_{32}, p_{33}]$, donde $[r_{31}, r_{32}, r_{33}]$ corresponde a la tercera fila de la matriz de rotación. Como esta matriz es ortonormal, la norma de sus vectores fila o columna es la unidad.

$$r_{31}^2 + r_{32}^2 + r_{33}^2 = 1$$

Por lo tanto imponer esta restricción en \mathbf{P} supone normalizarla, es decir, se dividen todos los elementos de la matriz por:

$$p_{31}^2 + p_{32}^2 + p_{33}^2$$

2. Estimación de los parámetros extrínsecos e intrínsecos a partir de la matriz de proyección.

Parámetro	Tipo	Valor
Vector de rotación r_3	Extrínseco	$r_3 = p_3$
Coordenada u_0 del centro óptico	Intrínseco	$u_0 = p_1 \cdot p_3$
Coordenada v_0 del centro óptico	Intrínseco	$v_0 = p_2 \cdot p_3$
Factor de escala $\alpha = f \cdot k_u$	Intrínseco	$\alpha = f \cdot k_u = -\ p_1^{p_2}\ $
Factor de escala $\beta = f \cdot k_v$	Intrínseco	$\beta = f \cdot k_v = -\ p_2^{p_3}\ $
Vector de rotación r_1	Extrínseco	$r_1 = \frac{1}{\alpha \cdot (p_1 - u_0 \cdot p_3)}$
Vector de rotación r_2	Extrínseco	$r_2 = \frac{1}{\beta \cdot (p_2 - v_0 \cdot p_3)}$
Vector de translación en el eje x, t_x	Extrínseco	$t_x = \frac{1}{\alpha \cdot (p_{14} - u_0 \cdot p_{34})}$
Vector de translación en el eje y, t_y	Extrínseco	$t_y = \frac{1}{\beta \cdot (p_{24} - u_0 \cdot p_{34})}$
Vector de translación en el eje z, t_z	Extrínseco	$t_z = p_{34}$

Tabla 2.1. Parámetros de la cámara calculados según el método de Faugeras.

La principal ventaja de este método es que es muy rápido de computar. Como principal desventaja presenta que no incluye la distorsión (radial y tangencial) por lo que los resultados son poco exactos. Además exige conocer el punto \mathbf{M}_i de la escena 3D.

En muchos métodos iterativos se pueden utilizar los resultados de este método como aproximación inicial para obtener la matriz de transformación (\mathbf{P}), como ocurre en [Zhang, 1998].

2.2.2. Método de Zhang [Zhang, 1998][Zhang, 2000]

Zhang propone una técnica de calibración basada en la observación de una plantilla plana como la de la figura 2.5 desde varias posiciones. A diferencia del método de Faugeras, el método de Zhang utiliza varias imágenes distintas.

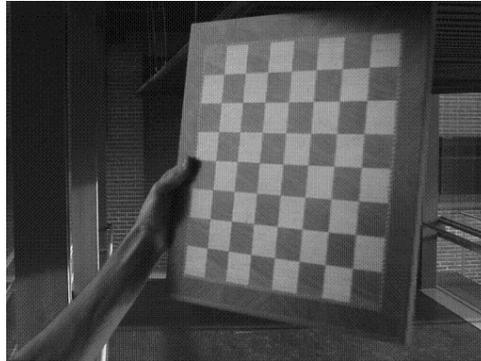


Fig. 2.5. Plantilla plana.

La ventaja de este método de calibración con respecto al resto de los métodos es que permite obtener los parámetros de la cámara fácilmente a partir de una plantilla plana sin necesidad de conocer las posición de los puntos \mathbf{M}_i de la escena que se van a utilizar para realizar la calibración. La razón es que se eligen las coordenadas del mundo de tal forma que dos de sus ejes coinciden con los lados de la plantilla, y el tercero (Z) es perpendicular a esta.

Esto hace que sea una técnica muy flexible ya que no requiere una preparación exhaustiva de la escena.

Para que el método funcione correctamente se necesitan al menos tres imágenes con la plantilla tomada en distintas orientaciones. Este número de imágenes puede ser inferior si se fijan los valores de algunos parámetros intrínsecos. Por ejemplo, si no se calcula la ortogonalidad del plano imagen, sólo son necesarias dos imágenes.

El modelo de cámara que utiliza es el *pinhole*. La proyección de un punto \mathbf{M} de la escena en un punto \mathbf{m} de la imagen viene dada por la ecuación (2.6.):

$$\lambda \cdot \mathbf{m} = \mathbf{P} \cdot \mathbf{M}$$

donde λ es un factor de escala dado que las coordenadas de los puntos son homogéneas.

Como ya se ha indicado en la ecuación (2.5), la matriz de proyección \mathbf{P} es el producto de la matriz de parámetros intrínsecos \mathbf{A} y la matriz de parámetros extrínsecos $[\mathbf{R} \ \mathbf{t}]$. Por lo que la ecuación (2.6) se puede escribir de la siguiente forma:

$$\lambda \cdot \mathbf{m} = \mathbf{A} \cdot [\mathbf{R} \ \mathbf{t}] \cdot \mathbf{M}$$

El método de Zhang no supone que los ejes del plano de la imagen sean ortogonales. Por ello, introduce un nuevo parámetro intrínseco que mide esta pérdida de ortogonalidad. En el caso de que los ejes fuesen perfectamente ortogonales, el valor de este parámetro sería la unidad (como supone el modelo tradicional *pinhole*).

Al añadir el parámetro que mide la pérdida de ortogonalidad (γ), la matriz de parámetros intrínsecos en el modelo de Zhang tiene la forma:

$$\mathbf{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

siendo $\alpha = (s \cdot d_u \cdot f)$ y $\beta = (d_v \cdot f)$ los factores de escala en el eje u y v ; (u_0, v_0) las coordenadas del punto principal de la imagen, y γ el parámetro que representa la pérdida de ortogonalidad de los ejes en la imagen. Se ha cambiado la notación de los factores de escala para utilizar la misma nomenclatura que utiliza Zhang al explicar su método.

Como el método asume que el sistema de coordenadas del mundo real se adapta a la propia plantilla, los puntos \mathbf{M}_i de la escena que se van a utilizar para realizar la calibración están colocados de forma que su coordenada $z = 0$.

El modelo $\lambda \cdot \mathbf{m} = \mathbf{A} \cdot [\mathbf{R} \ \mathbf{t}] \cdot \mathbf{M}$ se reduce eliminando la componente z . Definiendo los vectores columna de la matriz de rotación como $\mathbf{R} = [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3]$:

$$\mathbf{m} = \lambda \cdot \mathbf{A} \cdot [\mathbf{R} \ \mathbf{t}] \cdot \mathbf{M} = \lambda \cdot \mathbf{A} \cdot [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3 \ \mathbf{t}] \cdot \mathbf{M} = \lambda \cdot \mathbf{A} \cdot [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}] \cdot \mathbf{M} \quad (2.15.)$$

Es decir se transforma el modelo inicial en una homografía \mathbf{H} que relaciona los puntos de la plantilla plana del escenario con sus correspondientes en la imagen.

$$\lambda \cdot \mathbf{m} = \mathbf{H} \cdot \mathbf{M} \quad (2.16.)$$

Siendo

$$\mathbf{H} = \mathbf{A} \cdot [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}] = \mathbf{A} \cdot \begin{bmatrix} r_{11} & r_{21} & t_x \\ r_{12} & r_{22} & t_y \\ r_{13} & r_{23} & t_z \\ 0 & 0 & 1 \end{bmatrix} \quad (2.17)$$

El objetivo del método de Zhang es obtener los parámetros de la cámara a partir de $n \cdot m$ correspondencias $\mathbf{m}_{ij} \leftrightarrow \mathbf{M}_i$, ($i = 1, \dots, n$), ($j = 1, \dots, m$ con $m \geq 3$), donde el subíndice i indica el punto utilizado y el subíndice j la imagen utilizada.

Para ello, el método de Zhang sigue los siguientes pasos:

1. Cálculo de la homografía \mathbf{H} de cada imagen j .

Esta homografía puede ser calculada de varias formas. Se propone un método en dos pasos.

Primero se obtiene una aproximación inicial de la homografía (a la que denominaremos \mathbf{h}) mediante un método lineal. El método utilizado es el mismo que se ha descrito en el apartado 2.2.1. con la salvedad de que se ha eliminado la componente Z de las coordenadas de cada punto \mathbf{M}_i .

A partir de las correspondencias $\mathbf{m}_{ij} \leftrightarrow \mathbf{M}_i$ de una determinada imagen j , y siguiendo el mismo desarrollo que en el apartado 2.2.1. se llega a una ecuación similar a la (2.13) sin componente Z

$$\begin{bmatrix} X & Y & 1 & 0 & 0 & 0 & -u \cdot X & -u \cdot Y & -u \\ 0 & 0 & 0 & X & Y & 1 & -v \cdot X & -v \cdot Y & -v \end{bmatrix} \cdot \mathbf{h} = 0 \quad (2.18)$$

Si consideramos n puntos de interés, \mathbf{M}_i , se obtiene un sistema de $2n$ ecuaciones de la forma: $\mathbf{L} \cdot \mathbf{h} = 0$.

Una vez obtenida \mathbf{h} , se refina mediante un método iterativo para obtener \mathbf{H} . Este método trata de minimizar el error cometido al proyectar cada punto \mathbf{M}_i en la imagen, es decir, minimizar:

$$\sum_{i=1}^n \left\| m_i - \hat{m}_i \right\|^2 \quad (2.19)$$

donde \mathbf{m}_i son las coordenadas reales de cada punto del patrón en la imagen obtenida y que se obtienen marcándolos sobre la imagen, mientras que \hat{m}_i es la proyección en la imagen de \mathbf{M}_i aplicando los valores de la homografía que se va obteniendo en cada iteración (la primera es \mathbf{h}).

El paso 1 se aplica a todas las imágenes (n) que se van a utilizar para realizar el método.

2. Cálculo de la matriz cónica \mathbf{B} .

Para poder hallar la matriz de parámetros intrínsecos \mathbf{A} es necesario introducir un nuevo concepto: la “cónica absoluta”, \mathbf{B} . La “cónica absoluta” se define como:

$$\mathbf{B} = \mathbf{A}^t \cdot \mathbf{A}$$

Por lo que si se conoce \mathbf{B} se puede obtener la matriz de parámetros intrínsecos, \mathbf{A} , a partir de ella. Como \mathbf{B} es el producto de una matriz por su transpuesta, se trata de una matriz simétrica de 3×3 que se puede definir con un vector de 6 elementos de la forma: $\mathbf{b} = [b_{11} \ b_{12} \ b_{22} \ b_{13} \ b_{23} \ b_{33}]$.

A continuación se describe como se obtiene la “cónica absoluta” a partir de las n homografías \mathbf{H} de cada imagen obtenidas en el primer paso.

Dado que \mathbf{R} es una matriz de rotación, los vectores que la componen cumplen las restricciones de ortonormalidad, es decir:

$$\mathbf{r}_1^T \cdot \mathbf{r}_1 = \mathbf{r}_2^T \cdot \mathbf{r}_2 \quad \mathbf{r}_1^T \cdot \mathbf{r}_2 = 0 \quad (2.19)$$

Si se despejan los vectores de rotación, (\mathbf{r}_1 y \mathbf{r}_2) de la ecuación (2.17) y se sustituye su valor en las restricciones de la ecuación (2.19), se obtienen las dos restricciones básicas dada una homografía:

$$\mathbf{h}_1^T \cdot \mathbf{A}^{-T} \cdot \mathbf{A}^{-1} \cdot \mathbf{h}_1 = \mathbf{h}_2^T \cdot \mathbf{A}^{-T} \cdot \mathbf{A}^{-1} \cdot \mathbf{h}_2 \quad \mathbf{h}_1^T \cdot \mathbf{A}^{-T} \cdot \mathbf{A}^{-1} \cdot \mathbf{h}_2 = 0 \quad (2.20)$$

siendo \mathbf{h}_i la i -ésima columna de la homografía \mathbf{H} .

Si se escribe el producto $\mathbf{h}_i^T \cdot \mathbf{A}^{-T} \cdot \mathbf{A}^{-1} \cdot \mathbf{h}_i$ en función de la cónica absoluta se tiene:

$$\mathbf{h}_i^T \cdot \mathbf{B} \cdot \mathbf{h}_i = \mathbf{v}_{ij}^T \cdot \mathbf{b} \quad (2.21)$$

siendo

$$\mathbf{v}_{ij} = [h_{1i}h_{1j}, \quad h_{1i} \cdot h_{2j} + h_{2i}h_{1j}, \quad h_{2i}h_{2j}, \quad h_{3i} \cdot h_{1j} + h_{1i}h_{3j}, \quad h_{2i}h_{3j} + h_{3i}h_{2j}, \quad h_{3i}h_{3j}]^T$$

Las restricciones de la ecuación (2.20) se pueden escribir en función de \mathbf{v}_{ij} y de \mathbf{b} mediante dos ecuaciones homogéneas:

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ \mathbf{v}_{11}^T - \mathbf{v}_{22}^T \end{bmatrix} \cdot \mathbf{b} = 0 \quad (2.22)$$

Por cada imagen utilizada se obtienen dos ecuaciones homogéneas como las de la ecuación (2.22). Con las restricciones de n homografías correspondientes a n imágenes, ($n \geq 3$), se obtiene un sistema de $2n$ ecuaciones:

$$\mathbf{V} \cdot \mathbf{b} = 0 \quad (2.23)$$

donde \mathbf{V} es una matriz de dimensiones $2n \times 6$.

La solución es el vector propio de $\mathbf{V}^T \cdot \mathbf{V}$ asociado al valor propio más pequeño.

3. Cálculo de la matriz de parámetros intrínsecos \mathbf{A} , a partir de la matriz \mathbf{B} .

Una vez conocido el vector \mathbf{b} se pueden obtener los valores de la matriz de parámetros intrínsecos, \mathbf{A} (ver ecuación 2.1.). Los valores de los parámetros intrínsecos se muestran en la tabla 2.2.

Parámetro	Tipo	Valor
Coordenada v_0 del centro óptico	Intrínseco	$v_0 = \left(\frac{b_2 \cdot b_4 - b_1 \cdot b_5}{b_1 \cdot b_3 - b_2^2} \right)$
Factor de escala $\alpha = f \cdot k_u$	Intrínseco	$\alpha = \sqrt{\frac{\eta}{b_1}}$
Factor de escala $\beta = f \cdot k_v$	Intrínseco	$\beta = f \cdot k_v = -\ p_2^{p_3}\ $
Distorsión radial	Intrínseco	$\gamma = \frac{-b_2 \cdot \alpha^2 \cdot \beta}{\eta}$
Coordenada u_0 del centro óptico	Intrínseco	$u_0 = \frac{\gamma \cdot v_0}{\alpha} - \frac{b_4 \cdot \alpha^2}{\eta}$

Tabla 2.2. Parámetros intrínsecos calculados mediante el método de Zhang.

Siendo $\eta = b_6 - \left(\frac{b_4^2 + v_0 \cdot (b_2 \cdot b_4 - b_1 \cdot b_5)}{b_1} \right)$ un parámetro que ayuda a obtener expresiones compactas de los parámetros en función del vector \mathbf{b} .

4. Cálculo de los parámetros extrínsecos.

Los parámetros extrínsecos se obtienen despejando $[\mathbf{R} \ \mathbf{t}]$ de la definición de la homografía: $\lambda \cdot \mathbf{H}_j = \mathbf{A} \cdot [\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}]$ ya que se ha obtenido la homografía \mathbf{H} de cada imagen j , y la matriz de parámetros intrínsecos \mathbf{A} . El resultado obtenido es distinto para cada una de las imágenes seleccionadas.

Para despejar la matriz de parámetros extrínsecos se pasa la matriz de parámetros intrínsecos al otro lado de la definición de la homografía:

$$\lambda \cdot \mathbf{A}^{-1} \cdot \mathbf{H}_j = [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}] \quad (2.24)$$

Si se expresa la homografía \mathbf{H}_j en función de sus columnas \mathbf{H}_j^i , la ecuación (2.24) resulta de la siguiente forma, donde $\mathbf{A}^{-1} \mathbf{H}_j^i$ es la columna 'i' de la matriz resultante.

$$\lambda \cdot [\mathbf{A}^{-1} \cdot \mathbf{H}_j^1 \ \mathbf{A}^{-1} \cdot \mathbf{H}_j^2 \ \mathbf{A}^{-1} \cdot \mathbf{H}_j^3] = [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}] \quad (2.25)$$

En esta ecuación son conocidas las matrices \mathbf{A}^{-1} y \mathbf{H}_j , pero no el factor de escala λ . Para determinarlo, se tiene en cuenta que la norma de \mathbf{r}_1 es la unidad (ya que \mathbf{r} es la matriz de rotación).

$$\|\lambda \cdot \mathbf{A}^{-1} \cdot \mathbf{H}_j^1\| = \|\mathbf{r}_1\| = 1 \quad (2.26)$$

Se despeja $|\lambda|$ de la ecuación (2.24).

$$|\lambda| = \frac{1}{\|\mathbf{A}^{-1} \cdot \mathbf{H}_j^1\|} \quad (2.27)$$

Una vez obtenido $|\lambda|$, los parámetros extrínsecos se deducen directamente de la ecuación (2.25):

$$\mathbf{r}_1 = \frac{\mathbf{A}^{-1} \cdot \mathbf{H}_j^1}{\|\mathbf{A}^{-1} \cdot \mathbf{H}_j^1\|} \quad (2.28)$$

$$\mathbf{r}_2 = \frac{\mathbf{A}^{-1} \cdot \mathbf{H}_j^2}{\|\mathbf{A}^{-1} \cdot \mathbf{H}_j^1\|} \quad (2.29)$$

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \quad (2.30)$$

$$\mathbf{t} = \frac{\mathbf{A}^{-1} \cdot \mathbf{H}_j^3}{\|\mathbf{A}^{-1} \cdot \mathbf{H}_j^1\|} \quad (2.31)$$

5. Estimación de los coeficientes de distorsión radial por mínimos cuadrados.

El modelo de cámara en que se basa este método supone la corrección de la distorsión entre la aplicación de la matriz de parámetros extrínsecos y la matriz de parámetros intrínsecos. Es decir, primero se transforma del sistema de coordenadas del mundo real de la cámara (matriz de parámetros extrínsecos), luego se realiza la corrección de la distorsión, y por último, se obtienen las coordenadas 2D de la imagen (matriz de parámetros intrínsecos).

La diferencia con otros métodos estriba en que se corrige la distorsión antes de realizar la proyección de los puntos en el plano imagen, y por tanto las coordenadas no están afectas por la distancia focal. Por ese motivo, los valores de los coeficientes de distorsión no son comparables al del resto de los métodos que obtienen este parámetro.

Para ello se resuelven las siguientes ecuaciones para cada punto \mathbf{m}_i de cada imagen:

$$\begin{bmatrix} (u_p - u_0) \cdot r & (u_p - u_0) \cdot r^2 \\ (v_p - v_0) \cdot r & (v_p - v_0) \cdot r^2 \end{bmatrix} \cdot \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} u_i - u_p \\ v_i - v_p \end{bmatrix}$$

donde $r = (X_c^2 + Y_c^2) / Z_c^2$. (u_i, v_i) son las coordenadas de cada punto \mathbf{m}_i obtenido mediante marcado manual, y (u_p, v_p) son las coordenadas de cada punto \mathbf{m}_i obtenido aplicando el modelo de proyección con los valores de los parámetros obtenidos.

6. Se refinan todos los parámetros mediante un método iterativo que minimiza el error cuadrático medio.

Consiste en minimizar la diferencia entre las coordenadas de cada punto de cada una de las imágenes reales (u_p, v_p) con respecto a la proyección (u_i, v_i) realizada con los valores de todos los parámetros (incluidos los coeficientes de distorsión) obtenidos en la calibración de forma que todos ellos se puedan actualizar. Es decir, minimizar la función:

$$\sum_{i=1, p=1}^{n \cdot m} \left\| [u_i \quad v_i] - [u_p \quad v_p] \right\|^2 \quad (2.34)$$

2.2.3. Conclusiones sobre el estudio de los métodos de calibración.

La primera conclusión que se puede extraer es que existe un gran número de técnicas diferentes para la obtención de los parámetros del modelo *pinhole* de la cámara.

Las técnicas basadas en el conocimiento del escenario tienen un proceso de preparación más elaborado, ya que es necesario conocer la posición que ocupan, respecto al sistema de coordenadas del mundo, los puntos \mathbf{M}_i que se van a emplear para realizar la calibración. La exactitud de estas medidas hará que la estimación de los parámetros de la cámara sea más o menos ajustada. Se ha visto el método de Faugeras como ejemplo de técnica de calibración basada en el conocimiento del escenario.

En oposición a estos métodos, el método de Zhang no precisa de una preparación exhaustiva de la escena ya que el sistema de coordenadas del mundo se adapta a la propia plantilla que se utiliza para calibrar.

