

AUTOCALIBRACIÓN DE UNA CÁMARA PTZ

ANDRÉS MAURICIO DE ANGULO BARRERA

PONTIFICIA UNIVERSIDAD JAVERIANA

FACULTAD DE INGENIERÍA

DEPARTAMENTO DE ELECTRÓNICA

BOGOTÁ D.C

NOVIEMBRE DE 2010

AUTOCALIBRACIÓN DE UNA CÁMARA PTZ

ANDRÉS MAURICIO DE ANGULO BARRERA

TRABAJO DE GRADO PARA OPTAR POR EL TITULO DE INGENIERO ELECTRÓNICO

DIRECTOR

ING. JULIÁN QUIROGA SEPÚLVEDA, M.Sc

PONTIFICIA UNIVERSIDAD JAVERIANA

FACULTAD DE INGENIERÍA

DEPARTAMENTO DE ELECTRÓNICA

BOGOTÁ D.C

NOVIEMBRE DE 2010

PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE ELECTRÓNICA

RECTOR MAGNIFICO: P. JOAQUÍN EMILIO SÁNCHEZ GARCÍA, S.J.
DECANO ACADEMICO: Ing. FRANCISCO JAVIER REBOLLEDO MUÑOZ
DECANO DEL MEDIO UNIVERSITARIO: P. SERGIO BERNAL RESTREPO, S.J.
DIRECTOR DE CARRERA: Ing. JUAN MANUEL CRUZ
DIRECTOR DEL TRABAJO DE GRADO: Ing. JULIÁN QUIROGA SEPÚLVEDA

ARTÍCULO 23 DE LA RESOLUCIÓN N° 13 DE JUNIO DE 1946

“La universidad no se hace responsable de los conceptos emitidos por sus alumnos en sus trabajos de grado. Solo velara porque no se publique nada contrario al dogma y la moral católica y porque los trabajos no contengan ataques o polémicas puramente personales. Antes bien que se vea en ellos el anhelo de buscar la verdad y la justicia”

Ante todo quiero darle gracias a Dios por haberme dado la familia que tengo, gracias a ella he logrado llegar hasta este punto en mi vida.

A mis papás les doy las gracias por darme la oportunidad de estudiar y formarme como profesional en una Universidad tan importante como la Pontificia Universidad Javeriana. Espero contar como siempre con ese apoyo incondicional para la nueva etapa que comienzo.

A mis hermanos les doy las gracias por estar siempre ahí, ya sea para ayudar o simplemente para hacerme compañía.

A Juli le doy las gracias por tenerme paciencia y apoyarme durante toda esta época del desarrollo del trabajo.

Al Ingeniero Julian Quiroga le doy las gracias por guiarme y aconsejarme durante este proceso.

TABLA DE CONTENIDO

1. INTRODUCCIÓN	11
2. OBJETIVOS	12
2.1. OBJETIVO GENERAL	12
2.2. OBJETIVOS ESPECÍFICOS	12
3. MARCO CONCEPTUAL	12
3.1. GEOMETRÍA PROYECTIVA	12
3.2. TRANSFORMACIONES PROYECTIVAS 2D: HOMOGRAFÍAS	12
3.3. TRANSFORMACIONES PROYECTIVAS 3D	13
3.4. MODELO CÁMARA <i>PINHOLE</i>	14
3.5. PROYECCIÓN 3D-2D	16
3.6. TRANSFORMACIÓN CÁMARA-IMAGEN	17
3.7. TRANSFORMACIÓN ESCENA-CÁMARA	17
3.8. RANSAC	19
3.9. HARRIS CORNER DETECTOR	19
3.10. DESCOMPOSICIÓN EN VALORES SINGULARES (SVD)	20
3.11. FACTORIZACIÓN DE CHOLESKY	20
4. DESCRIPCIÓN GENERAL Y ESPECIFICACIONES	21
4.1. DESCRIPCIÓN DEL LUGAR Y DEL SISTEMA DE CAPTURA	21
4.1.1. <i>Movimiento pan de la cámara PTZ</i>	22
4.1.2. <i>Movimiento tilt de la cámara</i>	22
4.2. SOFTWARE EVI SERIES	23
4.3. DIAGRAMA DE BLOQUES	24
4.3.1. <i>Obtención de homografías Hij</i>	25
4.3.2. <i>Planteamiento de ecuaciones</i>	25
4.3.3. <i>Solución y obtención de la matriz A</i>	25
4.3.4. <i>Obtención de la matriz de Rotación</i>	26
4.4. DIAGRAMA DE FLUJO	27
4.5. IMPLEMENTACIÓN DEL MÉTODO DE ZHANG. TOOLBOX DE CALIBRACIÓN	28
5. DESARROLLO TEÓRICO	29
5.1. MÉTODOS DE CALIBRACIÓN UTILIZADOS	29
5.1.1. <i>Calibración clásica: método de Zhang</i>	29
5.1.2. <i>Autocalibración: método lineal de Agapito</i>	30
5.2. ESTIMACIÓN DE DISTANCIA	36
6. RESULTADOS ALGORITMO DE AUTOCALIBRACIÓN	37
6.1. PROTOCOLOS DE PRUEBAS	37
6.2. PARÁMETROS INTRÍNSECOS	37