Otra diferencia entre el método de Faugeras y el método de Zhang es el tiempo de computación de cada uno. El método de Faugeras es un método lineal que utiliza una sola imagen para calibrar por lo que su tiempo de computación será muy breve. Por el contrario, el método de Zhang es un método no lineal que utiliza algoritmos iterativos de optimización y, además, utiliza varias imágenes para realizar la calibración. Por todo ello, su tiempo de computación será mucho mayor que el de Faugeras.

En el siguiente capítulo se muestra la implementación de ambos métodos. La razón de realizar unos métodos que ya están implementados es la voluntad de contar con un software propio que permita aislar el problema concreto de calibración que se afronta en este trabajo de otros posibles problemas y contextos que hayan afectado a otras implementaciones ya existentes.

2.3. Implementación de distintos métodos de calibración de cámaras.

En este capítulo se explica la implementación de los métodos de calibración de Faugeras y de Zhang en software propio.

2.3.1. Implementación del método de Faugeras.

El método de Faugeras se ha implementado en código Matlab. Los programas necesarios para su implementación son dos:

- *ginput3.m*: función que devuelve la posición en píxeles que ocupa en una imagen el punto que marque el usuario. Esta función es necesaria para conocer las coordenadas de los puntos \mathbf{m}_i
- *calibración_faugeras.m*: es el programa principal. Es un *script* que pide al usuario una serie de datos de entrada y calcula los parámetros de la cámara.

Los datos que el programa solicita al usuario son los siguientes:

- Coordenadas de los puntos 3D de la escena (\mathbf{M}_i), referidos al sistema de coordenadas del mundo, que se van a utilizar para realizar la calibración de la cámara.
- El programa muestra al usuario la imagen utilizada para la calibración y le pide que marque los puntos 2D de la imagen (\mathbf{m}_i) correspondientes a los puntos 3D de la escena (\mathbf{M}_i). Cada vez que el usuario marca un punto, el programa llama a la función *ginput3.m* para obtener las coordenadas de los puntos \mathbf{m}_i . Es indispensable que el usuario marque los puntos en el mismo orden que ha introducido las coordenadas de los puntos \mathbf{M}_i correspondientes.

Una vez obtenidos estos datos, el programa calcula los parámetros extrínsecos e intrínsecos de la cámara siguiendo el método explicado en el apartado 2.2.1. Estos parámetros son mostrados por consola.

2.3.2. Implementación del método de Zhang.

El método de Zhang se ha implementado en código Matlab. Con el objetivo de independizar en la medida de lo posible el desarrollo del método y la forma de obtener los datos de entrada del algoritmo (el número de imágenes empleadas, las coordenadas de los puntos \mathbf{M}_i y sus proyecciones \mathbf{m}_{ij} en cada imagen), se ha dividido la implementación en dos grandes bloques.

- **Programas de interfaz de usuario:** se encargan de recoger los datos de entrada que introduce el usuario y de almacenarlos en las variables correspondientes.
- **Programas de ejecución del método:** se encargan de aplicar el método de Zhang para calcular los parámetros de la cámara.

2.3.2.1. Programas interfaz de usuario.

Uno de los objetivos del método de Zhang es conseguir que la preparación de la escena y la obtención de los datos de entrada requieran el mínimo trabajo posible. Siguiendo esta línea se ha desarrollado una interfaz de usuario que trata de automatizar el trabajo en la medida de lo posible.

El usuario debe conocer e introducir únicamente la siguiente información:

- Número de imágenes: el número de imágenes que se van a considerar para la calibración.
- Número de puntos por imagen: el número de puntos por imagen que se van a considerar para la calibración..
- Longitud del lado del damero (expresada en mm) que se va a usar como plantilla. Para evitar que el usuario deba introducir manualmente las coordenadas de los puntos \mathbf{M}_i , estas coordenadas las calcula el programa de forma automática a partir del número de puntos por imagen y la longitud del tablero. La forma en la que el programa calcula las coordenadas provoca que se puedan considerar un número determinado de puntos para realizar la calibración. El número de puntos que se pueden considerar son 4, 9, 25 y 81. En la figura 2.6 se muestra los puntos concretos que se consideran y el orden en que son almacenados en las variables.
- Nombre con el que se han guardado las imágenes: el programa muestra cada imagen j al usuario y le pide que marque los puntos \mathbf{m}_{ij} correspondientes a los puntos \mathbf{M}_i .
- Marcado de los puntos: el usuario debe marcar los puntos \mathbf{m}_{ij} en el mismo orden que se han almacenado las coordenadas de los puntos \mathbf{M}_i correspondientes. Es decir, en el orden que muestra la figura 2.6. Cada vez que el usuario marca un punto \mathbf{m}_{ij} , el programa calcula sus coordenadas y las guarda en las variables que corresponden.

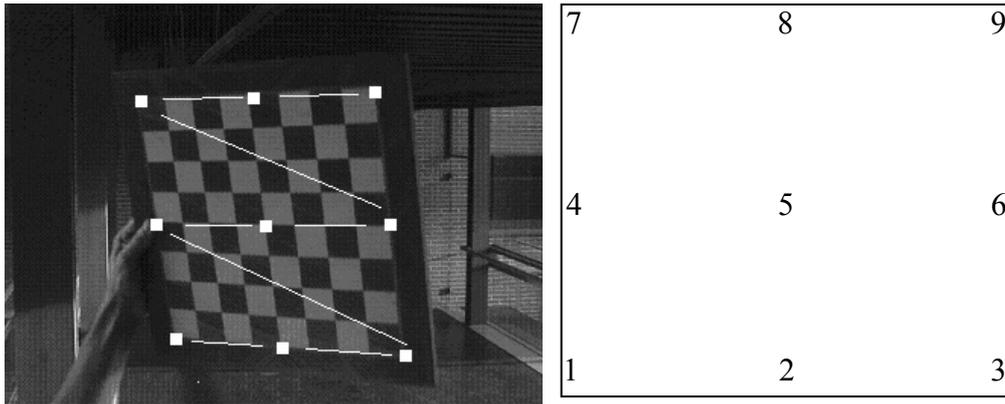


Fig. 2.6. Puntos que hay que marcar y orden en el que se marquen en el caso de considerar 9 puntos.

Con esta información se calculan todos los datos de entrada necesarios para la ejecución del método de Zhang. Los programas que se han desarrollado son:

- *zhang.m*: es un script que pide los datos necesarios al usuario, calcula las coordenadas 3D de los puntos \mathbf{M}_i , llama a las funciones que calculan las coordenadas 2D de los puntos \mathbf{m}_{ij} a partir de lo que marque el usuario en la imagen, llama al programa que implementa el método y muestra los resultados por pantalla.
- *ginput3.m*: función que devuelve la posición en píxeles que ocupa en una imagen el punto que marque el usuario. Esta función es necesaria para conocer las coordenadas de los puntos \mathbf{m}_{ij} .

2.3.2.2. Programas de ejecución del método.

El programa principal que implementa el método es una función Matlab, *calibración_zhang.m*, que tiene la siguiente interfaz:

Datos de entrada:

- **num_imag**: numero de imágenes con las que se va a calibrar la cámara.
- **coord_M**: coordenadas de los puntos de la escena referidas al sistema de coordenadas del mundo. Las unidades de las coordenadas son *mm*. El sistema de referencia de las coordenadas del mundo se elige de tal forma que dos de sus ejes coinciden con los lados de la plantilla, y el tercero (*Z*) es perpendicular a esta. El origen de este sistema de referencia podría ser cualquier punto de la plantilla. Sin embargo, el programa establece este punto en el centro del damero porque se ha comprobado que de esta forma se reduce el efecto de la distorsión radial.

- **coord_m**: coordenadas de los puntos \mathbf{m}_{ij} de la imagen referidos al sistema de coordenadas del plano normalizado de la imagen. Las unidades de las coordenadas son milímetros.

Datos de salida:

- **T**: vector de translación del sistema de coordenadas del mundo al sistema de coordenadas de la cámara. En este caso, como el sistema de coordenadas del mundo tiene su origen en el centro de la plantilla, el vector de translación mide la distancia entre la cámara y la plantilla. Las unidades del vector son mm.
- **R**: matriz de rotación del sistema de coordenadas del mundo al sistema de coordenadas de la cámara. Las unidades de cada elemento de la matriz son mm.
- **A**: matriz de parámetros intrínsecos. Las unidades de los factores de escala son *pixels* por mm. y las unidades de las coordenadas del punto principal son mm. El parámetro que mide la pérdida de ortogonalidad de la imagen es adimensional.
- **k**: coeficiente de distorsión.

Esta función implementa el método de Zhang siguiendo los seis pasos explicados en el apartado 2.2.2. *Método de Zhang [Zhang, 1998][Zhang, 2000]*. Durante su ejecución llama a dos pequeñas funciones que también se han implementado aparte:

- *zhang_hom.m*: función que calcula el error cometido entre la proyección teórica del punto en la imagen, (u_i, v_i) , y el punto que en realidad ocupa. Esta función es pasada como parámetro en la función iterativa que minimiza este error para refinar la aproximación inicial que se ha calculado de la homografía (paso 1), ecuación (2.19).
- *zhang_ref.m*: calcula la función $\sum_{i=1}^{n-m} \left\| \begin{bmatrix} u_i & v_i \end{bmatrix} - \begin{bmatrix} u_p & v_p \end{bmatrix} \right\|^2$, es decir, la diferencia entre las coordenadas de cada punto de cada una de las imágenes reales con respecto a la proyección, (u_i, v_i) , realizada con los valores de todos los parámetros (incluidos los coeficientes de distorsión) obtenidos en la calibración de forma que todos ellos se puedan actualizar (paso 6), ecuación (2.34).

2.4. Pruebas sobre cámaras fijas.

2.4.1. Cámaras fijas.

Las cámaras fijas con las que se van a realizar las pruebas del método de calibración de Zhang implementado en Matlab están ubicadas en la pasarela de la tercera planta que comunica los dos edificios de la EPS de la UAM.

El modelo de cámara es una SONY-DFW-X7. El modelo de lente es "Óptica montura C; distancia focal 8,5 mm; Apertura 1,5-C Modelo C815B de COSMICAR".



Fig. 2.7. Cámara fija SONY-DFW-X7 -

La resolución del CCD es de 1024 x 768 píxeles activos para una diagonal de la imagen de 5,952 mm., el tamaño de cada píxel es de 465 μ m. x 465 μ m. Es decir, los píxeles son cuadrados.

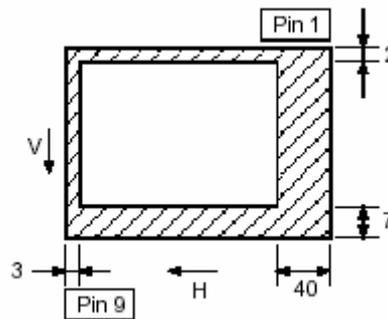


Fig. 2.8. Relación entre píxeles totales y píxeles activos de la cámara.

Con estos datos sobre la cámara se puede calcular los factores de escalado de cada eje:

- Como los píxeles son cuadrados, el factor de proporción es $s = 1$.
- La distancia entre el centro óptico y el plano imagen es de $f = 8,5$ mm.
- El número de píxeles por milímetro que utiliza la cámara es:
 - $d_u = 215,0538$ píxel/mm.
 - $d_v = 215,0538$ píxel/mm.
- Por tanto los factores de escala son:

$$\alpha = s \cdot f \cdot d_u = 1 \cdot 8,5 \text{ [mm]} \cdot 215,0538 \left[\frac{\text{píxel}}{\text{mm}} \right] = 1.828 \text{ [mm]}$$

$$\beta = f \cdot d_v = 8,5 \text{ [mm]} \cdot 215,0538 \left[\frac{\text{píxel}}{\text{mm}} \right] = 1.720,2 \text{ [mm]}$$

- El centro óptico de la cámara no se corresponde totalmente con el centro del plano imagen de la cámara.
 - $u_0 = 512 \text{ pixel}$.
 - $v_0 = 280 \text{ pixel}$

2.4.2. Patrón utilizado.

El patrón utilizado para la calibración (ver Fig. 2.9) es un damero de ajedrez de 320 mm. de lado. El damero tiene 64 casillas por lo que ofrece 81 esquinas como puntos de referencia para ser marcados.

2.4.3. Requisitos de las imágenes.

El método de calibración requiere una alta precisión al marcar los puntos para poder obtener resultados correctos. En concreto es necesario tomar las siguientes precauciones:

- El patrón debe ocupar casi toda la imagen para evitar que la obtención de las coordenadas 2D de los puntos en la imagen sea poco precisa. También es importante no apurar: que el patrón no ocupe la imagen entera porque la calibración se vería afectada por la distorsión de la lente, que es máxima en los bordes de la imagen (ver figura 2.4).
- Pese a que la teoría dice que con tres imágenes es posible obtener la solución, es recomendable que se utilicen 4 ó 5 vistas del patrón para que se produzca un resultado aceptable. Tomar más imágenes no es necesario según se ha comprobado porque no hay una mejora significativa.
- Se recomienda utilizar como origen de coordenadas del mundo el centro del patrón, ya que se producen menos errores de distorsión.
- El número mínimo de puntos necesarios por imagen para realizar la calibración es siete u ocho. En caso de utilizar más, se ha comprobado que las mejoras no son apreciables.

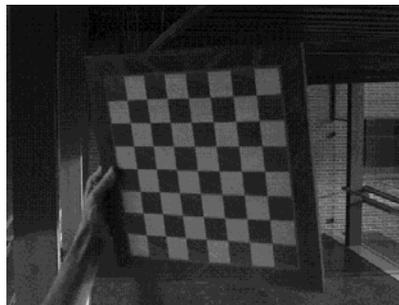


Fig. 2.9. Ejemplo de una imagen del patrón.

2.4.4. Desarrollo de las pruebas.

El desarrollo de las pruebas sobre cámaras fijas consiste en la toma de varias imágenes extraídas de una secuencia de vídeo grabada con la cámara, y la ejecución del método de Zhang implementado en Matlab.

Es necesario configurar la grabación de la secuencia con la resolución máxima (la que coincide con el tamaño del CCD) para realizar una calibración correcta. Ya que si no, las coordenadas de los puntos m_{ij} tendrán un error de cuantificación mucho mayor.

La resolución de las imágenes tomadas es de 1.024 x 768 píxeles (correspondiente al tamaño del CCD), estas imágenes se guardan en el mismo directorio en el que se encuentran los programas Matlab para su ejecución.

La ejecución del método (llamando al script *zhang.m*) ofrece la siguiente interfaz al usuario:

Indique el número de imágenes con las que se va a trabajar (al menos 3):

Indique el ancho del damero (mm):

Indique el número de puntos a considerar:

El número de imágenes que se ha escogido para realizar las pruebas es seis (ver figura 2.10) pues el método comienza a converger en realidad con 4 ó 5 imágenes y el hecho de aumentar el número de imágenes no mejora los resultados.

Estas imágenes han sido seleccionadas de una secuencia de vídeo grabada por la cámara SONY-DFW-X7 de la pasarela. En todas ellas el patrón está girado respecto de los tres ejes del espacio tridimensional, de forma que no existan dependencias lineales entre las posiciones relativas de los puntos que se van a seleccionar.

El ancho del damero de ajedrez utilizado es de 320 mm. Esta medida ha de hacerse desde la esquina de una casilla que esté en el extremo hasta la esquina de la casilla del extremo contrario. A partir de este dato (y del número de puntos seleccionado a partir de un protocolo), el método calcula de forma automática las coordenadas de los puntos de la escena que se van a marcar.

Como ya se ha indicado, el número de puntos que se pueden considerar por conveniencia ha de ser uno de los siguientes: {4, 9, 25, 81}. Se ha comprobado que tomar 9, 25 u 81 puntos no cambia los resultados. Sin embargo, con 4 puntos los resultados no convergen. Por lo tanto en este programa es recomendable utilizar 9 puntos por imagen.

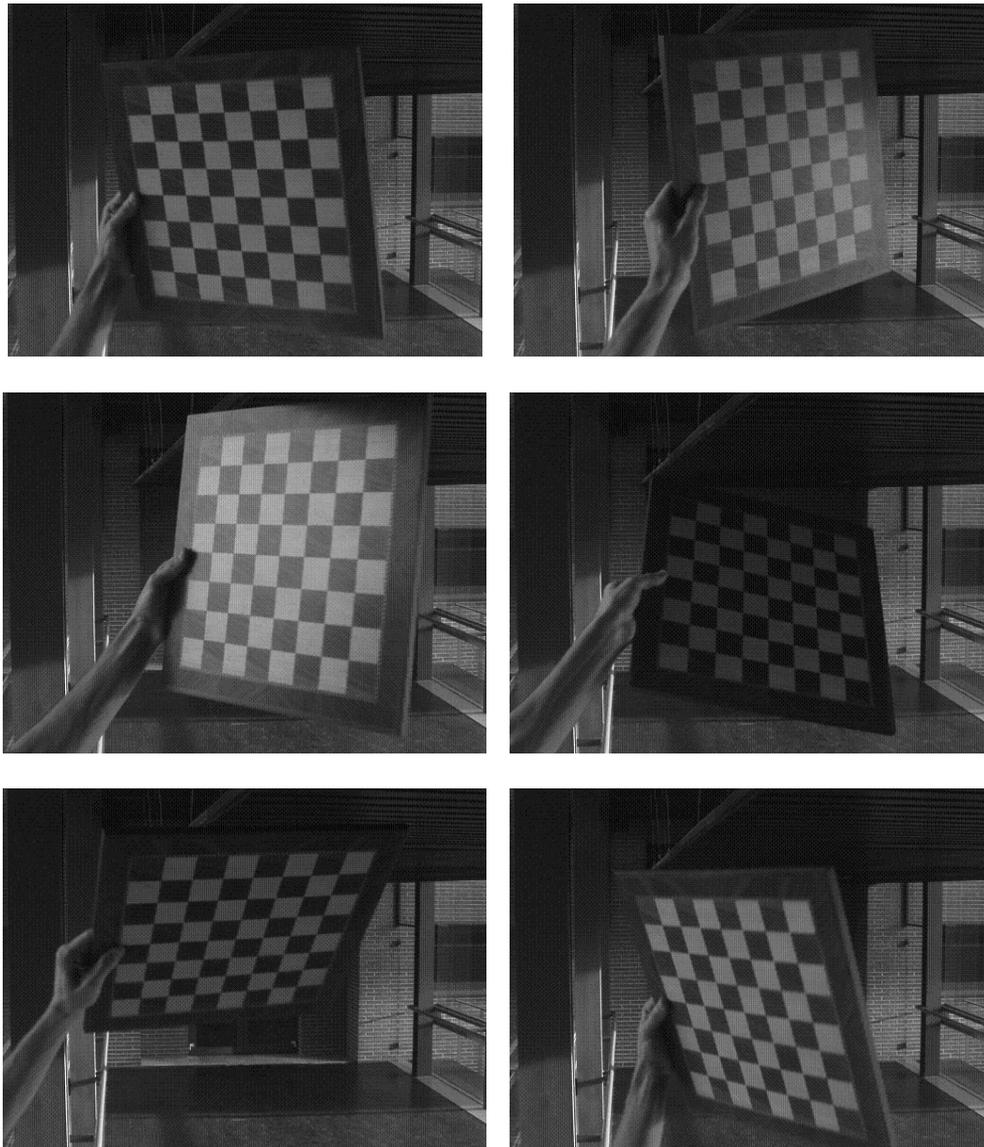


Fig. 2.10. Imágenes del patrón.

Una vez introducidos estos datos, el programa va a pedir tantas veces como número de imágenes se hayan introducido el siguiente dato:

Introduzca el nombre de la imagen con la que se va a calibrar la cámara:

Cada vez que se introduce el nombre de la imagen, el programa abre una figura con la imagen introducida y solicita al usuario que marque los puntos m_{ij} .

Estos puntos deben ser los indicados en la figura 2.11 en el caso, como se recomienda, de considerar 9 puntos:

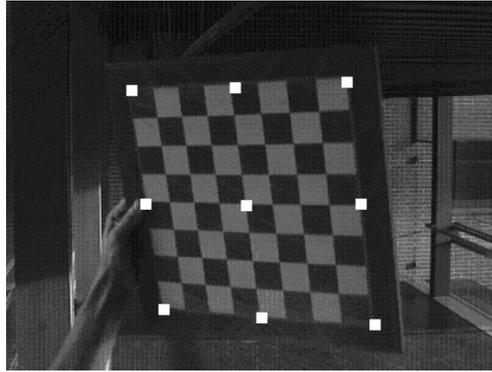


Fig. 2. 11. Puntos que se deben marcar cuando se consideran 9 puntos.

En el caso de considerar 25 puntos se marcan las esquinas que se señalan en la figura 3.12.

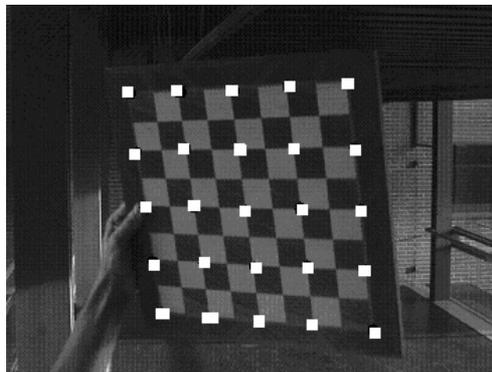


Fig. 2. 12. Puntos que se deben marcar cuando se consideran 25 puntos.