6.2.1.	<i>Video número uno.</i>	37
6.2.2.	<i>Video número dos.</i>	41
6.2.3.	<i>Video número tres.</i>	44
6.2.4.	<i>Mínimo nivel de zoom.</i>	48
6.2.5.	<i>Máximo nivel de zoom.</i>	50
6.2.6.	<i>Variación del número de imágenes.</i>	53
6.3.	PRUEBAS DE EVALUACIÓN DE ROTACIÓN.	55
6.4.	PRUEBA CON OTRO TIPO DE ESCENA	57
6.5.	ESTIMACIÓN DE DISTANCIA.	61
7.	CONCLUSIONES	65
8.	BIBLIOGRAFÍA	67
9.	ANEXOS	68
A.1	LISTADO DE FUNCIONES.	68
A.2	SECUENCIAS DE IMÁGENES	68
A.3	EVI SERIES SOFTWARE.	68
A.4	CÓDIGO DESARROLLADO.	68
A.5	VIDEOS UTILIZADOS.	68
A.1	LISTADO DE FUNCIONES.	69

Listado de Figuras

Figura 1 Ejemplo de transformación proyectiva, $x' = Hx$. Donde la cámara gira en torno a su centro. Tomado de [6].	13
Figura 2 Transformación Euclidiana entre coordenadas del mundo y de la cámara. Tomado de [6].	14
Figura 3 Geometría cámara Pinhole. Tomado de [4].	15
Figura 4 Cambio de coordenadas (c, u, v) Tomado de [4].	17
Figura 5 Ejemplo Detector de esquinas Harris.	20
Figura 6 Distintas escenas usadas para probar el algoritmo de autocalibración.	21
Figura 7 Cámara PTZ EVI-D70 con sus respectivas entradas y salidas. Tomado de [11]	22
Figura 8 Movimiento Pan de la cámara. Tomado de [11]	22
Figura 10 Valores Hexadecimales de la posición de la cámara. Tomado de [11]	23
Figura 9 Movimiento Vertical (Tilt) de la cámara. Tomado de [11]	23
Figura 11 Interfaz gráfica del Software de Sony.	24
Figura 12 Diagrama de Bloques.	24
Figura 13 Diagrama de Flujo.	28
Figura 14 Puntos que hay que marcar (las esquinas del tablero).	28
Figura 15 Plantilla utilizada para el método clásico	29
Figura 16 Barrido y cambios sobre un punto de interés.	32
Figura 17 El algoritmo detector de esquinas de Harris implementado para I_0, I_j	32
Figura 18 Ejemplo de ajuste de un modelo sobre un conjunto de puntos. En la figura a solo cuatro puntos son considerados como <i>inliers</i> , mientras que en la figura B la mayoría de puntos son <i>inliers</i> . Las líneas punteadas son el <i>threshold</i> y la distancia d es la distancia que hay entre la línea negreada hasta al punto. Tomado de [6]	33
Figura 19 Resultado de RANSAC, donde se superponen las imágenes y se grafican los puntos denominados <i>inliers</i> .	33
Figura 20 Relación de distancias con el ángulo recorrido por la cámara.	36
Figura 21 Recorrido realizado por la cámara PTZ	37
Figura 22 Recorrido de la escena.	38
Figura 25 Recorrido estimado de la cámara	39
Figura 23 Comparación de la distancia focal.	39
Figura 24 Valores del punto principal obtenidos para las distintas homografías. El círculo rojo indica el valor obtenido del Toolbox.	39
Figura 26 Error de la distancia focal obtenida para cada imagen.	40