El propio programa va indicando en el título de la figura cual es el siguiente punto que se ha de marcar, y colorea de blanco el punto marcado para ayudar al usuario en la tarea. La precisión con la que se marcan los puntos es determinante para conseguir resultados precisos.

2.4.5. Resultados obtenidos de las pruebas.

El objetivo de las pruebas sobre cámaras fijas es evaluar los datos obtenidos de los parámetros intrínsecos. Como se desconoce la posición del tablero al grabar las imágenes no tiene sentido comprobar los parámetros extrínsecos pues no hay datos medidos con los que comparar los resultados obtenidos.

A continuación se muestran los resultados obtenidos para diferentes números de puntos e imágenes considerados. El resultado obtenido en cada caso se compara con el resultado real, que se conoce gracias a las especificaciones de la cámara detalladas en la sección 2.4.1..

De los resultados obtenidos cabe considerar lo siguiente:

1. Pese a que la teoría del método de Zhang establece que el número de imágenes necesarias para poder implementar el método, la realidad de la implementación es que es a partir de 5 imágenes cuando se obtienen resultados correctos.
2. A partir de 9 puntos considerados, el aumentar el número de puntos no mejora los resultados obtenidos.
3. El error cometido se debe a la precisión con la que se marcan los puntos correspondientes a las esquinas. El punto correspondiente a una esquina no suele corresponderse con un píxel en concreto y es imposible marcar siempre el mismo píxel en todas las imágenes mediante un marcado manual.

1. Resultados obtenidos en función del número de imágenes utilizadas.

Resultados con tres imágenes y nueve puntos:

	k_u [mm]	k_v [mm]	u_0	v_0
Datos reales	1.828	1.828	512	280
Resultados	1.910	1.861	580	318

Como se puede apreciar, pese a que teóricamente con tres imágenes el método converge, en la práctica el programa ofrece unos resultados muy distintos a los datos reales con este número de imágenes.

Resultados con cinco imágenes y nueve puntos

	k_u [mm]	k_v [mm]	u_0	v_0
Datos reales	1.828	1.828	512	280
Resultados	1.823	1.820	520	293

Como se puede comprobar, los resultados al considerar 5 imágenes en lugar de las teóricas 3, mejoran de forma muy significativa. Aún así, los datos que se obtienen no son del todo precisos. Esta imprecisión es consecuencia del marcado manual de los puntos. El error cometido no supera el 0,5 %, aunque varía dependiendo de la precisión con la que se logre marcar las esquinas.

Resultados con seis imágenes y nueve puntos

	k_u [mm]	k_v [mm]	u_0	v_0
Datos reales	1.828	1.828	512	280
Resultados	1.815	1.824	519	271

Al considerar más de cinco imágenes los resultados mejoran muy poco, y esta mejoría no se aprecia porque el error que se comete al marcar las esquinas es mayor.

La conclusión en cuanto al número de imágenes necesarias para calibrar correctamente la cámara con el método de Zhang desarrollado en Matlab, es que son necesarias 5

imágenes. Un menor número de imágenes ofrece resultados erróneos, mientras que un mayor número de imágenes no mejora los resultados obtenidos.

2. Resultados obtenidos en función del número de puntos utilizados.

El número de puntos necesarios para que el método converja es de ocho. En esta implementación se utilizan 9 porque permite que el usuario tenga que conocer pocos datos aprovechando la simetría del damero.

Aumentar el número de puntos (25, 81) no mejora los resultados pues el riesgo de perder precisión al marcar las esquinas de forma manual se multiplica.

3. Precisión al marcar los puntos.

Una alternativa al marcado manual es utilizar un detector de esquinas, como el de Harris. Con este sistema se conseguiría que en todas las imágenes se considerase el mismo píxel para cada punto (sin la aleatoriedad del marcado manual).

3. Autocalibración de cámaras PTZ.

3.1. Estudio de distintos métodos de autocalibración de cámara.

3.1.1. Introducción.

Los métodos de calibración clásicos, como los que se han estudiado en el capítulo 2, no permiten que se pueda realizar una calibración “*on-line*” cuando se está utilizando la cámara en tareas visuales.

Si la cámara no es fija y realiza algún tipo de movimiento, sus parámetros extrínsecos variarán. Este hecho provoca la necesidad de tener que volver a calibrar la cámara cada vez que ésta realiza cualquier tipo de movimiento. Evidentemente tratar de calibrar la cámara cada vez que ésta se mueve es inviable.

Además, la configuración interna de una cámara puede variar (movimiento de zoom, enfoque, desajustes mecánicos, variaciones térmicas, etc.) durante la ejecución de tareas *on-line*.

Todo esto ha provocado que se haya orientado una línea importante de trabajo al desarrollo de métodos que permitan una calibración automática, es decir, métodos de autocalibración.

Los métodos de autocalibración **no necesitan tener conocimiento de ningún dato externo**. En su lugar se utiliza la correspondencia de puntos detectados en la escena a lo largo de una secuencia de imágenes. Esto permite que un sistema que lleve incorporado una cámara pueda realizar de forma automática la calibración de ésta en cualquier momento, ya que simplemente necesita adquirir una secuencia de imágenes de su entorno. Generalmente, los métodos de autocalibración involucran un algoritmo de seguimiento de diferentes características en múltiples imágenes.

3.1.2. Clasificación de los métodos de autocalibración.

Se pueden realizar dos tipos de clasificación de los métodos de autocalibración: una clasificación puede hacerse en función del tipo de movimiento que realiza la cámara, y la otra en función de la variación o no de los parámetros intrínsecos:

- **Clasificación según el tipo de movimiento:**

General: la cámara puede realizar cualquier tipo de movimiento al estar situada sobre una plataforma móvil.

Rotaciones puras: la cámara se encuentra fija en una posición y puede realizar giros (normalmente sobre sus ejes x e y, azimut y elevación). Ejemplos de este tipo de cámaras son las de vigilancia o las que se utilizan para retransmisiones deportivas.

Traslaciones puras: la cámara no puede realizar ningún giro, sólo efectuar translaciones.

- **Clasificación según la variación de los parámetros intrínsecos:**

Fijos: se considera que la configuración de la cámara permanece estable entre las diferentes imágenes captadas. El principal problema que presentan estos métodos es que si varía algún parámetro intrínseco se debe reiniciar el proceso de calibración.

Variables: se estudia principalmente el efecto de la variación de la distancia focal o uso del *zoom*. Unos métodos consideran que esta variación afecta sólo a la distancia focal mientras que otros suponen que también afecta al valor del punto principal de la imagen.

En este proyecto se están considerando cámaras PTZ (*Pan-Tilt-Zoom*) por lo que el estudio de los métodos se centrará en aquellos que consideran que el movimiento de la cámara se limita a rotaciones puras y que los parámetros intrínsecos pueden variar debido al efecto del *zoom*.

- El control del movimiento *pan* (P) es el responsable de la rotación horizontal de la cámara al modificar este ángulo.
- El control del movimiento *tilt* (T) es el responsable de la rotación vertical de la cámara al modificar el ángulo vertical.
- El *zoom* (Z) es el responsable de que se considere que los parámetros intrínsecos varíen de una imagen captada a otra.

Aunque se ha señalado que el estudio de los métodos de autocalibración se va a centrar en los que consideran que la cámara sólo tiene movimiento de rotación y *zoom*, se va a presentar antes uno de los primeros métodos de autocalibración. Este método es el de Hartley (1994), y constituye la base de los demás métodos de autocalibración basados en el modelo *pin-hole*. La particularidad de este método es que no contempla el efecto del *zoom*.

3.1.3. Método de Hartley [Hartley, 1994a].

El método de Hartley se basa en la suposición de que la configuración interna de las cámaras permanece constante durante largo tiempo y que sólo desajustes de tipo mecánico o térmico afectan a las cámaras. Por tanto, sólo se hace necesaria la recalibración cada cierto tiempo para compensar estas pequeñas variaciones, y sólo en

los parámetros intrínsecos. Además asume que el movimiento de las cámaras es sólo de rotación.

Modelo de cámara

Como se explica en el capítulo 2 de la memoria (sección 2.1), en el modelo de cámara *pin-hole*, un punto \mathbf{M} del espacio proyectivo P^3 (cuatro coordenadas homogéneas y tres cartesianas) se proyecta en un punto \mathbf{m} del plano proyectivo P^2 (tres coordenadas homogéneas y dos cartesianas). Esta proyección se puede representar mediante la matriz de proyección (expresada en coordenadas cartesianas) (ver ecuación 2.6):

$$\lambda \cdot \mathbf{m} = \mathbf{P} \cdot \mathbf{M}$$

Esta matriz \mathbf{P} se puede descomponer en la matriz de parámetros intrínsecos (\mathbf{A}) y la matriz de parámetros extrínsecos ($[\mathbf{R} \ \mathbf{t}]$). Ambas matrices fueron definidas en la ecuación 2.5 del capítulo 2.

Cámara con movimiento de rotación

El objetivo de este método es determinar la matriz de parámetros intrínsecos (\mathbf{A}) de una cámara. La cámara que se considera es una cámara estacionaria que rota según distintas orientaciones. Por conveniencia, el origen del sistema de coordenadas del mundo coincide con el centro óptico de la cámara. De esta forma, el vector de translación es nulo. Como el vector de translación es nulo, se pierde la utilidad de utilizar las coordenadas homogéneas que se explicó en el capítulo 1 (realizar la rotación y la translación en un solo paso), por lo que en este capítulo no se van a considerar coordenadas homogéneas sino cartesianas. Por lo tanto la matriz de proyección que se considera es:

$$\mathbf{P}_j = \mathbf{A} \cdot \mathbf{R}_j \quad (3.1)$$

Donde \mathbf{P}_j es matriz de proyección para una posición j de la cámara, la matriz de parámetros intrínsecos se mantiene constante para cualquier posición de la cámara y \mathbf{R}_j es la matriz de rotación para dicha posición.

Se consideran los puntos \mathbf{m} y \mathbf{m}' de la figura 3.1. Dichos puntos son las proyecciones del mismo punto \mathbf{M} en dos imágenes distintas. En particular, el punto \mathbf{m} es la proyección de un punto \mathbf{M} de la escena que se está grabando para una determinada configuración de la cámara; y el punto \mathbf{m}' es la proyección del mismo punto \mathbf{M} después de haber rotado la cámara.

La relación de los puntos \mathbf{m} y \mathbf{m}' con el punto \mathbf{M} viene determinada por la matriz de proyección (que es distinta en cada caso porque la cámara se ha rotado):

$$\mathbf{m} = \mathbf{A} \cdot \mathbf{R} \cdot \mathbf{M} \quad \mathbf{m}' = \mathbf{A} \cdot \mathbf{R}' \cdot \mathbf{M}$$

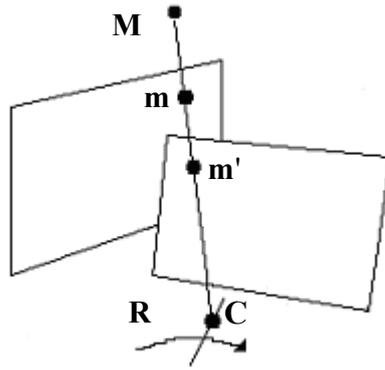


Fig. 3.1. Proyección del mismo punto M en dos planos de imagen diferentes.

Por lo que, despejando M de la primera relación, la relación entre los puntos m y m' es:

$$m' = A \cdot R' \cdot R^{-1} \cdot A^{-1} \cdot m \quad (3.2)$$

Se puede deducir que dadas dos imágenes tomadas por cámaras con los mismos parámetros intrínsecos, la misma localización y rotada una respecto a otra, existe una transformación proyectiva G que relaciona una imagen con la otra.

Esta proyección tiene la forma $G = A \cdot R'' \cdot A^{-1}$, siendo A la matriz de parámetros intrínsecos y R'' la matriz de rotación de una de las posiciones de la cámara con respecto a la otra. ($R'' = R' \cdot R^{-1}$)

En este apartado vamos a asociar el concepto cámara a una posición de rotación particular de la misma cámara. Consideremos pues que tenemos un número N de cámaras cuya matriz de proyección es $P_j = A \cdot R_j$ siendo $j = 0, \dots, N$. Por conveniencia asumimos que los ejes de coordenadas del mundo están elegidos para alinearse con la primera cámara que se toma como referencia. De forma que $R_0 = I$, la matriz identidad, y, por lo tanto, $P_0 = A$.

Se tiene que, dadas N imágenes tomadas por cámaras con los mismos parámetros intrínsecos la misma localización y rotada una respecto a otra, existen N transformaciones proyectivas, G_j , que relacionan la imagen de referencia con cada imagen.

$$G_j = A \cdot R_j \cdot A^{-1} \quad (3.3)$$

Siendo A la matriz de parámetros intrínsecos y R_j la matriz de rotación de la j -ésima cámara respecto de la primera.

Algoritmo de calibración

Se parte de una serie de N imágenes (J_0, \dots, J_N) solapadas, donde $N > 2$. Todas las imágenes han sido tomadas por la misma cámara con una rotación distinta en cada una.

Se pretende hallar la matriz de parámetros intrínsecos de la cámara. Los pasos del algoritmo son los siguientes:

1. Se establecen puntos correspondientes entre imágenes.
2. Para cada imagen rotada, $j = 1, \dots, N$ se calcula la transformación proyectiva \mathbf{G}_j (ecuación 3.3) entre la imagen de referencia, J_0 , y la j -ésima imagen, J_j .
3. Se calcula la matriz de parámetros intrínsecos \mathbf{A} que verifica para todo j : $\mathbf{A}^{-1} \cdot \mathbf{G}_j \cdot \mathbf{A} = \mathbf{R}_j$ siendo \mathbf{R}_j la orientación de la cámara cuando captura la j -ésima imagen con respecto a la posición de la cámara cuando captura la imagen de referencia.
4. Se refinan los parámetros estimados usando el proceso iterativo de Levenberg-Marquardt.

Mediante este método se podría obtener la matriz \mathbf{A} con los parámetros intrínsecos de la cámara. La ventaja que se obtiene es que no haría falta utilizar ningún patrón de calibración, ni tener ningún conocimiento sobre la escena. El mayor problema que presenta es la precisión a la hora de establecer los puntos correspondientes entre imágenes. Este paso es el primero del método, y el hecho de no lograr una buena precisión provoca que se arrastre un error significativo en el cálculo de los parámetros.

3.1.4. Métodos de autocalibración de cámaras con movimiento de rotación y con zoom.

El estudio de métodos de autocalibración de cámaras fijas con movimiento de rotación, como el de Hartley, está bastante desarrollado ya que existen muchas cámaras que sólo realizan rotaciones. Ejemplos de estas cámaras son las de vigilancia y las que cubren eventos deportivos.

Como la traslación es nula ($\mathbf{T} = \mathbf{0}$) se eliminan tres parámetros del problema y, por tanto, se simplifican los métodos con respecto a los que consideran que el movimiento de la cámara es general. El problema que puede presentar este tipo de métodos es que la traslación no se puede asegurar que sea nula, ya que, debido a la propia mecánica de las cámaras, su punto de giro puede no coincidir exactamente con su centro óptico.

En este apartado se incluye también, el estudio del efecto que tiene el zoom en los parámetros intrínsecos de la cámara. Debido al efecto del *zoom*, estos métodos incluyen en su modelo la variación de los parámetros intrínsecos durante la secuencia de imágenes utilizada para la autocalibración.

Los primeros métodos que consideraron este nuevo grado de libertad fueron realizados por Agapito [de Agapito et al., 1998] y Seo [Seo and Hong, 1998] [Seo and Hong, 1999].

Se trata de métodos iterativos que aplican las restricciones que impone la homografía en el infinito. Ambos métodos consideran que el punto principal y el factor de proporción son fijos, y que los ejes del plano imagen son ortogonales. Estos métodos se diferencian en el método de optimización utilizado: Agapito usa el de Levenberg-Marquardt mientras que Seo el método de Kantani.

Posteriormente, tanto Seo [Seo and K.S.Hong, 2001] como Agapito [de Agapito et al., 1999] han desarrollado variantes lineales de sus métodos.

Agapito desarrolla un método lineal muy similar al de Zhang y basado en la cónica absoluta. Este método precisa de, al menos, tres imágenes captadas para su funcionamiento, aunque pueden ser sólo dos si se conoce el punto principal de la imagen.

En su variante lineal Seo también considera conocido el centro del eje óptico, lo que simplifica las ecuaciones y consigue que sólo se necesiten dos imágenes para la calibración.

Agapito también ha desarrollado un método completo [Agapito et al., 2000] que se descompone en varias fases: obtención de una aproximación inicial mediante el método lineal [de Agapito et al., 1999]; aplicación del método iterativo [de Agapito et al., 1998]; y refinamiento de los parámetros limitando los errores geométricos, para lo que se minimiza la Estimación de la Máxima Probabilidad y la Estimación del Máximo a Posteriori.

En el siguiente apartado se va a estudiar en detalle el método de autocalibración de Agapito. Este método se plantea a raíz del método de Hartley ya estudiado e incluye el efecto que tiene el *zoom* en el cálculo de los parámetros intrínsecos de la cámara. En apariencia este método se plantea como el ideal para realizar la autocalibración de las cámaras PTZ, aunque a continuación estudiaremos los problemas que presenta.

3.1.5. Método lineal de Agapito.

El objetivo del método lineal de Agapito es hallar los parámetros intrínsecos de la cámara. Para ello es necesario disponer de una serie de imágenes que graben una misma escena y que hayan sido tomadas con variaciones de rotación y del *zoom*. Como en toda técnica de autocalibración no es necesario conocer ninguna información sobre la escena que se graba.

El método lineal de Agapito se basa en la obtención de la cónica absoluta a partir de la homografía entre imágenes mediante un algoritmo lineal.

Es necesario obtener al menos tres imágenes de la misma escena con distintas orientaciones de la cámara para poder alcanzar resultados correctos. Se considera, por tanto, que la escena permanece inmóvil mientras se mueve la cámara.

Es necesario establecer las siguientes restricciones:

1. Los ejes del plano imagen son perfectamente ortogonales por lo que las imágenes no sufren distorsión tangencial.
2. Los píxeles son cuadrados por lo que $\alpha = \beta$.

Los pasos que hay que realizar son los siguientes:

1. Obtención de la homografía de cada imagen respecto a una que se toma como referencia.

Un punto 2D de la i -ésima imagen, \mathbf{m} , se corresponde con un punto 3D de la escena. A partir de la ecuación 2.6 se tiene:

$$\mathbf{M} \overset{\circ}{=} \mathbf{P}_i^{-1} \cdot \mathbf{m} \quad (3.4)$$

El mismo punto \mathbf{M} de la escena se corresponde con otro punto 2D de la j -ésima imagen, \mathbf{m}' mediante la matriz de proyección \mathbf{P}_j :

$$\mathbf{m}' \overset{\circ}{=} \mathbf{P}_j \cdot \mathbf{M} \quad (3.5)$$

Por lo tanto, combinando las ecuaciones (3.4) y (3.5) se puede determinar que los puntos \mathbf{m} y \mathbf{m}' (correspondientes al mismo punto 3D de la escena) está relacionados de la siguiente forma:

$$\mathbf{m}' \overset{\circ}{=} \mathbf{P}_j \cdot \mathbf{P}_i^{-1} \cdot \mathbf{m} \quad (3.6)$$

La homografía entre dos puntos de dos imágenes correspondientes al mismo punto \mathbf{M} es:

$$\mathbf{H}_{ij} = \mathbf{P}'_j \cdot \mathbf{P}_i^{-1} \quad (3.7)$$

Para calcular la homografía entre dos imágenes, i y j , es necesario considerar al menos cuatro puntos de un mismo plano de cada imagen (ya que una homografía en P^n queda definida por $n + 2$ puntos, luego en P^2 se define con cuatro puntos). En [Hartley, 1997] se indica un método práctico para obtener la homografía a partir de las coordenadas de cuatro puntos de cada imagen.

2. Obtención de la “cónica absoluta”.

La cónica absoluta, \mathbf{w} , se define a partir de la matriz de parámetros intrínsecos, \mathbf{A} :

$$\mathbf{w} = (\mathbf{A}^{-1})^T \cdot \mathbf{A}^{-1} \quad (3.8)$$

Si descomponemos la matriz de proyección en la matriz de parámetros intrínsecos y la matriz de rotación ($\mathbf{P} = \mathbf{A} \cdot \mathbf{R}$ ya que $\mathbf{t} = 0$), la homografía de la ecuación (3.7) queda de la siguiente forma:

$$\mathbf{H}_{ij} = \mathbf{A}_j \cdot \mathbf{R}_j \cdot \mathbf{R}_i^{-1} \cdot \mathbf{A}_i^{-1} = \mathbf{A}_j \cdot \mathbf{R}_{ij} \cdot \mathbf{A}_i^{-1} \quad (3.9)$$

donde \mathbf{R}_{ij} es la rotación de j con respecto a i (según se ha visto en el estudio del método de Hartley).