Figura 27 Error del Punto Principal en pixels.....	41
Figura 28 Recorrido de la cámara.....	42
Figura 29 Resultados del punto principal, el círculo rojo es el valor obtenido utilizando el Toolbox.	42
Figura 30 Rotación estimada.	43
Figura 31 Distancia Focal estimada (puntos rojos), distancia focal de referencia (línea azul).....	43
Figura 32 Imágenes del recorrido realizado por la cámara.....	45
Figura 33 Puntos Principales estimados para cada homografía y el referencial (círculo rojo).....	45
Figura 34 Distancia Focal estimada (puntos rojos) y referencial	45
Figura 35 Rotación estimada.	46
Figura 36 Error de la Distancia Focal dado en pixeles.	47
Figura 37 Error en el punto principal.....	47
Figura 38 Imágenes del recorrido realizado.	48
Figura 39 Distancia Focal estimada (puntos rojos) y la distancia focal de referencia (línea azul).....	48
Figura 40 Puntos Principales estimados y punto principal referencial (círculo rojo).....	48
Figura 41 Rotación de la cámara estimada.	49
Figura 42 Error Distancia Focal.	50
Figura 43 Error del punto principal.	50
Figura 44 Imágenes del barrido de la escena.....	50
Figura 45 Distancia Focal estimada (Puntos rojos) y distancia focal de referencia (línea azul)	51
Figura 46 Gráfica de Distancia focal vs el Error porcentual.	53
Figura 47 Distancia vs Error porcentual.....	53
Figura 48 Valores de la Distancia Focal Respecto al número de imágenes utilizadas.....	54
Figura 49 Error Porcentual de la distancia focal respecto al número de imágenes utilizadas.....	55
Figura 50 Comparación de la rotación estimada (línea azul) y la real (línea verde).	55
Figura 51 Recorrido real (línea verde) y estimado de la cámara (línea azul).	56
Figura 52 Recorrido de la cámara PTZ.....	57
Figura 53 Imágenes del recorrido realizado.	58
Figura 54 Punto Principal estimado.....	58
Figura 55 Distancia Focal estimada (puntos rojos) y la distancia focal de referencia (línea azul).....	58
Figura 56 Rotación Estimada.....	59
Figura 57 Error en la distancia focal estimada.	60
Figura 58 Error en el punto principal estimado	60
Figura 59 Rotación estimada (línea azul) y rotación real (línea verde).....	60

Figura 60 Imágenes A y B para encontrar la homografía correspondiente. La primera imagen se ajusta para que coincida el borde la imagen capturada con el punto de origen de los 15cm. La segunda imagen se captura cuando el borde derecho de la imagen llega al punto final de los 15cm.....	62
Figura 61 Vista superior marcando los puntos A y B correspondientes a las imágenes tomadas.	62
Figura 62 Imágenes superpuestas y la homografía resultante.	62
Figura 63 Imágenes utilizadas para la homografía	63
Figura 64 Superposición de las imágenes y la homografía resultante	63
Figura 65 Imágenes utilizadas para encontrar la altura de la matera que se encuentra en la mitad de la imagen.	64
Figura 66 Homografía resultante	64

Listado de Tablas

Tabla 1 Desviación Estandar.....	40
Tabla 2 Desviación Estándar.....	44
Tabla 3 Desviación Estándar.....	46
Tabla 4 Desviación Estándar.....	49
Tabla 5 Resumen de resultados con todos los videos utilizados.....	52
Tabla 6 Tabla de valores de la distancia focal respecto al número de imágenes.	54
Tabla 7 Evaluación de los datos de rotación.	56
Tabla 8 Evaluación de los datos de rotación.	57
Tabla 9 Tabla comparativa de la rotación de la cámara.	61
Tabla 11 Resumen de los distancias estimadas y sus respectivos errores.....	64

1. Introducción

La autocalibración es el proceso de determinar los parámetros internos de una cámara directamente de un grupo de imágenes de una escena sin calibrar. Esta calibración permite la obtención de los parámetros que definen las condiciones de formación de la imagen dentro del campo de la visión artificial, entre estos parámetros encontramos los que definen la geometría interna y la óptica de la cámara (parámetros intrínsecos), así como los que definen su posición y orientación (parámetros extrínsecos) respecto a un objeto de referencia o patrón de calibración.

La gran ventaja que tiene la autocalibración es que se evita la complicada y larga tarea de calibrar la cámara usando objetos especiales para calibrar, en consecuencia se obtiene una gran flexibilidad ya que por ejemplo se puede calibrar una cámara directamente con una secuencia de imágenes sin saber el movimiento realizado o sus parámetros internos [6].

La calibración de cámaras tiene distintas aplicaciones, una aplicación es propuesta en *Self Calibration of rotating and zooming camera* donde utilizan los datos de calibración para generar una vista panorámica en 360° a partir de una secuencia de imágenes. Entre otros ejemplos de calibración encontramos que se puede utilizar para 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, como lo propuso Zhang [12]. Un área muy interesante es la visión por computador, donde también puede facilitar la navegación de un robot móvil, permitiendo evitar obstáculos, dirigirse a objetos determinados o facilitar la definición de la trayectoria más adecuada para alcanzar su destino.

La calibración de una cámara es un problema complejo ya que son muchos los parámetros a resolver y los factores que influyen sobre los resultados. En muchos métodos de autocalibración esta complejidad ha sido reducida porque se han empleado modelos ideales o simplificados de las cámaras, con estas simplificaciones se consiguen resultados bastante aceptables (aunque no exactos) con la ventaja de no tener que parametrizar muchos factores. El modelo de cámara de agujero o *pinhole*, que es el utilizado por la mayoría de los métodos de calibración y en este caso es el que se va a utilizar para el trabajo de grado, no parametriza aspectos ópticos tales 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 para este trabajo de grado no se parametriza [4].