Como $\mathbf{R}_{ij} \cdot \mathbf{R}_{ij}^T = \mathbf{I}$, la ecuación (3.9) queda de la siguiente forma:

$$\mathbf{H}_{ij} \cdot \mathbf{w}_i \cdot \mathbf{H}_{ij}^T = \mathbf{w}_j \quad (3.10)$$

donde $\mathbf{w}_i = \mathbf{A}_i \cdot \mathbf{A}_i^T$. Se expresa la ecuación (3.10) en función de la cónica absoluta de cada imagen, para ello se aplica la inversa y se obtiene:

$$\mathbf{H}_{ij}^{-T} \cdot \mathbf{w}_i \cdot \mathbf{H}_{ij}^{-1} = \mathbf{w}_j \quad (3.11)$$

Por lo que se obtiene una ecuación que permite relacionar la cónica de una imagen con la cónica de otra. A partir de la ecuación 3.10 si se toma una de las imágenes como referencia (cuya cónica es $\mathbf{w}_j = \mathbf{w}_0$), se puede obtener la cónica de cada imagen a partir de la cónica de la imagen de referencia y la homografía de la imagen i con respecto a la imagen de referencia:

$$\mathbf{w}_i = \mathbf{H}_{i0}^{-T} \cdot \mathbf{w}_0 \cdot \mathbf{H}_{i0}^{-1} \quad (3.12)$$

Teniendo en cuenta las restricciones del método, la cónica absoluta expresada en función de los parámetros intrínsecos es (a partir de la ecuación 3.8):

$$\mathbf{w} = \begin{bmatrix} 1/\alpha^2 & 0 & -u_0/\alpha^2 \\ 0 & 1/\alpha^2 & -v_0/\alpha^2 \\ -u_0/\alpha^2 & -v_0/\alpha^2 & 1 + (u_0^2 + v_0^2)/\alpha^2 \end{bmatrix}$$

A la vista de esta equivalencia se pueden deducir dos ecuaciones para cada imagen:

$$\mathbf{w}_i(1, 2) = 0 \quad \mathbf{w}_i(1, 1) = \mathbf{w}_i(2, 2). \quad (3.13)$$

El conjunto de todas las ecuaciones, incluyendo las ecuaciones sobre la imagen de referencia se almacena en una matriz \mathbf{E} , donde cada fila es una ecuación. Éstas se forman aplicando estas restricciones a la ecuación (3.12).

Con las dos restricciones que se obtienen por cada imagen, se forma un sistema de ecuaciones de la forma $\mathbf{E} \cdot \mathbf{a} = \mathbf{0}$, donde \mathbf{a} es un vector formado por los elementos de la cónica absoluta, \mathbf{w} .

Para resolverlo, basta con encontrar el vector \mathbf{a} que minimice $\|\mathbf{E}\mathbf{a}\|$ tal que $\|\mathbf{a}\| = 1$. Esto se consigue descomponiendo en valores singulares la matriz \mathbf{E} , y tomando como solución el autovector correspondiente al auto valor más pequeño.

Al obtener \mathbf{a} se reconstruye la matriz \mathbf{w}_0 . Luego se aplica la ecuación (3.12) para obtener las matrices \mathbf{w}_i de cada una de las imágenes.

A partir de la cónica absoluta se haya la matriz de parámetros intrínsecos, \mathbf{A} , que era el objetivo inicial del método.

3.1.5. Conclusiones sobre el estudio de distintos métodos de autocalibración.

Es este apartado se han estudiado las ventajas que ofrecen las técnicas de autocalibración sobre las técnicas de calibración. Estas son fundamentalmente no tener que conocer ningún dato de la escena que se graba y poder realizar una calibración *online*.

El método de autocalibración de Agapito parece, en un principio, la técnica que mejor se adapta a nuestro contexto (cámaras con movimiento de rotación y con *zoom*). Mediante este método se pueden calcular, a priori, los parámetros intrínsecos de las cámaras.

En el siguiente apartado se va a estudiar el estado del arte del proceso de autocalibración de cámaras PTZ. En este apartado se van a conocer las inconsistencias del método de Agapito cuando la cámara realiza un movimiento *pan-tilt*.

3.2. Estudio del estado del arte del proceso de autocalibración de cámaras PTZ.

3.2.1. Autocalibración de un sistema “*arm-eye*”.

Una cámara PTZ es un sistema formado por una cámara con distancia focal variable (*zoom*) y una plataforma que permite un movimiento de rotación con dos grados de libertad (*pan-tilt*). Este sistema recibe el nombre de sistema “*arm-eye*” en la literatura científica.



Fig. 3.2. Cámara PTZ.

En el apartado referente al estudio de los métodos de autocalibración, se ha señalado que de entre todos los métodos de autocalibración los que más se relacionan con un sistema *pan-tilt-zoom* son aquellos que consideran que la cámara es fija con movimiento de rotación, realiza movimientos de rotación pura y la distancia focal puede variar. Se ha visto que los métodos más destacados con estas características son los de Agapito y Seo.

El problema que presentan estos métodos es que si la cámara realiza algunos movimientos particulares, es imposible determinar algunos de los parámetros intrínsecos.

En las evaluaciones experimentales de estos métodos se ha comprobado que los resultados obtenidos son muy inestables a no ser que el movimiento de la cámara contenga alguna rotación sobre el eje principal (eje Z) de la cámara. Si existe alguna rotación sobre el eje Z, los resultados obtenidos son buenos.

Precisamente, si el movimiento de la cámara es *pan-tilt* **es imposible determinar los parámetros intrínsecos β y α** , pues un movimiento puro de rotación sobre el eje x o sobre el eje y no aporta ninguna información sobre este parámetro. Por tanto la restricción que se impone en [Agapito et al.1999] (ejes perfectamente ortogonales por lo que no existe distorsión tangencial) es insuficiente para determinar todos los parámetros intrínsecos de la cámara.

La explicación de esto debe hacerse a raíz del estudio del método de calibración de Agapito: el movimiento de una cámara sobre un sistema *pan-tilt* con respecto a su posición de referencia puede denotarse como $\mathbf{R}_x \cdot \mathbf{R}_y$. Es decir una rotación sobre el eje vertical (*panning*) seguido de una rotación sobre el eje horizontal (*tilting*). Esta matriz de rotación tiene la propiedad que su elemento (1, 2) es cero. Por lo que la homografía entre las dos imágenes tiene la siguiente forma (3.9):

$$\mathbf{H}_{0i}^{-1} = \mathbf{A}_0 \cdot \mathbf{R}_x \mathbf{R}_y^T \cdot \mathbf{A}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix}$$

Si tomamos la ecuación que relaciona la cónica de una imagen i con la cónica de la imagen de referencia (ver ecuación (3.12)):

$$\mathbf{w}_i = \mathbf{H}_{0i}^{-T} \cdot \mathbf{w}_0 \cdot \mathbf{H}_{0i}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix} \cdot \begin{bmatrix} a_1 & a_2 & a_3 \\ a_2 & a_4 & a_5 \\ a_3 & a_5 & a_6 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix}$$

Siguiendo el método de Agapito, se puede deducir que el elemento (1,2) de \mathbf{w}_i no depende del elemento a_4 de \mathbf{w}_0 . Esto implica que el factor de escala β no puede ser determinado. Se puede hacer el mismo razonamiento considerando que el movimiento *pan-tilt* con respecto a su posición de referencia puede denotarse como $\mathbf{R}_y \cdot \mathbf{R}_x$ para demostrar que el factor de escala α tampoco puede ser determinado.

Este problema ha sido estudiado en [Kahl et al.2000] pero en el contexto de cámaras que también pueden realizar un movimiento de translación. Una contribución importante de este trabajo es la determinación de las inconsistencias que se cometen al calcular los parámetros intrínsecos con respecto a los ángulos de rotación.

En [Hayman 2000] se ofrece una clasificación de las inconsistencias que se producen en los métodos de autocalibración cuando la cámara tiene un movimiento pan-tilt, un movimiento pan puro o un movimiento tilt puro. En este trabajo se vuelve a demostrar que no es suficiente considerar constante el factor de escala para solucionar el problema, sino que éste debe ser conocido.

En conclusión, aunque en un principio se podía plantear el método de Agapito como el más idóneo para calcular los parámetros intrínsecos de la cámara y la posición de ésta (parámetros extrínsecos), la realidad es que no se puede determinar el factor de escala, por lo que no se pueden calcular los parámetros intrínsecos. Por esta razón el método de Agapito, aunque a priori parecía el más idóneo, no puede ser utilizado en nuestro contexto.

En el siguiente apartado se va a proponer una forma de calibrar una red de cámaras PTZ.

3.2.2. Calibración de una red de cámaras PTZ.

En este capítulo se discute el problema de calibrar una red de cámaras PTZ. Este estudio ha sido realizado por Studipta N. Sinha y Marc Pollefeys [Sinha y Pollefeys, *Towards Calibrating a Pan-Tilt-Zoom Camera Network*] con el objetivo de poder modelar y reconstruir una escena 3D.

Los parámetros intrínsecos de cada cámara dependen del *zoom* empleado y se estiman en dos pasos. Primero se calculan los parámetros intrínsecos correspondientes al *zoom* más pequeño y después se estiman los parámetros intrínsecos al aumentar el *zoom* de forma discreta. Ambos pasos son automáticos y no es necesario conocer ningún dato de la estructura de la escena.

El estudio describe un método automático para calibrar una red de cámaras PTZ sin necesitar ningún acceso físico a las cámaras o a la escena donde se van a capturar las vistas. Se van a estimar los valores de los parámetros intrínsecos para todo el rango posible de niveles de *zoom*. El método no requiere ningún patrón ni otra estructura para realizar la calibración.

Modelo de cámara

El modelo de cámara asume que el centro de rotación de la cámara es fijo y coincide con el centro de proyección de la cámara durante la operación de calibración.

Como se explica en el capítulo 2 de la memoria, en el modelo de cámara pin-hole, un punto \mathbf{M} del espacio proyectivo P^3 (cuatro coordenadas homogéneas y tres cartesianas) se proyecta en un punto \mathbf{m} del plano proyectivo P^2 (tres coordenadas homogéneas y dos cartesianas). Esta proyección se puede representar mediante la matriz de proyección (ecuación (2.6) expresada en coordenadas cartesianas añadiendo el parámetro s de pérdida de ortogonalidad):

$$\lambda \cdot \mathbf{m} = \mathbf{P} \cdot \mathbf{M} \quad \mathbf{P} = \mathbf{A} \cdot [\mathbf{R} \quad \mathbf{t}] \quad \mathbf{A} = \begin{pmatrix} k_u & s & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

Esta matriz \mathbf{P} se puede descomponer en la matriz de parámetros intrínsecos (\mathbf{A}) y la matriz de parámetros extrínsecos ($[\mathbf{R} \quad \mathbf{t}]$).

Como las cámaras PTZ tienen movimientos de rotación puros referidos al centro del eje óptico, se elige este punto como centro de las coordenadas del mundo de forma que la translación es nula, $\mathbf{t} = \mathbf{0}$.

El objetivo es estimar los parámetros intrínsecos desconocidos en función del *zoom* y las rotaciones *pan* y *tilt*.

El punto principal, (u_0, v_0) , y la distancia focal, f , dependen únicamente del nivel del *zoom* y no dependen del estado de las rotaciones *pan* y *tilt*. Por lo que la matriz de parámetros intrínsecos que se desea estimar es función únicamente del nivel del *zoom* z :

$$\mathbf{A}(z) = \begin{pmatrix} k_u(z) & 0 & u_0(z) \\ 0 & k_v(z) & v_0(z) \\ 0 & 0 & 1 \end{pmatrix} \quad (3.14)$$

Como se ha visto, la proyección real del punto \mathbf{M} de la escena en un punto \mathbf{m} de la imagen difiere de la proyección ideal (m_{ideal}) del modelo *pin-hole* por causa de la distorsión radial.

Si definimos r como la distancia radial del punto $\mathbf{m} = (u, v)$ al punto principal de la cámara (u_0, v_0):

$$r = \sqrt{u^2 + v^2}$$

Entonces $L(r)$ es el factor de distorsión del punto real respecto al punto ideal:

$$L(r) = 1 + a_1 \cdot r^2 + a_2 \cdot r^4 + a_3 \cdot r^6 + \dots$$

Este factor de distorsión depende de los parámetros k_i , y de la posición del punto de la imagen con respecto al punto principal (r). Los parámetros k_i dependen de la distancia focal, de forma que el efecto de la distorsión radial disminuye según aumenta la distancia focal, por lo que el modelo de distorsión (al que llamaremos RD) radial también depende del zoom.

$$\begin{pmatrix} u_{ideal} \\ v_{ideal} \end{pmatrix} = L(r) \cdot \begin{pmatrix} u \\ v \end{pmatrix} \quad RD(z) = (a_1(z), a_2(z), a_3(z), u_0(z), v_0(z))$$

El proceso de calibración consistirá en calcular $\mathbf{A}(z)$ y $\mathbf{RD}(z)$ para todo el rango posible de *zooms* desde el *zoom* mínimo hasta el máximo avanzando en pasos discretos según una escala logarítmica.

Cámaras con zoom y con movimiento de rotación

Se consideran los puntos \mathbf{m} y \mathbf{m}' . Dichos puntos son las proyecciones del mismo punto \mathbf{M} en dos imágenes distintas. El punto \mathbf{m} es la proyección de un punto \mathbf{M} de la escena que se está grabando para una determinada configuración de la cámara; y el punto \mathbf{m}' es la proyección del mismo punto \mathbf{M} después de haber rotado, cambiado el *zoom* de la cámara o ambas cosas.

La figura 3.3 la correspondencia de un mismo punto (\mathbf{m}_a) en dos imágenes distintas en las que se ha variado el *zoom*, y la correspondencia de un mismo punto (\mathbf{m}_b) en dos imágenes distintas en las que se ha rotado la cámara

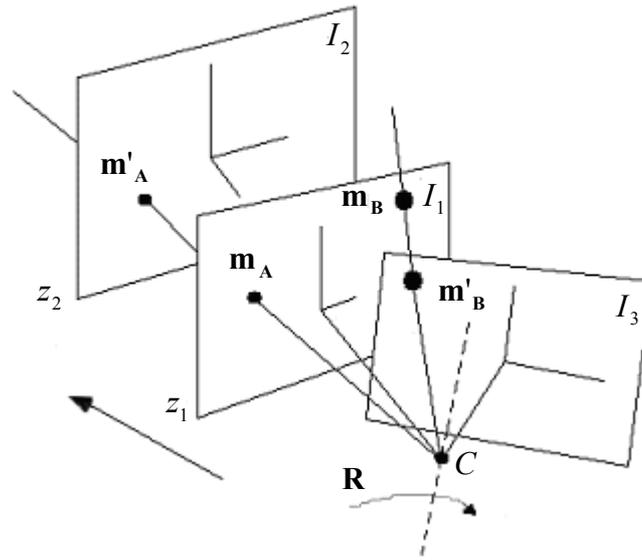


Fig. 3.3. Correspondencia entre puntos de distintas imágenes cuando hay un movimiento de rotación y de zoom.

La relación de los puntos \mathbf{m} y \mathbf{m}' con el punto \mathbf{M} viene determinada por la matriz de proyección (que es distinta en cada caso porque la cámara se ha rotado y/o se le ha cambiado el nivel de zoom):

$$\mathbf{m} = \mathbf{A} \cdot \mathbf{R} \cdot \mathbf{M} \quad \mathbf{m}' = \mathbf{A}' \cdot \mathbf{R}' \cdot \mathbf{M}'$$

Por lo que la relación entre los puntos \mathbf{m} y \mathbf{m}' es:

$$\mathbf{m}' = \mathbf{A}' \cdot \mathbf{R}' \cdot \mathbf{R}^{-1} \cdot \mathbf{A}^{-1} \cdot \mathbf{m} \quad (3.15)$$

Los parámetros intrínsecos permanecen iguales para una rotación pura si se mantiene constante el nivel del *zoom*. Por lo que si consideramos que de una vista a otra no variamos el nivel del *zoom*, la ecuación (3.15) se reduce a:

$$\mathbf{m}' = \mathbf{A}' \cdot \mathbf{R}_{\text{ref}} \cdot \mathbf{A}^{-1} \cdot \mathbf{m} \quad (3.16)$$

Donde $\mathbf{R}_{\text{ref}} = \mathbf{R}' \cdot \mathbf{R}^{-1}$ representa la rotación de la cámara relativa al centro de proyección entre las dos vistas; y \mathbf{A} es la matriz de parámetros intrínsecos para un nivel de *zoom* determinado.

De la misma forma si la cámara sólo varía su *zoom* y no efectúa ninguna rotación entre la toma de las dos vistas, la ecuación (3.15) se reduce a:

$$\mathbf{m}' = \mathbf{A}' \cdot \mathbf{A}^{-1} \cdot \mathbf{m} \quad (3.17)$$

Ambas operaciones representan homografías que vienen dadas por \mathbf{H}_{rot} y \mathbf{H}_{zoom} .

$$\mathbf{H}_{\text{rot}} = \mathbf{A}' \cdot \mathbf{R} \cdot \mathbf{A}^{-1} \quad (3.18)$$

$$\mathbf{H}_{\text{zoom}} = \mathbf{A}' \cdot \mathbf{A}^{-1} \quad (3.19)$$

Procedimiento de calibración

Como ya se ha indicado, el algoritmo de calibración tiene dos fases:

1. Se estiman los parámetros intrínsecos correspondientes al menor nivel de *zoom*, con los que se obtiene \mathbf{A} .
2. Se calculan los parámetros intrínsecos correspondientes a varios incrementos discretos del *zoom*, con los que se obtiene \mathbf{A}' .

La notación que se va a seguir en este desarrollo es la siguiente:

- \mathbf{I}_i son las imágenes capturadas por la cámara. Estas imágenes se guardan en una rejilla según sea la posición de rotación de la cámara (ver figura 3.5).
- \mathbf{H}_i y \mathbf{V}_j representan las homografías (\mathbf{H}_{rot}) entre dos imágenes adyacentes horizontal y verticalmente.
- \mathbf{T}_i representa la homografía (\mathbf{H}_{rot}) entre cualquier imagen y la imagen que se escoja como referencia (\mathbf{I}_r).

1. Estimar los parámetros intrínsecos para el nivel de zoom mínimo.

El primer paso para hallar los parámetros intrínsecos es determinarlos para el nivel de *zoom* más bajo. Esto se hace a partir de imágenes capturadas por la cámara cuando efectúa un movimiento de rotación (no se varía el *zoom* en ningún momento). Se debe calcular la homografía, \mathbf{H}_{rot} , entre cada una de las imágenes capturadas y una que se escoja como referencia.

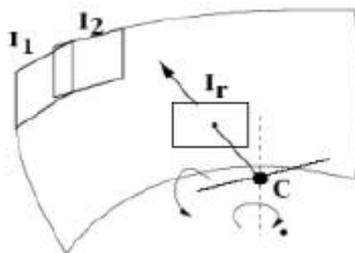


Fig. 3.4. Imágenes capturadas por la cámara cuando se efectúa un movimiento de rotación.

Durante la fase de captura, las imágenes adquiridas se disponen en una rejilla para cada paso *pan* y *tilt*. La figura 3.5. muestra 18 imágenes capturadas en 6 pasos *pan* y 3 pasos *tilt* con el mismo nivel de *zoom*:

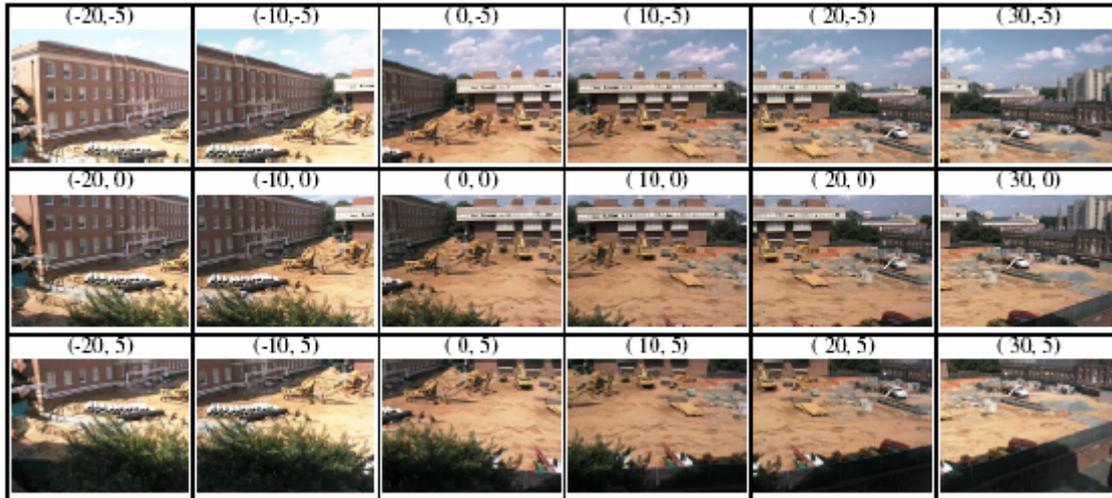


Fig. 3.5. Figura extraída de [*Towards Calibrating a PTZ Camera Network* (Sinha y Pollefeys)].

Se calculan las homografías (\mathbf{H}_{rot}) entre todas las parejas de imágenes adyacentes horizontales, \mathbf{H}_i , y todas la parejas de imágenes adyacentes verticales, \mathbf{V}_i siguiendo el método expuesto en la ecuación 3.18.

Una de las imágenes, \mathbf{I}_r , es elegida como imagen de referencia, y se calculan las homografías \mathbf{T}_i entre cada imagen y la imagen de referencia. Cada homografía \mathbf{T}_i se calcula como una composición de secuencias de transformaciones ($\dots \mathbf{H}_a, \mathbf{H}_b, \dots \mathbf{V}_c, \mathbf{V}_d \dots$) a través de las imágenes que conectan cada imagen con la imagen de referencia.

Una estimación exacta de las homografías \mathbf{T}_i para todas las imágenes permitiría el alineamiento multi-imagen en el plano de la imagen de referencia. Pero si se acumulan errores residuales en la composición de las homografías el mosaico final que se obtendrá al alinear las imágenes tendrá errores significativos.

La alineación global de imágenes se logra a través de un conjunto de ajustes (es necesario usar una implementación eficiente para tratar con un alto número de imágenes).

El ajuste del conjunto se inicializa a partir el conjunto de homografías calculadas, \mathbf{T}_i , y construyendo una lista de características globales a partir de las coincidencias entre las parejas de imágenes horizontales y verticales.

El ajuste del conjunto realiza una minimización global que produce la estimación de máxima igualdad de los parámetros del modelo cuando se asume que el error cometido entre el punto real de la imagen y el punto ideal (ver capítulo 2) es un ruido gaussiano de media cero.

Esto es usado primeramente para estimar todas las homografías \mathbf{T}_i , los parámetros de distorsión global, $RD(z_{min})$, y las características de la imagen, I_0 que minimiza el error cuadrático medio entre las características de la imagen y los puntos proyectados.

$$\min_{T_i, RD(z_{min}), M^j} \sum_{k=1}^m \sum_{i=1}^n D(m_i^j, RD(T_i I_0))^2$$

A partir de las homografías calculadas, \mathbf{T}_i , se estima la matriz de parámetros intrínsecos correspondiente al nivel de zoom mínimo, $\mathbf{A}(z_{min})$, utilizando el algoritmo lineal de Hartley para calcular parámetros intrínsecos de una cámara con movimiento de rotación.

El valor de $\mathbf{A}(z_{min})$ y $\mathbf{R}(z_{min})$ obtenidos se refina minimizando el error de proyección:

$$\min_{A(z_{min}), RD(z_{min}), M^j} \sum_{k=1}^m \sum_{i=1}^n D(m_i^j, A(RD(R_i I_0)))^2$$

2. Calibración de una secuencia con zoom.

La calibración de todo el rango de *zooms* se realiza construyendo un mosaico y repitiendo el proceso descrito en el paso 1 con múltiples pasos dentro del rango del zoom de la cámara.

Para ello se sitúa la cámara en una dirección fija y se adquiere una secuencia de imágenes aumentando el zoom de forma progresiva y sin rotar la cámara.

A continuación se estima las homografías, \mathbf{H}_{zoom} , entre cada par de imágenes dentro de la secuencia. Esto requiere un menor número de imágenes y el proceso es considerablemente más rápido que en el primer paso.

El siguiente paso es calcular la matriz de parámetros intrínsecos, $\mathbf{A}(z_i)$, y la distorsión radial, $RD(z_i)$, para cada nivel de zoom z_i registrado en cada paso de la secuencia. Esto se realiza recorriendo por parejas las imágenes de la secuencia y ajustando los valores para cada nivel z_i a partir de los valores de $\mathbf{A}(z_{i-1})$ y $RD(z_{i-1})$, calculados en el paso anterior, minimizando el error de proyección:

$$\min_{A(z_i), RD(z_i), M^j} \sum_{k=1}^m \sum_{i=1}^n D(m_i^j, A(RD(I_0)))^2$$

Los parámetros obtenidos se refinan minimizando el mismo error de proyección sobre todas las imágenes de la secuencia.

3.3. Conclusión.

En este capítulo se ha estudiado que los métodos de autocalibración ofrecen importantes ventajas sobre los métodos de calibración que se han estudiado en el capítulo 2 de la memoria. Las principales ventajas son que permiten calibrar los parámetros de la cámara sin la necesidad de conocer la posición de los puntos 3D de la escena (M) respecto a algún sistema de coordenadas del mundo; y que permiten una calibración *on-line* de la cámara.

Por otra parte, se ha comprobado que el estudio de los métodos de autocalibración es una línea de investigación que está muy abierta en la actualidad. Existen fuertes inconsistencias en los métodos dependiendo del contexto y del tipo de cámaras que se utilicen. Un ejemplo de estas inconsistencias es la imposibilidad de calibrar las cámaras PTZ mediante el método de Agapito si se desconoce el factor de escala.

En el capítulo 2 se concluyó que el método de Zhang es un método muy flexible y que, por tanto, resulta útil para calibrar los parámetros intrínsecos de la cámara. En este capítulo, se ha estudiado que es posible desarrollar un método de autocalibración de cámaras PTZ basándose en el método de Hartley y evitando los problemas que se han descrito en el método de Agapito. Mediante este método sería posible determinar los parámetros intrínsecos de la cámara, pero no los parámetros extrínsecos pues no se puede averiguar el factor de escala mediante un método de autocalibración con una cámara fija que sólo tienen movimiento de rotación. Por esta razón no se puede conocer la posición de una cámara PTZ respecto a un sistema de referencia (objetivo principal de este trabajo) mediante técnicas de autocalibración. Hay que señalar que actualmente existen estudios de métodos de autocalibración que tratan de conseguir los parámetros extrínsecos mediante un mínimo de 4 cámaras y conociendo los parámetros intrínsecos de éstas. Nuestro contexto es el de 3 cámaras PTZ por lo que no se puede avanzar en esta línea.

No obstante, el método descrito para autocalibrar cámaras PTZ y determinar los parámetros intrínsecos es todavía más flexible que el de Zhang, por lo que su implementación constituiría una línea de mejora del presente trabajo.

4. Calibración de varias cámaras PTZ que graban una misma escena.

4.1. Introducción.

En este capítulo se va a desarrollar el objetivo principal de este proyecto. El objetivo es desarrollar un método que permita conocer la posición de varias cámaras que graban una misma escena con respecto a un sistema de referencia común.

El contexto en el que se va a desarrollar el método es un sistema formado por tres cámaras PTZ que están ubicadas en el vestíbulo de la Escuela Politécnica Superior de la Universidad Autónoma de Madrid. Con el fin de desarrollar aplicaciones futuras que permitan la obtención de la posición 3D de los objetos que graban las tres cámaras, se va a desarrollar el método inverso: conocer la posición de las tres cámaras PTZ respecto a un sistema de referencia que graben las tres cámaras.

En el segundo capítulo se ha implementado el método de calibración de Zhang con el objetivo de calibrar los parámetros intrínsecos de las cámaras y se han hecho pruebas sobre cámaras fijas para evaluar la fiabilidad del método implementado. Las pruebas realizadas han permitido concluir que el método de Zhang es un método flexible y fiable para calibrar los parámetros intrínsecos de la cámara por lo que se puede emplear para obtener los de las cámaras PTZ.

El principal problema que presenta el método de Zhang a la hora de calibrar los parámetros extrínsecos de la cámara (es decir los que nos van a dar la posición de la cámara respecto del sistema de referencia elegido), es que necesita que el patrón de calibración ocupe casi toda la imagen por lo que, a priori, no parece un método práctico para aplicarlo en nuestro contexto (ya que el patrón ha de ser visible simultáneamente por todas las cámaras).

En el capítulo 3 se han estudiado las posibilidades de obtener los parámetros extrínsecos e intrínsecos de la cámara mediante técnicas de autocalibración. Se han analizado las inconsistencias que tienen estas técnicas en cámaras PTZ debido a sus particulares movimientos de rotación. Además se ha comprobado que es imposible determinar el factor de escala con un número tan reducido de cámaras. La idea inicial de este proyecto era desarrollar el objetivo principal mediante técnicas de autocalibración. Este estudio ha mostrado que no es posible determinar la posición de las cámaras respecto a un sistema de referencia mediante técnicas de autocalibración en este particular contexto.

Como solución se va a plantear una aplicación basada en el método de Zhang que permita aplicarse de forma práctica (sin necesidad de que el patrón de calibración ocupe la mayor parte de la imagen). Esta aplicación consta de dos pasos bien diferenciados:

1. Cálculo de los parámetros intrínsecos de la cámara mediante el método de calibración de Zhang.

2. Cálculo de los parámetros extrínsecos de la cámara en escenas en las que el patrón de calibración sea muy sencillo y ocupe una pequeña parte de la imagen.

Antes de desarrollar la aplicación, en la siguiente sección se van a presentar las especificaciones de la cámaras PTZ con las que se va a desarrollar el método.

4.2. Cámaras PTZ.

El modelo de las tres cámaras PTZ es la SONY SNC RZ50P. Estas tres cámaras están conectadas a una red IP.



Fig. 4.1. Cámara SONY SNC RZ50P.

4.2.1. Especificaciones de la cámara.

La cámara SONY SNC RZ50P es una cámara motorizada con *zoom* óptico integrado (26x) y función día/noche. Alcanza un nivel de 25 imágenes por segundo y tiene una resolución VGA de 640x480.

Los parámetros intrínsecos de la cámara varían según el nivel de *zoom* como se ha explicado en el tercer capítulo. A continuación se describen estos parámetros.

4.2.1.1. Factor de escala.

Como se ha explicado en el segundo capítulo, el factor de escala indica la proporción de tamaño de un objeto visto en la realidad respecto a su proyección en el plano imagen y depende de los siguientes parámetros:

$$k_u(z) = s \cdot d_u \cdot f(z) \quad k_v(z) = d_v \cdot f(z)$$

Donde s es el factor de proporción, f es la distancia focal y (d_u, d_v) son los factores de conversión píxel-milímetros. Se expresa el factor de escala y la distancia focal como función de z para especificar que dependen del nivel de *zoom*.

El dispositivo sensor es un CCD en color de tipo ¼. El tamaño del píxel del píxel es de $3,85\mu m(H) \cdot 3,5\mu m(V)$.

De este dato se deduce que el factor de proporción s es:

$$s = \frac{3,5}{3,85} = 0,90909$$

Y que los factores de conversión son:

$$d_u = \frac{1}{0.00385} = 259,74 \left[\frac{\text{píxeles}}{\text{mm.}} \right]$$

$$d_v = \frac{1}{0,0035} = 285,71 \left[\frac{\text{píxeles}}{\text{mm.}} \right]$$

La distancia focal depende del nivel de *zoom* óptico utilizado. Por tanto, los factores de escala también. Los valores que toma son los siguientes:

Zoom óptico	Valor de la distancia focal f [mm.]	Valor de los factores de escala (k_u, k_v).
x1	3,5	843,36
x2	7	1.686,72
x3	10,5	2.530,08
x4	14	3.373,44
x5	17,5	4.216,8
x6	21	5.060,16
x7	24,5	5.903,52
x8	28	6.746,88
x9	31,5	7.590,24
x10	35	8.433,6
x11	38,5	9.276,96
x12	42	10.120,32
x13	45,5	10.963,68
x14	49	11.807,04
x15	52,5	12.650,4
x16	56	13.493,76
x17	59,5	14.337,12
x18	63	15.180,48
x19	66,5	16.023,84
x20	70	16.867,2
x21	73,5	17.710,56
x22	77	18.553,92
x23	80,5	19.397,28
x24	84	20.240,64
x25	87,5	21.084
x26	91	21.927,36

Tabla 4.1. Valores de la distancia focal y el factor de escala para cada nivel de zoom.

4.2.1.2. Punto principal de las cámaras PTZ.

El punto principal de la cámara (punto donde el eje óptico corta el plano principal de la imagen) también depende del nivel del *zoom*. Cuando el nivel de *zoom* es el mínimo (x_1), en la fabricación se intenta que el punto principal sea el centro de la imagen.

En realidad, el punto principal no suele ser el centro exacto de la imagen, y cada cámara fabricada tiene un punto principal diferente.

Cuando el nivel de *zoom* aumenta, el punto principal varía según el nivel aumentado. El parámetro de la cámara “*Area Zoom*” define la posición del punto principal. Los argumentos de este parámetro son “*AreaZoom = x,y,w,h*”, siendo x , y , w , h los valores del siguiente dibujo:

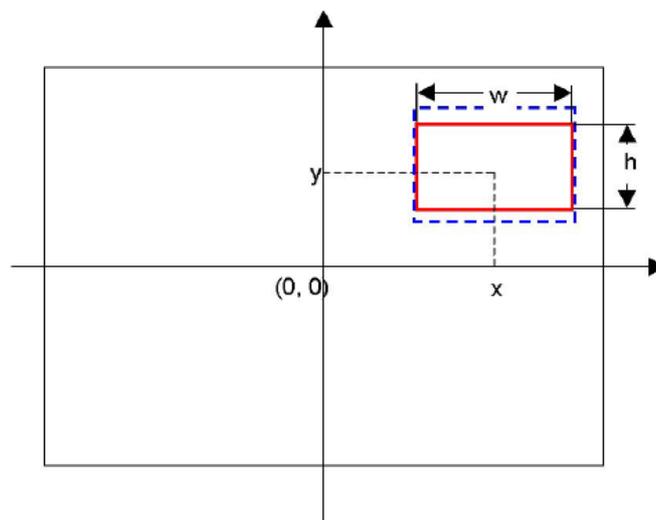


Fig. 4.2. Plano imagen. Valores del parámetro “*AreaZoom*”.

Es decir los valores “ x ” e “ y ” definen la posición del punto principal de la cámara con respecto al centro de la imagen para cada distinto nivel de *zoom*. Cuando los valores de “ w ” y “ h ” son cero, el nivel de *zoom* es el mínimo. Por lo tanto, es necesario calibrar el punto principal para cada nivel de *zoom* empleado.

4.2.1.3. Movimiento pan de la cámara.

Los parámetros extrínsecos de la cámara dependen de la posición del sistema de referencia del mundo, de la posición de la cámara y del movimiento *pan* y *tilt* que tenga la cámara en cada instante.

Como la posición de las cámaras es fija, y la posición del sistema de referencia del mundo que se va a considerar es fija también, el vector de translación va a permanecer constante con independencia del movimiento *pan-tilt* que tenga la cámara.

La matriz de rotación, sin embargo, sí va a variar conforme varíen los movimientos *pan* y *tilt* de la cámara. A continuación se describe el rango dinámico de estos movimientos.

La rotación sobre el eje horizontal de la cámara puede modificar el ángulo en un rango de 340 grados según indica la figura 4.3:

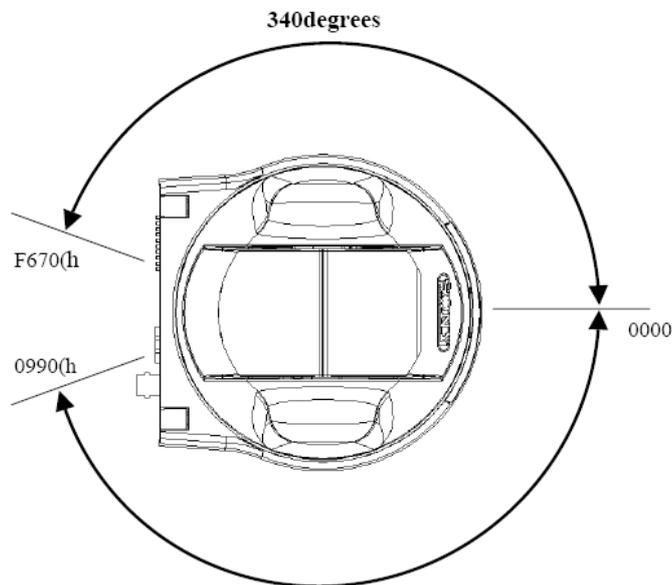


Fig. 4.3. Vista superior de la cámara. Rango de los ángulos horizontales que puede tomar la cámara.

La posición *pan* de la cámara puede ser escrita o leída mediante el parámetro *AbsolutePanTilt*. Este parámetro está formado por tres argumentos, el primero de los argumentos indica la posición *pan* de la cámara. Este argumento se representa con un valor hexadecimal. Como se puede ver en la figura 4.3, el valor 0000 corresponde a la posición de referencia de la cámara, el valor F670h corresponde a una rotación horizontal de 170° hacia la derecha, y el valor 0990h corresponde a una rotación horizontal de 170° hacia la izquierda.

4.2.1.4. Movimiento *tilt* de la cámara.

La rotación sobre el eje vertical de la cámara puede modificar el ángulo en un rango de 115 grados (25 hacia arriba y 90 hacia abajo) según indica la figura 4.4.

La posición *tilt* de la cámara también puede ser escrita o leída mediante el parámetro *AbsolutePanTilt*. El segundo de los argumentos de este parámetro indica la posición *tilt* de la cámara. Este argumento se representa con un valor hexadecimal. Como se puede ver en la figura 4.4, el valor 0000 corresponde a la posición de referencia de la cámara (que no es la posición correspondiente al eje óptico en posición horizontal), el valor 033Ch corresponde a una rotación vertical de 57,5° hacia arriba, y el valor FCC4h corresponde a una rotación vertical de 57,5° hacia abajo.

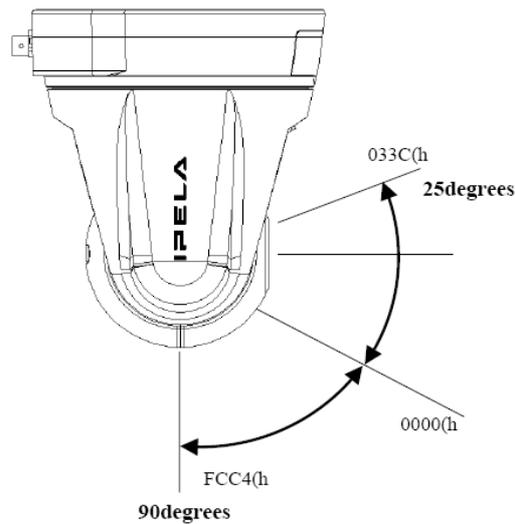


Fig. 4.4. Vista lateral de la cámara. Rango de los ángulos horizontales que puede tomar la cámara.

4.2.2. Posición de las tres cámaras.

Las tres cámaras PTZ se encuentran distribuidas en el vestíbulo de entrada de la EPS de la UAM.

Como centro de coordenadas del mundo se va tomar el centro de una DIN-A4 colocada en la entrada de la EPS y que se encuentra en la escena que graban las tres cámaras.

Los parámetros extrínsecos de las cámaras estarán referidos respecto de este centro de coordenadas.

Como las cámaras son fijas (están sujetas al techo y sólo tienen los dos movimientos de rotación pura *pan* y *tilt*) y el centro de coordenadas del sistema de referencia del mundo también es fijo, los tres parámetros del vector de translación $(t_x \ t_y \ t_z)$ también son fijos. Los parámetros extrínsecos que pueden variar en todo momento son los de la matriz de rotación, y lo harán conforme al movimiento *pan-tilt* de la cámara.



Fig. 4.4bis. Vestíbulo donde están las cámaras.

4.3. Cálculo de los parámetros intrínsecos de las cámaras mediante el método de calibración de Zhang.

En este apartado se va a calibrar una de las tres cámaras PTZ mediante el método de Zhang. Como las tres cámaras son iguales, los parámetros intrínsecos calculados serán los mismos para las tres cámaras.

El procedimiento utilizado para la calibración es el mismo que se explicó en el capítulo 2. Como los parámetros intrínsecos varían dependiendo del nivel de *zoom* (ver apartado 4.2.1.), será necesario calibrar la cámara con distintos niveles de *zoom*, obteniendo resultados diferentes para cada nivel.

4.3.1. Calibración con el nivel más bajo de *zoom*.

Como se puede ver en la tabla 4.1., la distancia focal correspondiente al nivel más bajo de *zoom* es $f = 3,5 \text{ mm}$. En el capítulo 2 se estudió que la distorsión radial era mayor cuanto menor era la distancia focal por lo que en este nivel de *zoom* se tendrá la mayor distorsión.

Para la calibración se han considerado 5 imágenes del patrón y 9 puntos por cada imagen y se han seguido los criterios explicados en el capítulo 2.4. a la hora de obtener las imágenes. En la Fig. 4.5. se muestran las 5 imágenes que se han utilizado.



Fig. 4.5. Imágenes utilizadas para la calibración con zoom x1.

En la tabla 4.2 se muestran los valores de los parámetros intrínsecos de las tres cámaras obtenidos en comparación a los valores reales. Nos referimos a cada cámara para distinguirlas con el último número de sus dirección IP (cámara 100, cámara 101 y cámara 102).