El algoritmo utilizado en este trabajo de grado es el método lineal planteado por Agapito, Hartley y Hayman en el artículo *Linear calibration of a rotating and zooming camera*, donde se realiza la calibración de una cámara PTZ planteando y resolviendo un sistema de ecuaciones que son lineales, esto se realiza a partir de homografías y utilizando el concepto de la cónica absoluta [1].

2. Objetivos

2.1. Objetivo general

El objetivo principal del trabajo de grado es implementar un método que permita la calibración automática de una cámara PTZ a partir de un conjunto de imágenes sin calibrar de la escena.

2.2. Objetivos específicos

Los objetivos específicos de este trabajo de grado se enumeran a continuación:

- Encontrar la homografía en un par de imágenes a partir de la correspondencia de un conjunto de puntos característicos.
- Plantear un sistema de ecuaciones para expresar los elementos de la matriz de la cónica absoluta en términos de las homografías entre cada par de vistas.
- Resolver el sistema de ecuaciones previamente planteado usando SVD, encontrar la matriz de calibración de la cámara K y las respectivas matrices de rotación R .
- Evaluar la estimación del punto principal, distancia focal y rotación de la cámara para un barrido con zoom constante.

3. Marco conceptual

En este capítulo se presentan los conceptos de geometría proyectiva que están asociados al proceso de captación de una imagen. Esto es muy importante porque el proceso de autocalibración se basa en el modelo de una cámara *pinhole* que su base es la proyección de puntos de una escena en la imagen.

3.1. Geometría proyectiva

Se llama geometría proyectiva a una estructura matemática que estudia las incidencias de puntos y rectas sin tener en cuenta la medida. La geometría proyectiva suministra un modelo lineal del proceso de captación de imágenes (si no hay distorsiones) 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.

3.2. Transformaciones proyectivas 2D: homografías.

Una homografía es la transformación proyectiva de un punto que se encuentra en una imagen al mismo punto en otra imagen. Por definición, se entiende que la geometría proyectiva en dos dimensiones es el estudio de las propiedades del plano proyectivo P^2 que son invariantes bajo un grupo de transformaciones conocidas como 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:

$$h(\mathbf{x}) = \mathbf{x}' = \mathbf{H} \cdot \mathbf{x} \text{ donde } \mathbf{x}, \mathbf{x}' \in P^2$$

Matricialmente se tiene:

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

Se tiene que \mathbf{H} es una matriz 3×3 no singular, \mathbf{x}' es la transformación lineal \mathbf{H} de \mathbf{x} . Esta transformación es biunívoca entre dos planos 2D, cuyos puntos son representados homogéneamente por \mathbf{x} y \mathbf{x}' . Es decir, un punto en un plano 2D tiene una única correspondencia en un punto de otro plano 2D, y viceversa [9].

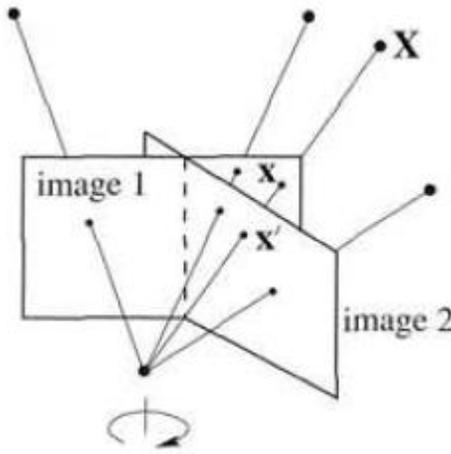


Figura 1 Ejemplo de transformación proyectiva, $\mathbf{x}' = \mathbf{H}\mathbf{x}$. Donde la cámara gira en torno a su centro. Tomado de [6].

3.3. Transformaciones proyectivas 3D.

La única transformación proyectiva 3D que se va a estudiar en el marco teórico es la transformación 3D Euclídea, se va a estudiar este tipo de transformación ya que representa los cambios de coordenadas que pueden sufrir los objetos rígidos al pasar de un sistema de coordenadas a otro. Este tipo de transformación es el utilizado en el modelo de cámara *Pinhole*.

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 $x = \frac{X}{T}, y = \frac{Y}{T}, z = \frac{Z}{T}$

Una forma sencilla de pasar de coordenadas no homogéneas a homogéneas es agregando un uno al final del vector, es decir $\mathbf{M} = [X, Y, Z, 1]$.

Dado un sistema de coordenadas 3D (x, y, z) que ha sufrido una rotación y una traslación, el espacio 3D en el nuevo sistema de coordenadas (x', y', z') queda definido por una transformación 3D Euclídea definida por:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathbf{R} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \mathbf{t}$$

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.

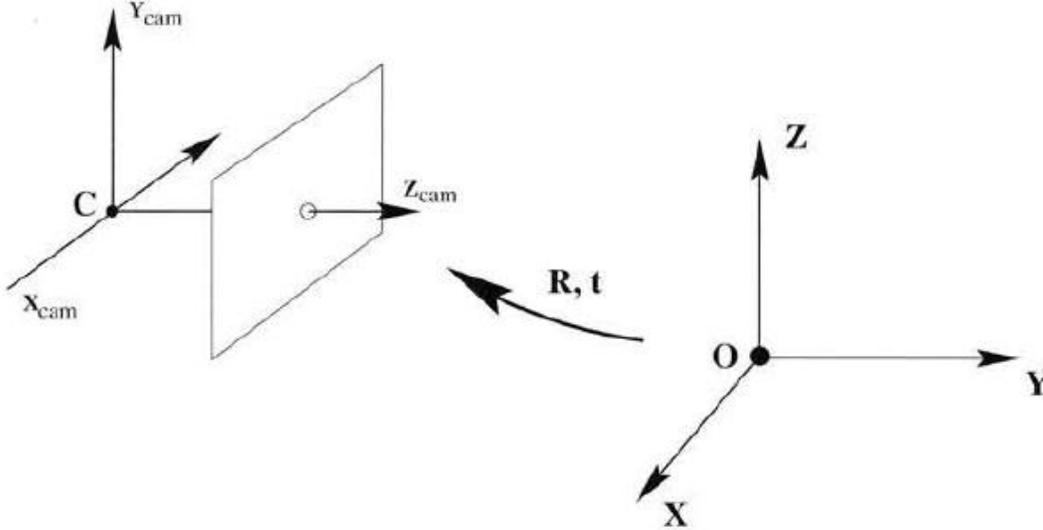


Figura 2 Transformación Euclidiana entre coordenadas del mundo y de la cámara. Tomado de [6]

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

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

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

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

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 traslación \mathbf{t}) en un solo paso (multiplicar por la matriz $\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} = [\mathbf{R} \ \mathbf{t}]$ [4].

3.4. Modelo cámara *pinhole*

La gran mayoría de los procedimientos de calibración se basan en el modelo de cámara de agujero o *pinhole*, este modelo es el más simple que se puede obtener de una cámara. El modelo *Pinhole* es el que necesita un menor número de parámetros para ser representado. Se basa en que 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 (Figura 3). Básicamente, este modelo aplica una matriz de proyección para transformar las coordenadas 3D de los puntos del objeto en coordenadas 2D de la imagen:

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

Donde $M = [X_w, Y_w, Z_w, 1]^T$ es el vector que contiene las coordenadas del punto en el sistema de referencia exterior a la cámara, \mathbf{P} es una matriz de 3×4 denominada matriz de proyección, $m = [u, v, 1]^T$ es el vector de las coordenadas del punto proyectado en la imagen y transformado a píxeles.

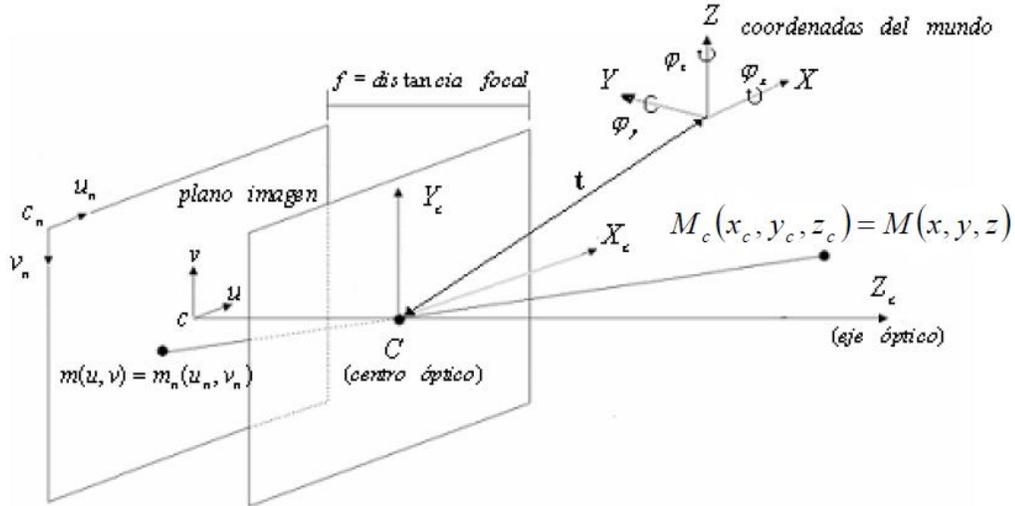


Figura 3 Geometría cámara Pinhole. Tomado de [4]

Como se ilustra en la Figura 3., 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 3, los ejes de este sistema de coordenadas son X, Y, Z.

- *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 3, los ejes de este sistema de coordenadas son $X_c Y_c Z_c$. Su origen es el centro óptico de la cámara, el punto \mathbf{M} y el punto \mathbf{M}_c son el mismo punto.

- *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 3). Como se ve en esta Figura, los ejes de este sistema de coordenadas son u, v .

- *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 3.). 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 3 [9].

Los parámetros que caracterizan una matriz de proyección en el modelo de la cámara *pinhole* se clasifican en dos tipos:

Parámetros intrínsecos.