	Parámetros	k_u	k_v	u_0	v_0
	Datos reales	843	843	320	240
Cámara 100	Resultados	840	869	322	244
	Error	0,3 %	3 %	0,6 %	1 %
Cámara 101	Resultados	840	870	330	247
	Error	0,3 %	3 %	1,2 %	1,2 %
Cámara 102	Resultados	840	870	330	245
	Error	0,3 %	3 %	1,2 %	1,2 %

Tabla 4.2. Valores obtenidos con nivel de *zoom* x1 ($f = 3,5 \text{ mm.}$)

Se puede apreciar que el error cometido es similar al que se ha obtenido en las pruebas sobre cámaras fijas del capítulo 2. Un aspecto importante de los resultados es comprobar que el punto principal de la imagen depende de la fabricación de cada cámara y que no coincide exactamente con el centro de la imagen.

4.3.2. Calibración con distintos niveles de *zoom*.

Actuando de la misma forma que en el apartado anterior se han calibrado los parámetros intrínsecos con los siguientes niveles de *zoom*: x2, x4 y x8. A partir de este valor del *zoom* la calibración no se puede realizar con precisión debido al enfoque de las cámaras que provocan la obtención de imágenes muy distorsionadas.

El nivel de *zoom* de la cámara se programa enviándole por IP el siguiente comando:

http://192.168.1.100//command/ptzf.cgi?AbsoluteZoom=CÓDIGO_ZOOM

donde CÓDIGO_ZOOM toma el valor que corresponda al *zoom* deseado según la tabla 4.3 extraída del manual “SNC-RZ50N/P CGI command manual” de SONY:

Zoom Ratio	Optical Zoom Position Data
x1	0000
x2	1760
x3	214C
x4	2722
x5	2B22
x6	2E20
x7	3080
x8	3278
x9	3426
x10	359E
x11	36EE
x12	381C
x13	392E
x14	3A26
x15	3B08
x16	3BD4
x17	3C8C
x18	3D2E
x19	3DBC
x20	3E58
x21	3EA2
x22	3F00
x23	3F4E
x24	3F92
x25	3FCC
x26	4000

Tabla 4.3. Código asociado a cada *zoom*.

Así el comando

<http://192.168.1.100//command/ptzf.cgi?AbsoluteZoom=1760>

fuerza que el *zoom* de la cámara sea x2.

En las siguientes tablas se recogen los valores de los parámetros intrínsecos obtenidos para los distintos niveles de *zoom* que se han considerado. Estos valores son comparados con los valores reales. La comparación no se realiza con respecto al punto principal, pues ya se ha indicado que al aumentar el *zoom*, éste se desplaza del centro de la cámara y se desconoce su valor. Si se constata que el desplazamiento producido es proporcional en ambos ejes.

	k_u	k_v	u_0	v_0
Datos reales	1.686	1.686	-	-
Resultados	1.640	1.700	361	294
Error cometido	0,3 %	3 %	-	-

Tabla 4.4. Valores obtenidos con nivel de *zoom* x2 ($f = 7 \text{ mm.}$)

	k_u	k_v	u_0	v_0
Datos reales	2.350	2.360	-	-
Resultados	2.400	2.400	361	294
Error cometido	2 %	1,6 %	-	-

Tabla 4.5. Valores obtenidos con nivel de *zoom* x3 ($f = 10,5 \text{ mm.}$)

	k_u	k_v	u_0	v_0
Datos reales	3.372	3.372	-	-
Resultados	3.369	3.390	361	294
Error cometido	0,01 %	0,05%	-	-

Tabla 4.6. Valores obtenidos con nivel de *zoom* x4 ($f = 14 \text{ mm.}$)

Se observa que el error cometido al obtener los resultados es similar al producido en el capítulo 2 cuando se probó el método de Zhang con cámaras fijas. La comparación de los datos obtenidos con los datos reales sólo puede realizarse respecto a los factores de escala ya que el valor del punto principal varía con el nivel de *zoom*, y este dato no figura en las especificaciones.

4.4. Cálculo de los parámetros extrínsecos de las cámaras conocida la matriz de parámetros intrínsecos.

4.4.1. Introducción.

En el capítulo 2 se estudió con detalle el desarrollo del método de Zhang y su implementación. En dicho capítulo se indicó que teóricamente eran necesarias al menos tres imágenes (en la práctica eran más) para calibrar la cámara.

Sin embargo esta restricción sólo afecta a la hora de calcular los parámetros intrínsecos de la cámara. Como se verá a continuación, para el cálculo de los parámetros extrínsecos sólo es necesario disponer de una imagen j , si se conocen los parámetros intrínsecos de la cámara y la matriz de proyección, \mathbf{P}_j , de un determinado número de puntos de la imagen. De hecho, el método de Zhang calcula una matriz de parámetros extrínsecos para cada imagen que se utiliza para la calibración (pues esta matriz varía para cada imagen al mover el patrón de calibración).

La idea de este apartado consiste en tomar como patrón un elemento plano de dimensiones conocidas, calcular la matriz de proyección \mathbf{P}_j de ciertos puntos del patrón en una imagen siguiendo el método de Zhang y calcular la matriz de parámetros extrínsecos a partir de la matriz de proyección, \mathbf{P}_j , y la matriz de parámetros intrínsecos, \mathbf{A} (que es conocida de antemano).

Si se escoge un patrón que permita obtener los puntos de forma automática, será posible conseguir la posición de las cámaras con respecto a un sistema de referencia (donde se colocará el patrón), también de forma automática.

Como ya se comprobó en el capítulo 2 para que el método se pueda resolver numéricamente es necesario que el patrón de calibración ocupe casi toda la imagen. De lo contrario, si los puntos considerados están muy cerca la relación señal a ruido que se introduce es muy baja y los resultados tienen un error muy grande.

Utilizar un patrón de calibración que ocupe casi toda la imagen es inviable para la mayor parte de las situaciones (recuérdese que el patrón ha de ser visible simultáneamente por todas las cámaras) de la aplicación que se desea desarrollar, ya que se pretende que las tres cámaras estén grabando el mismo patrón.

La solución que se propone consiste en considerar varias imágenes de la misma escena grabadas con distinto *zoom*. En este caso la matriz de parámetros extrínsecos no varía para ninguna de las imágenes. No obstante, exige conocer la matriz de parámetros intrínsecos \mathbf{A} para cada nivel de *zoom*.

Si aplicamos el método en cada una de las imágenes, obtendremos tantas homografías como imágenes. Cada una de las imágenes tendrá su propio ruido, pero los ruidos de cada imagen son linealmente independientes, por lo que si promediamos los valores de cada imagen el resultado tenderá a converger al valor real (teorema del límite central).

Como la cámara es PTZ (está fija) una solución alternativa consiste en aplicar este mismo concepto a varias imágenes rotando la cámara para obtener un mayor número de homografías. En este caso no se podrá obtener la matriz de rotación (ya que ésta varía al

variar la cámara) pero sí se podrá obtener el vector de translación (que permanece constante), y bastará con disponer de \mathbf{A} para un solo nivel de *zoom*.

El objetivo de este apartado es evaluar si es posible obtener resultados convergentes para un número razonable de imágenes.

4.4.2. Desarrollo del método.

Como ya se ha explicado en el capítulo 2, el cálculo de la matriz de proyección, \mathbf{P} , es análogo al del apartado 2.2.1., pero eliminando la coordenada z .

De esta forma el sistema de ecuaciones que se obtiene es (ecuación 2.18):

$$\begin{bmatrix} X & Y & 1 & 0 & 0 & 0 & -u \cdot X & -u \cdot Y & -u \\ 0 & 0 & 0 & X & Y & 1 & -v \cdot X & -v \cdot Y & -v \end{bmatrix} \cdot \mathbf{p} = 0 \quad (4.17)$$

donde \mathbf{p} es una matriz columna que contiene los 9 elementos de la matriz \mathbf{P} :

$$\mathbf{p} = [p_{11} \ p_{12} \ p_{13} \ p_{21} \ p_{22} \ p_{23} \ p_{31} \ p_{32} \ p_{33}]^T \quad (4.18)$$

Al igual que en el apartado 2.2.2., se consideran n puntos de la misma imagen para establecer $2n$ ecuaciones. Este sistema se resuelve hallando el valor singular más pequeño de la SVD de la matriz del sistema, y utilizándolo como punto de partida del método no lineal que busca minimizar la diferencia entre el punto medido y el punto proyectado.

Una vez obtenida la matriz de proyección, \mathbf{P}_j , se calculan los parámetros extrínsecos, despejándolos de la homografía del modelo de cámara de Zhang (capítulo 2.2.2, ecuación 2.15):

$$\lambda \cdot \mathbf{P}_j = \mathbf{A} \cdot [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}] \quad (4.19)$$

donde λ es el factor de escala, \mathbf{A} la matriz de parámetros intrínsecos y $[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}]$ la matriz de parámetros extrínsecos.

Para despejar la matriz de parámetros extrínsecos se pasa la matriz de parámetros intrínsecos al otro lado de la ecuación (4.19):

$$\lambda \cdot \mathbf{A}^{-1} \cdot \mathbf{P}_j = [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}] \quad (4.20)$$

Si se expresa la matriz de proyección \mathbf{P}_j en función de sus columnas \mathbf{P}_j^i , la ecuación (4.20) resulta de la siguiente forma, donde $\mathbf{A}^{-1} \cdot \mathbf{P}_j^i$ es la columna 'i' de la matriz resultante.

$$\lambda \cdot [\mathbf{A}^{-1} \cdot \mathbf{P}_j^1 \ \mathbf{A}^{-1} \cdot \mathbf{P}_j^2 \ \mathbf{A}^{-1} \cdot \mathbf{P}_j^3] = [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}] \quad (4.21)$$

En esta ecuación son conocidas las matrices \mathbf{A}^{-1} y \mathbf{P}_j , pero no el factor de escala λ . Para determinarlo, se tiene en cuenta que la norma de \mathbf{r}_1 es la unidad (ya que \mathbf{r} es la matriz de rotación).

$$\|\lambda \cdot \mathbf{A}^{-1} \cdot \mathbf{P}_j^1\| = \|\mathbf{r}_1\| = 1 \quad (4.22)$$

Se despeja $|\lambda|$ de la ecuación (4.22).

$$|\lambda| = \frac{1}{\|\mathbf{A}^{-1} \cdot \mathbf{P}_j^1\|} \quad (4.23)$$

como ya se ha indicado, el problema que presenta el método de Zhang es que precisa que los puntos considerados en la imagen estén bastante alejados entre sí para que no haya una relación señal a ruido muy baja en la homografía calculada. El último paso consiste en promediar los valores de $\lambda_j \cdot \mathbf{A}^{-1} \cdot \mathbf{P}_j$ obtenidos de cada imagen j ($j = 1, \dots, N$) que se utilice, para disminuir esta relación señal a ruido. De esta forma los valores que se obtendrán de los parámetros extrínsecos serán más exactos.

$$\lambda \cdot \mathbf{A}^{-1} \cdot \mathbf{P} = \sum_{j=1}^N \frac{\lambda_j \cdot \mathbf{A}^{-1} \cdot \mathbf{P}_j}{N} \quad (4.24)$$

Una vez promediados los valores obtenidos, se pueden calcular los parámetros extrínsecos a partir de la ecuación (4.21). En el caso de haber utilizado imágenes con la misma posición y variando únicamente el *zoom* (lo que exige matrices \mathbf{A} distintas para cada nivel de *zoom*), todos los valores son válidos. Si además se han utilizado imágenes obtenidas al rotar la cámara, únicamente es válido el valor del vector de translación.

$$\mathbf{r}_1 = \frac{\mathbf{A}^{-1} \cdot \mathbf{P}^1}{\|\mathbf{A}^{-1} \cdot \mathbf{P}^1\|} \quad (4.25)$$

$$\mathbf{r}_2 = \frac{\mathbf{A}^{-1} \cdot \mathbf{P}^2}{\|\mathbf{A}^{-1} \cdot \mathbf{P}^1\|} \quad (4.26)$$

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \quad (4.27)$$

$$\mathbf{t} = \frac{\mathbf{A}^{-1} \cdot \mathbf{P}^3}{\|\mathbf{A}^{-1} \cdot \mathbf{P}^1\|} \quad (4.28)$$

4.4.3. Implementación del método.

Este método se ha implementado en código Matlab al igual que todos los demás.

Como patrón de calibración se ha utilizado una DINA-4, y se han considerado 9 puntos del patrón. El centro de coordenadas del sistema de referencia es el centro de la DINA-4. Los puntos considerados se muestran en la siguiente imagen.

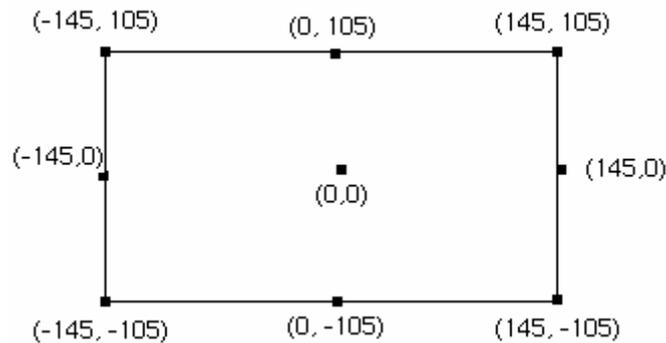


Fig. 4.6. Puntos considerados de la DINA-4.

La ventaja de considerar este patrón es que permite encontrar y calcular los puntos de forma automática. Basta con indicar cuál es la referencia (el DINA-4) para hallar la posición de las cámaras con respecto a esta referencia.

La forma de hallar los puntos es la siguiente. El usuario sólo tiene que marcar un punto de la DINA-4 y mediante un algoritmo de detección de regiones implementado en Matlab (*growRegion.m*) se calculan los píxeles que ocupan la DINA-4. Una vez conocida la región, es sencillo calcular la posición de las esquinas y de los puntos medios de los lados.

Esta forma de hallar los puntos se puede mejorar considerando un detector de puntos que encuentre las esquinas de la DINA-4. De esta forma, ni siquiera haría falta que el usuario marcar un punto de la región.

4.4.4. Pruebas sobre el método y conclusiones.

El método puede tener tres variantes distintas a la hora de elegir las imágenes que se van a utilizar.

1. Considerar distintas imágenes variando la rotación de la cámara (barrido). De esta forma sólo se puede calcular el vector de translación, ya que la matriz de rotación varía de una imagen a otra.
2. Considerar distintas imágenes variando el *zoom* de la cámara. De esta forma se pueden calcular todos los parámetros extrínsecos (tanto la matriz de rotación como el vector de translación) ya que no varían de una imagen a otra.
3. Combinar imágenes variando el *zoom* y la rotación de la cámara.

Con el único fin de comprobar la validez del método propuesto se va a calcular únicamente el vector de translación (ya que es el único parámetro extrínseco que se puede medir en la escena). Por lo tanto se van a considerar distintas imágenes variando la rotación de la cámara (barrido) por una mayor comodidad

Si se deseara calcular también la matriz de rotación, sólo se podrían considerar imágenes en las que se ha variado el *zoom* (y utilizar una matriz de parámetros intrínsecos para cada una).

No obstante, la matriz de rotación puede ser calculada una sola vez con una posición *pan-tilt* de referencia. Para determinar la matriz de rotación que se obtiene con otra posición de la cámara, basta con recoger los valores *pan* y *tilt* del parámetro *AbsolutePanTilt*, y calcular la matriz de rotación con respecto a la que se ha calculado como referencia.

En este apartado se va a analizar cuantas imágenes es necesario utilizar para que los valores obtenidos del vector de translación converjan en los valores reales. La viabilidad del método dependerá de la cantidad de imágenes que se necesite.

4.4.4.1. Secuencia de imágenes grabadas con un barrido pan.

Se toma una serie de imágenes de la DIN-A4 mediante un barrido pan de la cámara. Para forzar una posición determinada de la cámara se le envía el siguiente comando por IP:

`http://192.168.1.100//command/ptzf.cgi? AbsolutePanTilt=F93F,0138,24`

Para realizar el barrido pan es necesario ir cambiando el argumento “pan” (el primero) de la instrucción *AbsolutePanTilt*. El ángulo con el que se rota la cámara se puede conocer gracias a la relación entre el valor hexadecimal del argumento “pan” y su correspondencia en grados.

Esta correspondencia se puede extraer de la Fig. 4.3. en la que se muestra que el valor “F670h” corresponde con un ángulo de 170°. En este caso, cada imagen ha sido tomada rotando la cámara 0,054° con respecto a su posición anterior.

La figura 4.7 muestra la secuencia de imágenes grabadas mediante el barrido pan. Se aprecia que el patrón utilizado tiene un tamaño bastante reducido.

La distancia (vector de translación \mathbf{t}) entre la cámara y el centro del patrón (origen del sistema de referencia) es de 3,5 m.

El conjunto de imágenes que se ha tomado es mayor que 12 pero se han utilizado distintos subgrupos para validar los resultados. El método implementado se ha ejecutado para un número progresivo de imágenes con el fin de estudiar como disminuye el error cometido al medir el vector de translación a medida que aumenta el número de imágenes utilizadas.

Como ya se ha indicado, únicamente es necesario marcar un punto cualquiera de la DIN-A4 en cada imagen para indicar cuál es el sistema de referencia.

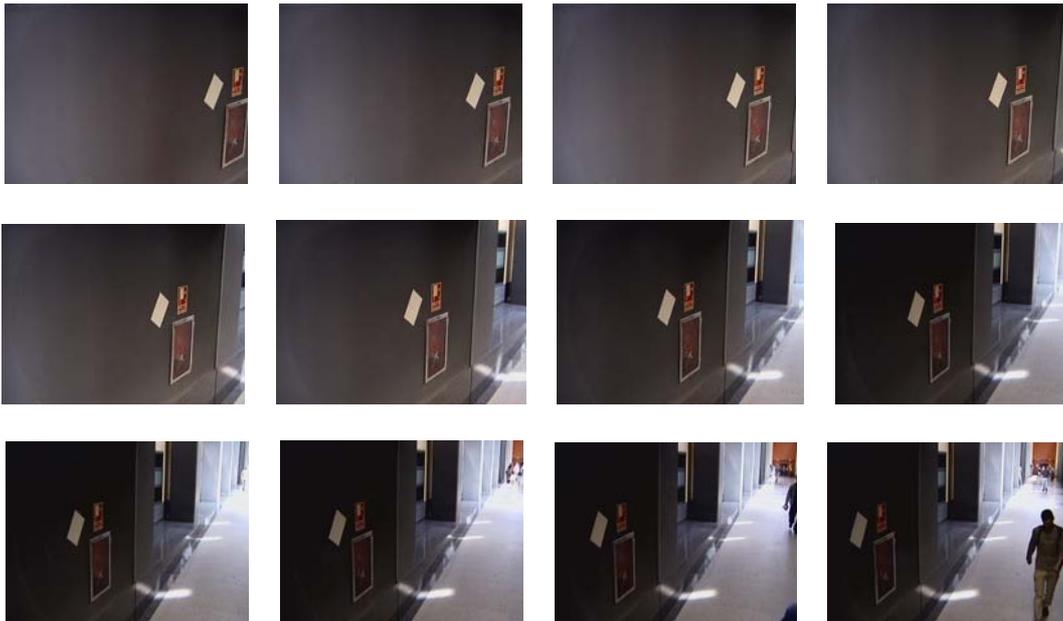


Fig. 4.7. Secuencia de imágenes grabadas mediante un barrido pan.

La siguiente gráfica muestra el valor del vector T obtenido para cada número de imágenes utilizadas.

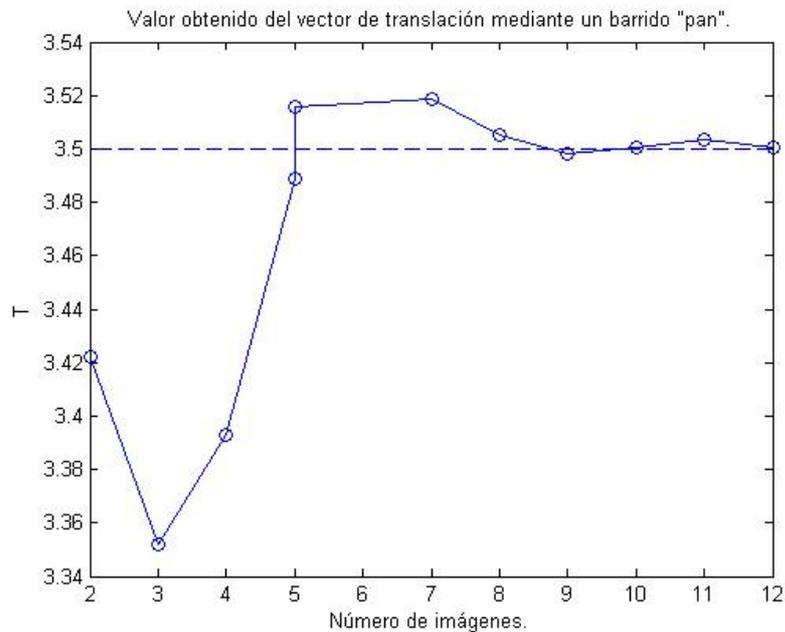


Fig. 4.8. Valor obtenido del vector de translación t (m)

Como se puede apreciar en la gráfica, el error cometido si se considera un número muy pequeño de imágenes es relativamente alto. Este error llega a valer 15 cm. ($\pm 4\%$). Como se ha indicado en la introducción de este apartado, el error que se comete es debido a que los puntos del patrón están muy cerca entre sí, haciendo que la relación señal a ruido no sea muy alta.