Son aquellos que definen el modelo de cámara utilizado, estableciendo sus características geométricas y ópticas. Los parámetros intrínsecos básicos son:

- Centro del eje óptico o punto principal (P_x, P_y). Define el punto donde el eje óptico atraviesa el plano imagen. Las coordenadas de este punto vienen dadas en píxeles.
- Distancia focal en píxeles (f_x, f_y). Es un parámetro que indica 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. Frecuentemente este parámetro se descompone a su vez en:
 - Factores de conversión pixel-milímetros (d_u, d_v). Indica 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 foco y el plano imagen, viene dada en mm.
 - Factor de proporción (S_x). Indica la relación de tamaño entre la dimensión horizontal y vertical de un pixel.
 - La relación entre ellos (cuando no existe distorsión) viene dada por:
 $f_x = S_x d_u f$ y $f_y = d_v f$.

Parámetros extrínsecos. Son los que definen la relación (traslación y rotación) entre un sistema de coordenadas absoluto (exterior a la cámara) y otro ligado a la cámara. Se incluyen 6 parámetros: tres para la traslación (T_x, T_y, T_z) y tres para los ángulos rotados sobre cada uno de los ejes ($\varphi_x, \varphi_y, \varphi_z$).

Para determinar la forma de la matriz de proyección se estudia cómo se relacionan las coordenadas del punto M , expresadas en base a las coordenadas del mundo (ejes X, Y, Z de la figura 3), con las coordenadas del punto m , expresadas en referencia a las coordenadas normalizadas de la imagen (ejes $u_n v_n$) [4]. A continuación se explica la manera en que se proyecta un punto en las coordenadas del mundo al sistema de coordenadas de la imagen.

3.5. 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:

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

3.6. 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 4.

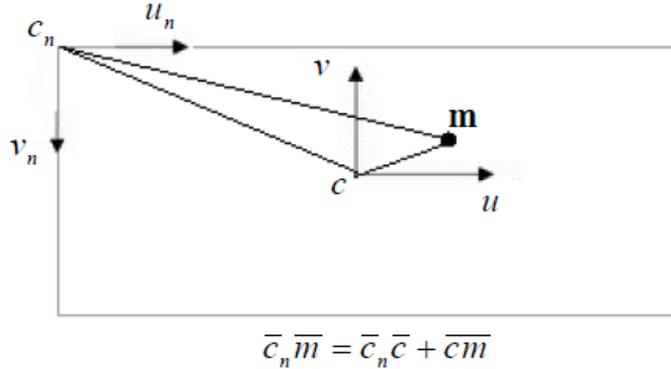


Figura 4 Cambio de coordenadas (c, u, v) Tomado de [4]

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

$$u_n = d_u \cdot u + P_x \quad v_n = d_v \cdot v + P_y$$

Con estos dos primeros pasos queda constituida la matriz de parámetros intrínsecos (en adelante A) que relaciona el punto 3D M_c (referido a las coordenadas de la cámara) con el punto m_n (referido a las coordenadas normalizadas de la imagen). En coordenadas homogéneas:

$$\begin{bmatrix} U_n \\ V_n \\ S_n \end{bmatrix} = \underbrace{\begin{bmatrix} f \cdot d_u & 0 & P_x & 0 \\ 0 & f \cdot d_v & P_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_A \cdot \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (1.1)$$

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.

3.7. 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:

$$\begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (1.2)$$

$$M_r = R \cdot M$$

Seguida de una translación:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (1.3)$$

$$M_c = M_r + \mathbf{t}$$

Combinando las ecuaciones 1.2 y 1.3 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 $[\mathbf{R} \ \mathbf{t}]$):

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1.4)$$

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 1.1 (\mathbf{A}) y la matriz de parámetros extrínsecos de la ecuación 1.4. $[\mathbf{R} \ \mathbf{t}]$:

$$\mathbf{A} = \begin{bmatrix} f \cdot d_u & 0 & P_x & 0 \\ 0 & f \cdot d_v & P_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (1.5)$$

$$[\mathbf{R} \ \mathbf{t}] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_x \\ r_{31} & r_{32} & r_{33} & t_x \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1.6)$$

$$\mathbf{P} = \begin{bmatrix} f \cdot d_u & 0 & P_x & 0 \\ 0 & f \cdot d_v & P_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_x \\ r_{31} & r_{32} & r_{33} & t_x \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1.7)$$

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:

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

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

Tomado de [4][6][9][10].

El modelo que se ha introducido no tiene en cuenta los efectos de distorsión de la lente.

A continuación se explicarán otros conceptos que se utilizan para poder autocalibrar una cámara. Para encontrar la homografía entre dos imágenes se utilizan algoritmos como el de RANSAC y el *Harris corner detector* mientras que para la solución del sistema de ecuaciones se utiliza el *Singular Value Decomposition* y la factorización de Cholesky.

3.8. RANSAC

RANSAC es una abreviatura para *Random Sample Consensus*. Es un método iterativo para estimar parámetros de un modelo matemático de un sistema de datos observados. RANSAC es un algoritmo no determinístico en el sentido que da resultados con cierta probabilidad, esta probabilidad se incrementa con el número de iteraciones realizadas por el algoritmo. El algoritmo primero fue publicado por Fischler y Bolles en 1981. El método se basa en ir cogiendo pequeños conjuntos de puntos y en calcular el modelo a partir de esos puntos. Además estima cuántos puntos del conjunto completo están dentro de una distancia mínima con respecto a ese modelo. Este proceso se repite hasta obtener un conjunto de puntos grandes que cumplan con el modelo.

En este algoritmo se asume que los datos contienen *inliers*, es decir, los datos que pueden ser ajustados a un modelo, y los *outliers* que son los datos que no pueden ser encajados en el modelo. Los *outliers* pueden existir por ejemplo por los valores extremos del ruido, o por medidas erróneas o de hipótesis incorrectas sobre la interpretación de datos [10].

Este algoritmo es aplicado a la correspondencia de puntos encontrados con el detector de esquinas de Harris. En resumen, se selecciona al azar un grupo de cuatro correspondencias y RANSAC encuentra el modelo que más se ajuste a estos puntos, se realizan m iteraciones y al final se escoge la homografía que tenga el mayor número de aciertos, mejor dicho, el modelo que más se ajuste al grupo de puntos.

3.9. Harris corner detector

El detector de esquinas de Harris permite detectar esquinas o puntos aislados en una imagen. En estos puntos existe una variación de intensidad en todas las direcciones. Esta variación de intensidad se obtiene luego de medir los cambios cuando se desliza una ventana alrededor del punto de interés. Por ejemplo, si la ventana se desliza alrededor de un área plana no existirán cambios en la intensidad lo cual indica que no es una esquina, por otro lado si la ventana se desliza sobre un borde existirá un cambio en la intensidad sólo en una dirección y en consecuencia tampoco se dirá que es una esquina. Si se desliza la ventana sobre una esquina habrá cambios en todas las direcciones y en consecuencia se concluirá que el punto es una esquina. El detector de esquinas de Harris mide estos cambios de intensidad y marca como esquinas los valores que sobrepasan un umbral. Este umbral puede ser modificado para obtener un mayor o menor número de esquinas en la imagen [5].



Figura 5 Ejemplo Detector de esquinas Harris.

3.10. Descomposición en valores singulares (SVD)

La descomposición en valores singulares es una de las más útiles descomposiciones matriciales. Su aplicación más frecuente es en la solución de sistemas de ecuaciones sobre determinados.

Dada una matriz cuadrada A , el SVD es una factorización de A como $A = UDV^T$, donde U y V son matrices ortogonales y D es una matriz diagonal con valores no negativos. La descomposición SVD también puede ser utilizada para matrices que no son cuadradas. Los valores diagonales de la matriz D son conocidos como los valores singulares de la matriz A . Estos valores no son los mismos que los valores propios. Los valores singulares de A son las raíces cuadradas de los valores propios de $A^T A$, U son los vectores propios de AA^T y V son los vectores propios de $A^T A$ [6].

3.11. Factorización de Cholesky

La factorización o descomposición de Cholesky toma su nombre del matemático André-Louis Cholesky, quien encontró que una matriz simétrica definida positiva puede ser descompuesta como el producto de una matriz triangular inferior y la traspuesta de la matriz triangular inferior. La matriz triangular inferior es el triángulo de Cholesky de la matriz original positiva definida. El resultado de Cholesky ha sido extendido a matrices con entradas complejas. Es una manera de resolver sistemas de ecuaciones matriciales y se deriva de la factorización LU con una pequeña variación.

Cualquier matriz cuadrada A con pivotes no nulos puede ser escrita como el producto de una matriz triangular inferior L y una matriz triangular superior U ; esto recibe el nombre de factorización LU. Sin embargo, si A es simétrica y definida positiva, se pueden escoger los factores tales que U es la traspuesta de L , y esto se llama la descomposición o factorización de Cholesky. Tanto la descomposición LU como la descomposición de Cholesky son usadas para resolver sistemas de ecuaciones lineales. Cuando es aplicable, la descomposición de Cholesky es dos veces más eficiente que la descomposición LU [13].

4. Descripción general y especificaciones

4.1. Descripción del lugar y del sistema de captura.

El sistema que compone el trabajo de grado para realizar la autocalibración es una cámara PTZ y un computador donde se ejecuta el código de autocalibración. Dada la flexibilidad del sistema, las escenas que se escogieron para realizar las pruebas fueron lugares donde los objetos permanecieran inmóviles y donde se pudiera evitar cualquier perturbación externa sobre la escena, para tal fin se escogieron diferentes lugares de una casa como se ilustra en la Figura 6. Se habla de objetos inmóviles porque en el proceso de autocalibración es necesario que los objetos de la escena no cambien de posición. Además de poder controlar muy bien las escenas, en la casa también se pudo tener una conexión fácil, cercana y directa de la cámara con el computador, esta conexión es necesaria para capturar las imágenes y realizar el control de la cámara.