También se puede ver que a medida que aumenta el número de imágenes, el valor de T tiende a converger rápidamente a su valor real. Así entre 5 y 8 imágenes el error que se comete es de 2 cm. ($\pm 0,5\%$); y a partir de 9 imágenes este error no supera el medio centímetro ($\pm 0,14 \%$). La razón por la que el resultado obtenido tiende a converger a su valor real es que el ruido introducido por cada imagen es linealmente independiente, por lo que si se promedian los valores obtenidos en cada imagen, el resultado tenderá a converger.

En conclusión, es posible obtener la distancia a la que la cámara se encuentra del patrón de referencia con la precisión que se necesite para cada aplicación. Únicamente es necesario aumentar el número de imágenes si se requiere una precisión mayor.

4.4.4.2. Secuencia de imágenes grabadas con un barrido tilt.

Con el fin de validar los datos obtenidos en el apartado 4.4.4.1. se ha realizado la misma prueba con una nueva secuencia de imágenes. Esta vez tomadas mediante un barrido tilt.

Para realizar el barrido *tilt* es necesario ir cambiando el argumento “*tilt*” (el segundo) de la instrucción *AbsolutePanTilt*. El ángulo con el que se rota la cámara se puede conocer gracias a la relación entre el valor hexadecimal del argumento “*tilt*” y su correspondencia en grados.

Esta correspondencia se puede extraer de la Fig. 4.4. en la que se muestra que el valor “678h” corresponde con un ángulo de 115° . En este caso, cada imagen ha sido tomada rotando la cámara $1,11^\circ$ con respecto a su posición anterior.

La siguiente figura muestra la secuencia de imágenes grabadas mediante el barrido *tilt*. De nuevo, el tamaño del patrón es bastante reducido en relación al tamaño de la imagen.

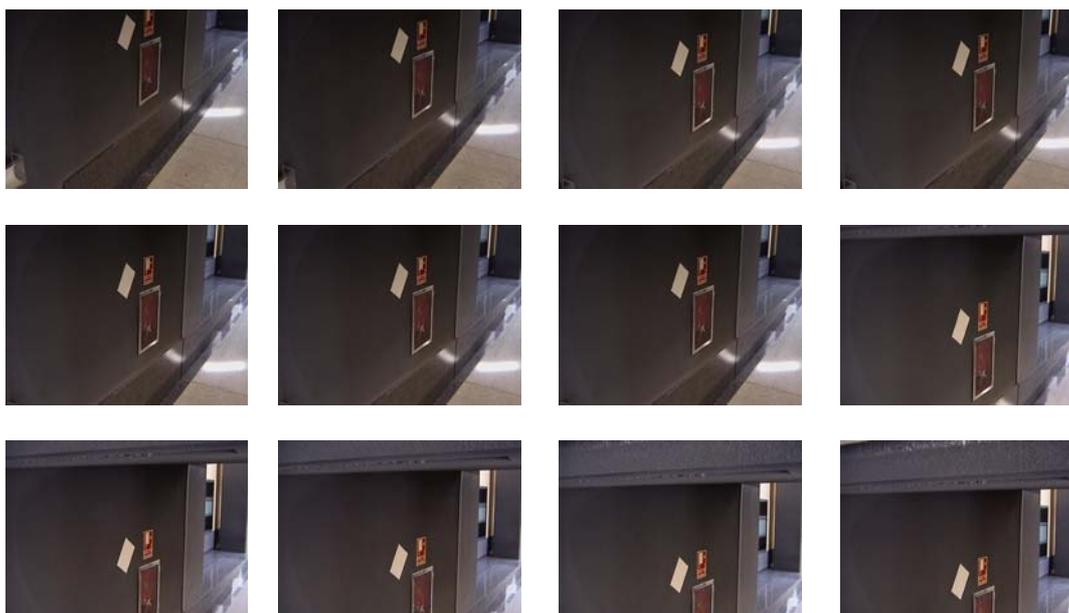


Fig. 4.9. Secuencia de imágenes grabadas mediante un barrido *tilt*.

De nuevo, se ha ejecutado el método implementado para un número progresivo de imágenes con el fin de estudiar como disminuye el error cometido al medir el vector de translación a medida que aumenta el número de imágenes utilizadas. Además se va a verificar si los resultados coinciden con los de un barrido “*pan*”.

La siguiente gráfica muestra el valor del vector t obtenido para cada número de imágenes utilizadas.

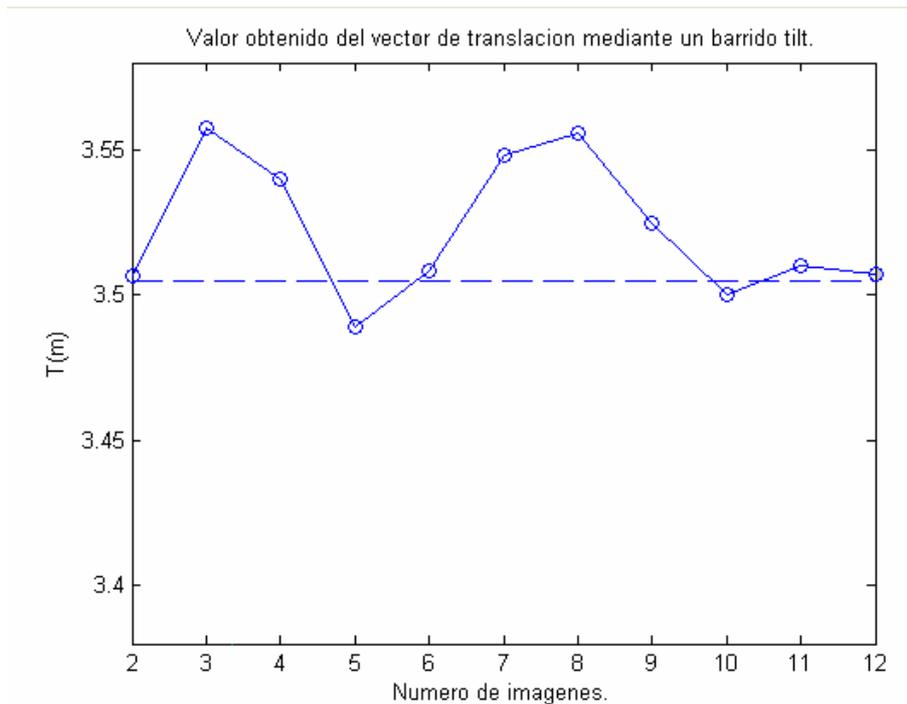


Fig. 4.10. Valor obtenido del vector de translación t (m).

Como se puede apreciar en la gráfica, el valor al que tiende a converger el vector de translación es el mismo que el del barrido “*tilt*”.

También se puede ver que el error tiende a disminuir considerablemente a partir de las 10 imágenes como ocurría con las imágenes del apartado anterior.

Mediante este apartado se pueden validar los resultados obtenidos en el apartado anterior y extraer las mismas conclusiones, es decir, es posible conocer la distancia a la que la cámara se encuentra del sistema de referencia con la precisión que se necesite.

4.4.4.3. Análisis del error de cuantificación.

En los apartados anteriores se ha estudiado como se puede reducir el error cometido al calcular la matriz de proyección debido a la cercanía entre sí de los puntos de calibración. Sin embargo, existe un error de cuantificación que no puede ser reducido.

Este error es consecuencia de que la máxima precisión subpíxel que se puede obtener es de 0,5 píxeles.

Pero sí es posible determinar el error de cuantificación aprovechando el patrón de la escena. Como las medidas del patrón son conocidas, basta medir la distancia en mm. que representa un píxel de la imagen para hallar este error.

En este apartado se ha medido el error de cuantificación cometido en los apartados anteriores. El resultado obtenido es el siguiente:

$$error_q = \pm 3,7mm. = 0,105\%$$

En conclusión, aunque se trate de promediar el máximo número de imágenes, existirá siempre una incertidumbre cercana al 0,1% en los resultados obtenidos. Este error de cuantificación depende directamente de la distancia a la que la cámara se encuentra del patrón. Cuanto más alejado esté, mayor error se comete. También depende inversamente de la resolución de la imagen.

4.5. Desarrollo de la aplicación.

Utilizando el método desarrollado en el apartado anterior, en este capítulo se va a demostrar una aplicación en la que las tres cámaras PTZ graban una misma escena. El objetivo de la aplicación es determinar la distancia a la que están las tres cámaras PTZ de un punto de referencia común.

La siguiente figura muestra el esquema de una vista aérea del vestíbulo de la EPS donde están las tres cámaras.

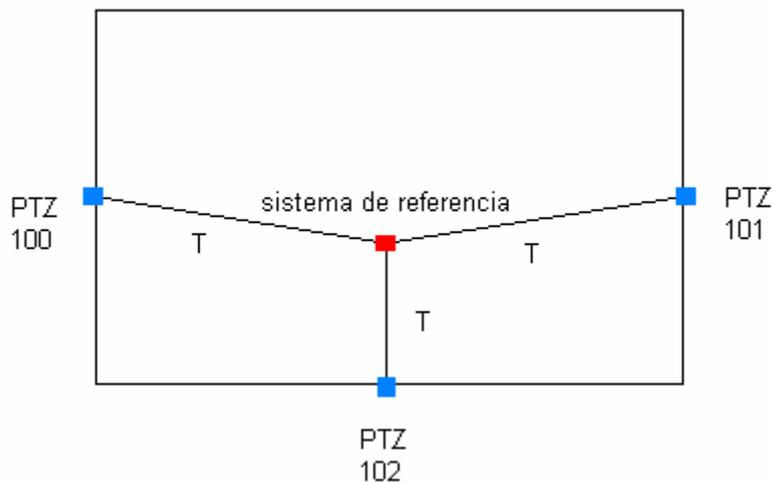


Fig. 4. 11. Esquema de la posición de las tres cámaras respecto al sistema de referencia.

Como se puede apreciar en la imagen, las cámaras PTZ 100 y 101 están a una distancia similar del sistema de referencia. Esto servirá para validar los resultados de la aplicación.

El método para averiguar la distancia a la que está cada cámara sigue los siguientes pasos:

1. Capturar una secuencia de imágenes del patrón mediante un barrido pan, un barrido zoom, un barrido tilt o una combinación de todos.
2. Registrar la información sobre el zoom que ha utilizado la cámara para utilizar la matriz de parámetros intrínsecos adecuada.
3. Aplicar el método descrito en el apartado anterior sobre las imágenes grabadas.

En esta aplicación se van a grabar todas las imágenes con el nivel más bajo de zoom y efectuando rotaciones de la cámara. Por lo tanto, únicamente se va a medir el vector de translación t .

4.5.1. Imágenes capturadas por las cámaras.

Las imágenes de la cámara PTZ 100 han sido capturadas mediante un barrido pan de la escena. Debido a la distancia a la que se encontraba la DINA-4 ha sido necesario coger 2 folios como referencia. Aún así, el patrón ocupa un tamaño bastante pequeño en relación a la imagen.

La siguiente figura muestra las 18 imágenes tomadas mediante el barrido pan para ser utilizadas en la calibración.

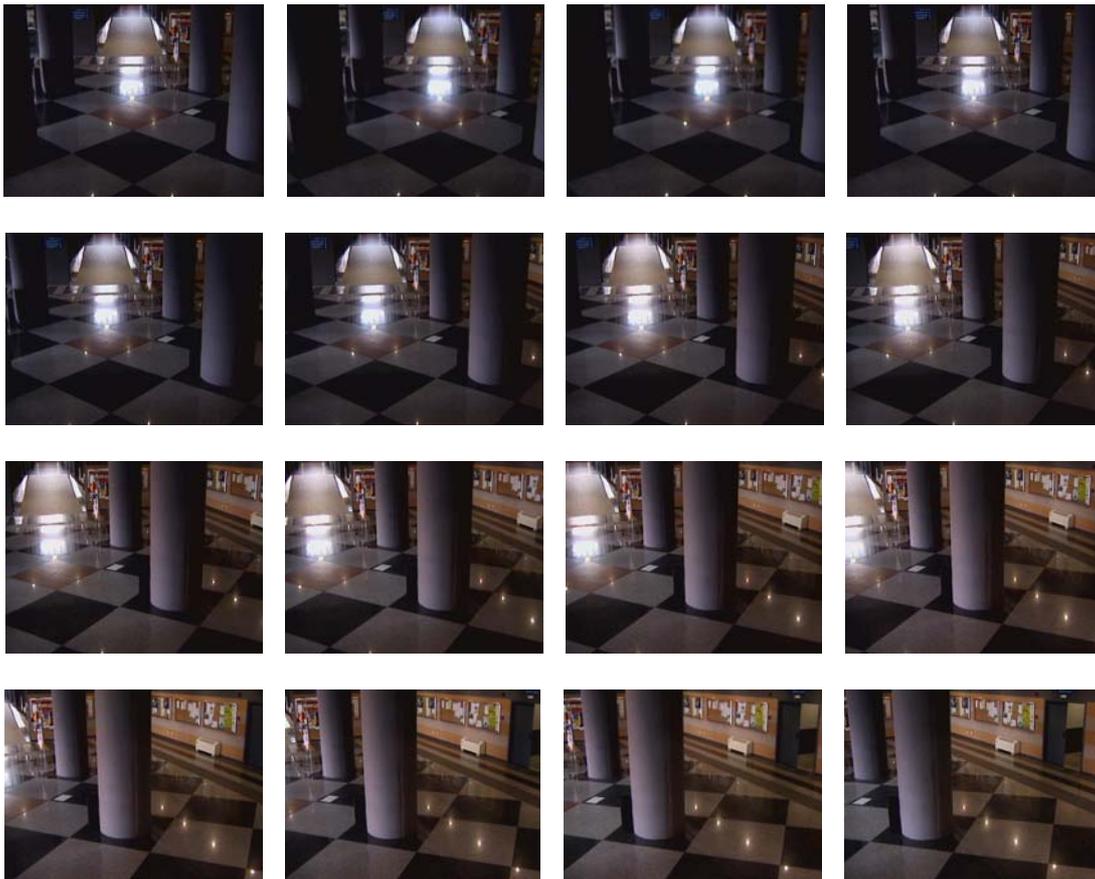


Fig. 4.12. Imágenes de la escena grabadas por PTZ 100 (barrido pan).

Se puede apreciar que el tamaño que ocupan los folios en la imagen es muy pequeño. El nivel de zoom con el que se han grabado las imágenes es el más bajo por lo que es necesario utilizar la matriz de parámetros intrínsecos correspondiente. También hay que tener en cuenta que las dimensiones del patrón son 420 x 290 en lugar de 290 x 210.

Con estas imágenes y estos datos se ejecuta el método explicado en el apartado 4.4 para hallar los parámetros extrínsecos de la cámara., el resultado obtenido del vector de translación es el siguiente:

$$\mathbf{T}_{100} = 10,51 \text{ m.}$$

Las imágenes de la cámara PTZ 101 también han sido capturadas mediante un barrido *pan* de la escena. Debido a la distancia a la que se encontraba la DIN-A4 de nuevo ha sido necesario coger 2 folios como referencia.

La siguiente figura muestra las 16 imágenes tomadas mediante el barrido *pan* para ser utilizadas en la calibración.

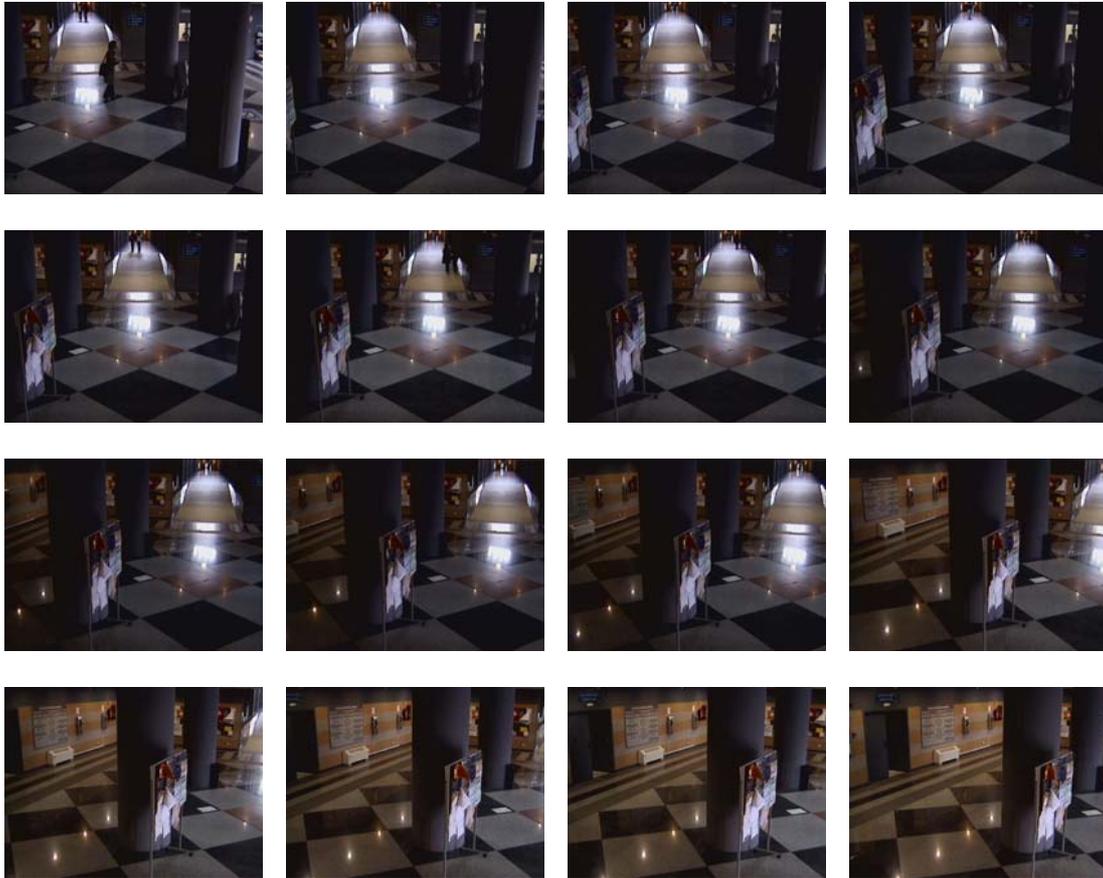


Fig. 4.13. Imágenes de la escena grabadas por PTZ 101 (barrido *pan*).

En este caso también hay que tener en cuenta que las imágenes han sido grabadas con el nivel de *zoom* más bajo y que las medidas del “DIN-A3” son de 420 x 290.

Con estas imágenes y estos datos se ejecuta el método explicado en el apartado 4.4 para hallar los parámetros extrínsecos de la cámara., el resultado obtenido del vector de translación es el siguiente:

$$t_{101} = 10,52 \text{ m.}$$

Se comprueba que los resultados indican que las cámaras están a la misma distancia del patrón de referencia como realmente ocurre. Con este hecho se pueden validar los resultados obtenidos.

Por último se muestran las imágenes de la cámara PTZ 102. Estas imágenes han sido capturadas mediante un barrido *tilt* de la escena. Como la distancia a la que se encuentra esta cámara es algo menor que a la que se encuentran las otras 2, en esta ocasión ha sido posible utilizar una DIN-A4 como referencia.

La siguiente figura muestra las 16 imágenes tomadas mediante el barrido *tilt* para ser utilizadas en la calibración.

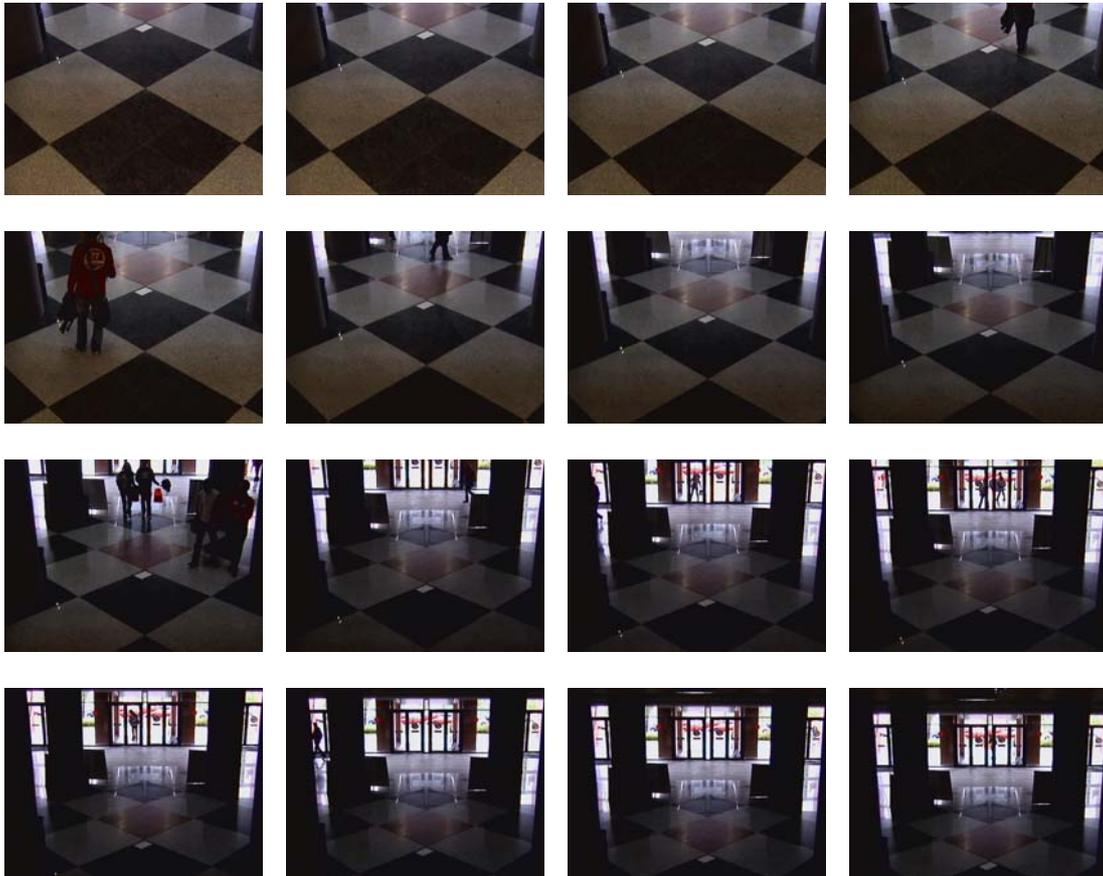


Fig. 4.14. Imágenes de la escena grabadas por PTZ 102 (barrido tilt).

En este caso también hay que tener en cuenta que las imágenes han sido grabadas con el nivel de zoom más bajo, aunque que las medidas del patrón son de 290 x 210.

Con estas imágenes y estos datos se ejecuta el método explicado en el apartado 4.4 para hallar los parámetros extrínsecos de la cámara., el resultado obtenido del vector de translación es el siguiente:

$$t_{102} = 7,11 \text{ m.}$$

Se comprueba que los resultados indican que cámara PTZ 102 está más cerca del sistema de referencia que las otras dos cámaras como realmente ocurre.

4.5.2. Conclusiones sobre la aplicación.

Con esta aplicación se ha comprobado nuevamente que el método desarrollado para calcular los parámetros extrínsecos de la cámara proporciona unos resultados correctos.

Se puede sintetizar la idea de este método en que es posible utilizar el método de Zhang para conocer las posiciones de las cámaras aunque el patrón de calibración que se utilice ocupe muy poco espacio en la imagen.

Esto es debido a que el error numérico cometido no se escala al no tener que calcular la cónica absoluta; y a que se puede minimizar tanto como se quiera el error numérico cometido (hasta el error de cuantificación) utilizando un mayor número de imágenes y promediando los valores obtenidos puesto que el error de cada uno es linealmente independiente del de los demás (teorema central del límite).

5. Conclusiones y trabajos futuros.

En este capítulo se desarrollan las conclusiones resultantes del presente PFC. Se va a presentar las conclusiones a las que se ha llegado sobre cada uno de los objetivos parciales que se plantearon en la introducción (capítulo 0) de esta memoria. Así mismo se van a plantear las líneas de mejora del presente proyecto para posibles trabajos futuros. Finalmente se exponen las conclusiones finales sobre el objetivo principal de este proyecto.

5.1. Conclusiones sobre el estudio de los métodos de calibración.

Se ha comprobado que la calibración de cámaras es un campo muy estudiado, en el que existe una gran cantidad de métodos diferentes que parten de distintos modelos. No obstante, también se ha comprobado que los métodos de calibración son muy inestables y que ofrecen resultados muy sensibles a una gran cantidad de factores.

En este trabajo se ha optado por estudiar e implementar un método de calibración en concreto: el método de Zhang. La razón por la que se ha escogido este método entre todos los posibles es su flexibilidad. A continuación se van a exponer las conclusiones a las que se ha llegado después de su implementación.

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.

5.1.1. Conclusiones sobre la autonomía del método de Zhang.

Los procesos de calibración no pueden ser, por definición, totalmente autónomos. En todos ellos existe un patrón sobre el que es necesario conocer sus medidas e introducirlas como datos de entrada. No obstante, el método de Zhang ofrece un grado de autonomía mayor que los demás métodos de calibración, al no ser necesario conocer la ubicación que ocupa el patrón en la escena. Esto hace que el método de Zhang sea mucho más flexible que cualquier otro método de calibración pues no precisa de una preparación exhaustiva de la escena.

En el método concreto que se ha implementado es necesario que el usuario marque los puntos del patrón que se van a utilizar para la calibración. Esto le resta cierta autonomía y constituye la primera de las mejoras que se pueden añadir a este trabajo: la detección automática de las esquinas del patrón mediante el método de detección de Harris.

5.1.2. Conclusiones sobre la eficiencia del método de Zhang.

El coste computacional del método de Zhang no es muy elevado. Únicamente precisa para su convergencia un número de 4 ó 5 imágenes, lo que lo convierte en un método altamente eficiente.

No obstante, por su sencillez, se ha optado por implementar el método en código *matlab*, por lo que otra posible línea de mejora del presente trabajo sería pasar el algoritmo del método a código C, mucho más eficiente computacionalmente.

5.1.3. Conclusiones sobre la versatilidad del método de Zhang.

Una de las principales ventajas del método de Zhang es su versatilidad, ya que utiliza un patrón plano para realizar la calibración. Además permite utilizar las esquinas como características de correspondencia, y las esquinas son características presentes en muchos tipos de escenas.

Estas dos consideraciones han sido fundamentales a la hora de desarrollar de forma eficiente el método que permite conocer la posición de las cámaras con respecto a un sistema de referencia común, puesto que permiten que se puedan considerar un gran número de escenas para calibrar sin preparaciones exhaustivas.

5.1.4. Conclusiones sobre la precisión del método de Zhang.

La precisión constituye la característica más importante de los métodos de calibración. También se ha indicado que es la más complicada de obtener. Obtener unos resultados precisos y estables ha sido uno de los trabajos más costosos del proyecto. A continuación se detallan las conclusiones a las que se ha llegado sobre las consideraciones que es necesario tener en cuenta para conseguir unos resultados numéricamente correctos. Estas conclusiones se han obtenido a partir de los resultados de las pruebas hechas sobre cámaras fijas y descritas en el capítulo tres.

En primer lugar hay que señalar una ventaja del método de Zhang sobre los demás métodos. Gracias al tipo de patrón utilizado, el método de Zhang gana en precisión, ya que al utilizar planos, reduce el número de fuentes de ruido en las coordenadas 3D, puesto que, el resto de los métodos necesitan que los puntos estén distribuidos en, al menos, dos planos, generalmente en forma de diedro, por lo que es necesario conocer con exactitud el ángulo que forman los dos planos.

A la hora de capturar las imágenes es necesario tener en cuenta que el patrón debe ocupar casi toda la imagen para evitar que la obtención de las coordenadas 2D de los puntos en la imagen sea poco precisa. Esto es debido a que si los puntos se encuentran muy cerca entre sí, el ruido introducido es muy alto y la relación señal a ruido muy baja.

Al mismo tiempo también es importante que el patrón no ocupe la imagen entera, porque la calibración se ve afectada por la distorsión radial, que es máxima en los bordes de la imagen.

Otro resultado al que se ha llegado mediante la realización de las pruebas se refiere al número de imágenes que es necesario emplear. Pese a que la teoría del método explica que con tres vistas es posible realizar el algoritmo, es recomendable que se utilicen 4 ó 5 vistas del patrón para que se produzca un resultado que no se vea muy afectado por el ruido. Tomar más vistas no es necesario porque no hay una mejora significativa.

En cuanto a los puntos se ha comprobado que es recomendable utilizar como origen de coordenadas del mundo el centro del patrón, ya que se producen menos errores de distorsión radial. Además, el número mínimo de puntos necesarios para realizar la calibración es siete u ocho. En caso de utilizar más, las mejoras no son apreciables.

Pero se ha comprobado que la característica más importante de precisión se consigue 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 se puede producir una estimación errónea de la cónica absoluta, lo que provoca que no se obtengan buenos resultados de precisión.

Para evitar falsas correspondencias es necesario disponer de un proceso de extracción de características estable y fiable. El marcado manual de las esquinas del patrón es un proceso fiable pero no estable ya que introduce cierta aleatoriedad. El detector de esquinas de Harris (posible elemento de mejora del trabajo que ya ha sido presentado) constituiría un proceso de extracción de características estable.

A la luz de los resultados obtenidos en las pruebas y que se han expuesto en el capítulo 3, se puede concluir que se han obtenido unos resultados fiables pero que se pueden mejorar mediante las líneas de mejora indicadas.

5.2. Conclusiones sobre el estudio de los métodos de autocalibración.

La primera conclusión que se extrae del estudio de los métodos de autocalibración es que se trata de una línea que todavía se está desarrollando y que, a diferencia de los métodos de calibración, constituye un campo de investigación que está muy abierto.

Las ventajas que ofrecen los métodos de autocalibración sobre los métodos de calibración son principalmente dos. La primera es que se pueden calibrar los parámetros de la cámara sin necesidad de conocer ningún dato de la escena y, por tanto, no requiere una preparación exhaustiva de la misma. La segunda ventaja es que permite una calibración *online* de los parámetros intrínsecos de la cámara si estos no permanecen constantes.

5.2.1. Conclusiones sobre el estudio de método de Hartley.

En este trabajo se han estudiado tres métodos de autocalibración. El primer método, (Hartley), sólo contempla el uso de cámaras que tienen movimiento de rotación y distancia focal fija. El problema que presentan los métodos de autocalibración con las cámaras que sólo tienen movimiento de rotación y no utilizan *zoom* es que es imposible determinar el

factor de escala λ , ya que un movimiento de rotación puro no aporta ninguna información sobre dicho parámetro.

5.2.2. Conclusiones sobre el estudio de método de Agapito.

Siguiendo la misma línea que el método de Hartley, el método de Agapito contempla el mismo tipo de cámaras pero con distancia focal variable. Este método podía constituir a priori el método ideal para alcanzar el objetivo principal del proyecto, pero tiene grandes inconsistencias para algunos movimientos determinados. En particular, si no existe ninguna rotación sobre el eje principal (eje z) de la cámara los resultados numéricos presentan un error muy alto. Este es el caso del movimiento *pan-tilt* que sólo tiene movimientos de rotación en los ejes horizontal (*pan*) y vertical (*tilt*). Además para este movimiento específico es imposible determinar algunos parámetros intrínsecos de la cámara como el factor de escalado k_u . Esta es la razón por lo que no se ha podido avanzar en esta línea para desarrollar el objetivo principal del proyecto.

5.2.3. Conclusiones sobre el estudio de método que calcula los parámetros intrínsecos de una red de cámaras PTZ.

Por último se ha presentado un método para calcular los parámetros intrínsecos de una red de cámaras PTZ basado en el método de Hartley. La implementación de este método constituye una de las posibles líneas de mejora de este trabajo.

5.3. Conclusiones sobre el estudio de la calibración de varias cámaras PTZ que graban una misma escena.

El objetivo principal de este proyecto era desarrollar un método que permitiese conocer la posición de las cámaras que graban una escena con respecto a un sistema de referencia común.

Como solución se ha planteado una aplicación basada en el método de Zhang que se pudiera aplicar de forma práctica (sin necesidad de que el patrón de calibración ocupe la mayor parte de la imagen). Esta aplicación se ha realizado mediante dos pasos. El primero consiste en calcular los parámetros intrínsecos de la cámara mediante el método de calibración de Zhang; y el segundo consiste en calcular los parámetros extrínsecos de la cámara en escenas en las que el patrón de calibración sea muy sencillo y ocupe una pequeña parte de la imagen.

El problema que planteaba esta solución es que para que mediante el método de Zhang se obtengan unos resultados correctos, es necesario que el patrón de calibración, ocupe casi toda la imagen. De lo contrario, si los puntos considerados están muy cerca, la relación señal a ruido que se introduce es muy baja y los resultados tienen un error muy grande.

Evidentemente utilizar un patrón de calibración que ocupe casi toda la imagen era inviable para la aplicación que se deseaba desarrollar, ya que se pretendía que las tres cámaras estén grabando el mismo patrón.

La solución que se ha planteado es considerar varias imágenes de la misma escena grabadas con distinto zoom. En este caso la matriz de parámetros extrínsecos no varía para ninguna de las imágenes. De esta forma se puede pretender utilizar un patrón lo suficientemente pequeño y obtener unos resultados precisos.

Las razones por las que es viable utilizar un patrón pequeño para calcular los parámetros extrínsecos son dos. La primera es que al no tener que calcular la matriz de parámetros intrínsecos, no hay que calcular la cónica absoluta por lo que no se escala el error. La segunda es que si se consideran varias imágenes, el error cometido entre cada una de ellas es linealmente independiente, por lo que si se promedian los resultados obtenidos en cada imagen estos tienden a converger a su valor real (por el teorema central del límite).

Se ha comprobado mediante pruebas calculando el vector de translación que el método de calibración propuesto es viable ya que a partir de un número de diez imágenes los resultados obtenidos tienen un error muy pequeño (próximo al error de cuantificación) para una distancia de cuatro metros.

Estos resultados han permitido concluir que el método propuesto es versátil y fiable aunque no es autónomo y su tiempo de computación depende linealmente del número de imágenes utilizadas.

5.3.1. Conclusiones sobre la autonomía del método propuesto.

Al igual que el método de Zhang, el método que se ha propuesto no es autónomo por definición ya que utiliza un patrón de calibración. No obstante este patrón es muy sencillo (un DIN-A4) y casi puede ser utilizado en cualquier contexto.

5.3.2. Conclusiones sobre la eficiencia del método propuesto.

En primer paso del método propuesto es utilizar el método de Zhang, por lo que en este caso cabe hacer las mismas consideraciones que en el apartado 5.3.1.

En el segundo paso no es necesario calcular la cónica absoluta por lo que el tiempo de computación de una sola imagen es más reducido. No obstante sigue siendo alto al tratarse de un método iterativo. Además este tiempo se multiplica por cada imagen utilizada.

5.3.3. Conclusiones sobre la versatilidad del método propuesto.

El método es versátil ya que como patrón de calibración se utiliza una DIN-A4 (o DIN-A3 si la distancia es muy grande) que ocupe una parte muy pequeña de la imagen. Además, únicamente es necesario marcar un punto del patrón y, a partir de él, el método calcula el resto de los puntos de la escena. Otra posible línea de mejora de este proyecto es diseñar un

método a partir del creado que no necesite marcar ningún punto del patrón sino que lo detecte directamente.

5.3.4. Conclusiones sobre la precisión del método propuesto.

El método es fiable ya que se obtienen unos resultados muy precisos con un número de imágenes relativamente bajo. Si se quiere aumentar esta precisión basta con aumentar el número de imágenes obtenidas. El límite máximo de precisión viene dado por el error de cuantificación de la imagen que depende del tamaño en pixels del plano imagen y de la distancia a la que se encuentra el patrón. La precisión también depende de la fiabilidad de los resultados obtenidos al calcular los parámetros intrínsecos mediante el método de Zhang. Esta precisión ha sido comentada en el apartado 5.1.4. La fiabilidad del método también ha sido probada con una aplicación que mide la distancia de las tres cámaras a un punto.

Anexo A. Descomposición en valores singulares (SVD)

Una de las operaciones más robustas que se usan en álgebra lineal es la Singular Value Decomposition (SVD). Suponer una matriz $\mathbf{C} \in R^{m \times n}$. El teorema de la SVD dice que existen tres matrices $\mathbf{D} \in R^{m \times n}$, $\mathbf{U} \in R^{m \times m}$, y $\mathbf{V} \in R^{n \times n}$, con las que se puede escribir:

$$\mathbf{C} = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{V}^T$$

Estas matrices tienen las siguientes características:

1. La matriz $\mathbf{D} \in R^{m \times n}$ es diagonal con valores positivos o cero (valores singulares).
2. La matriz $\mathbf{U} \in R^{m \times m}$ es ortonormal, $\mathbf{U}^T \cdot \mathbf{U} = \mathbf{I}$.
3. La matriz $\mathbf{V} \in R^{n \times n}$ es ortonormal, $\mathbf{V}^T \cdot \mathbf{V} = \mathbf{I}$.

Nótese que la j -ésima fila de la matriz \mathbf{V}^T (es decir la j -ésima columna de \mathbf{V}) se relaciona con el j -ésimo valor singular de \mathbf{D} y lo mismo se puede decir de la j -ésima columna de \mathbf{U} .

La SVD siempre se puede realizar y es “única” (en el sentido de que las únicas diferencias pueden ser el orden de las columnas de dichas matrices). La SVD permite resolver muchos de los problemas habituales que se encuentran en álgebra lineal: determinar el condicionamiento de una matriz, calcular la inversa de matrices, calcular EOFs, y resolver sistemas lineales.

Kernel y Rango de una aplicación

Sea una aplicación lineal $R^n \rightarrow R^n$ caracterizada por la ecuación:

$$\mathbf{C} \cdot \mathbf{x} = \mathbf{y}$$

con \mathbf{C} una matriz cuadrada. Si \mathbf{C} es singular, quiere decir que existe un subconjunto de soluciones diferentes a la trivial para las cuales:

$$\mathbf{C} \cdot \mathbf{x} = \mathbf{0}$$

El conjunto de todas las soluciones $\mathbf{x} \in R^n$ de dicha ecuación es el Kernel (o espacio nulo) de dicha aplicación.

El poder de la SVD es que directamente construye bases de los subespacios nulo y rango. Las columnas de \mathbf{U} relacionadas con valores singulares diferentes de cero forman una base del subespacio del rango. Las columnas de \mathbf{V} asociadas a valores singulares nulos forman una base de espacio nulo. De esta manera la SVD permite encontrar todas las soluciones de la ecuación

$$\mathbf{C} \cdot \mathbf{x} = \mathbf{0}$$

Bibliografía

- [1] R. Hartley and A. Zisserman, *Multiple View Geometry In Computer Vision*, Cambridge University Press, 2000.
- [2] Yao Wang, Jörn Ostermann y Ya-Qin Zhang. *Video Processing and communications*. Prentice-Hall, 2001.
- [3] Olivier Faugeras. *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, 1993.
- [4] Olivier Faugeras y Quang-Tuan Luong. *The geometry of multiples images*. MIT Press, 2001.
- [5] [Agapito et al., 2000] Agapito, L. D., Hayman, E., and Reid, I. (2000). Selfcalibration of rotating and zooming cameras. Technical Report OUEL 0225/00, Department of Engineering Science, University of Oxford.
- [6] [Hartley, 1994b] Hartley, R. (1994b). Self-calibration from multiple views with a rotating camera.
- [7] [Hartley, 1997] Hartley, R. Self-calibration of stationary cameras (context).
- [8] [Yongduek Seo K, 1998] Yongduek Seo K. Auto -calibration of a Rotating and Zooming Camera .
- [9] [Sinha, Pollefeys]] Sudipta N. Sinha and Marc Pollefeys. Towards Calibrating a Pan-Tilt-Zoom Camera Network. Department of Computer Science, University of North Carolina at Chapel Hill.
- [10] [Zhang, 2000] Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactios on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334.

PRESUPUESTO

1) Ejecución Material	
• Ordenador personal	2.000 €
• Material de oficina	150 €
• Total de ejecución material	2.150 €
2) Gastos generales	
• 16 % sobre Ejecución Material	344 €
3) Beneficio Industrial	
• 6 % sobre Ejecución Material	129 €
4) Honorarios Proyecto	
• 640 horas a 15 € / hora	9600 €
5) Material fungible	
• Gastos de impresión	60 €
• Encuadernación	200 €
6) Subtotal del presupuesto	
• Subtotal Presupuesto	12010 €
7) I.V.A. aplicable	
• 16% Subtotal Presupuesto	1921.6 €
8) Total presupuesto	
• Total Presupuesto	13931,6 €

Madrid, Mayo de 2007

El Ingeniero Jefe de Proyecto

Fdo.: Javier García Ocón
Ingeniero Superior de Telecomunicación

PLIEGO DE CONDICIONES

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de un sistema para calibrar los parámetros de una red de cámaras PTZ. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego.

Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

Condiciones generales

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.
2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.
3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.
4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.
5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.
6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.
7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.
8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4% del presupuesto y la provisional del 2%.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la

misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad "Presupuesto de Ejecución de Contrata" y anteriormente llamado "Presupuesto de Ejecución Material" que hoy designa otro concepto.

Condiciones particulares

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.