El sistema de captura consta de la cámara PTZ conectada al computador, en el computador se procesan los videos de las escenas para extraer las imágenes necesitadas para calibrar la cámara. En la figura 6 se puede observar dos imágenes correspondientes a escenas utilizadas en la autocalibración, donde los objetos permanecen inmóviles mientras se toma la secuencia de imágenes.



Figura 6 Distintas escenas usadas para probar el algoritmo de autocalibración.

Se utilizaron imágenes obtenidas con la cámara Sony EVI-D70 que se ilustra en la Figura 7. El tamaño de las imágenes capturadas fue de 320 x 240 píxeles. Esta cámara permite realizar *zoom* variando la distancia focal entre 4.1 y 73.8 mm y dispone también de enfoque y apertura automáticos. Además cuenta con motores que permiten movimientos de azimut y elevación.

Los movimientos realizados para las pruebas fueron aleatorios y controlados con el software de Sony además de apoyarse también con el control remoto de la cámara. La cámara se puede controlar desde el computador a partir de su puerto de entrada RS-232C, el cual permite el control de la cámara a través del computador utilizando los comandos VISCA. Para la señal de video se utilizó la salida de S-Video y se conectó a un instrumento de captura de video en el computador. Este elemento captura el video con una resolución de 320x240 píxeles, el cual corresponde al tamaño de imagen utilizado.



Figura 7 Cámara PTZ EVI-D70 con sus respectivas entradas y salidas. Tomado de [11]

A continuación se explica la mecánica de la cámara.

4.1.1. Movimiento pan de la cámara PTZ.

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 la cámara es fija, y la posición del sistema de referencia del mundo que se va a considerar es fijo 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 8:

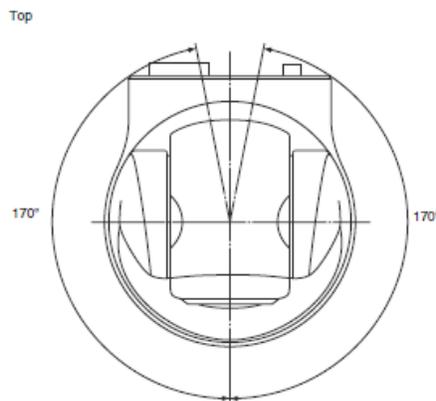


Figura 8 Movimiento Pan de la cámara. Tomado de [11]

La posición *pan* de la cámara puede ser obtenida mediante la casilla *Pan Position* del software de Sony, este parámetro está formado por un valor hexadecimal. El valor 0000 corresponde a la posición de referencia de la cámara, el valor F725 corresponde a una rotación horizontal de 170° hacia la derecha, y el valor 08DB corresponde a una rotación horizontal de 170° hacia la izquierda [11].

Para saber en qué posición (grados) se encuentra la cámara con respecto a su punto de referencia simplemente se encontró la relación entre el valor hexadecimal mostrado en el software de Sony y los grados a que corresponde ese valor. El resultado fue una relación lineal donde cada 0.075 grados equivalen a una unidad en base decimal. En consecuencia, para encontrar los grados a partir de un valor hexadecimal simplemente se pasa este valor a base decimal para luego multiplicarlo por 0.075.

4.1.2. Movimiento tilt de la cámara.

La rotación sobre el eje vertical de la cámara puede modificar el ángulo en un rango de 120 grados (30° hacia abajo y 90° hacia arriba) según indica la Figura 9. La posición *tilt* de la cámara también puede ser obtenida utilizando el Software de Sony. La posición se representa con un valor hexadecimal. 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 04B0 corresponde a una rotación vertical de 90° hacia arriba, y el valor FE70 corresponde a una rotación vertical de 30° hacia abajo [11]. Al igual que en el movimiento pan de la cámara la relación es de 0.075 grados equivalen a uno en base decimal.

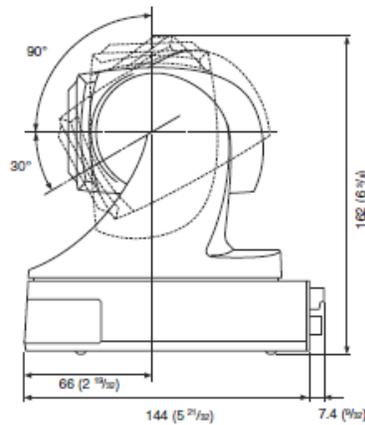


Figura 9 Movimiento Vertical (Tilt) de la cámara. Tomado de [11]

A continuación se muestra la tabla con los valores para pan y tilt en hexadecimal.

Pan/Tilt Position (Reference values)

		Pan Position Data	Tilt Position Data (Image Flip: OFF)	Tilt Position Data (Image Flip: ON)
Angle	+170°	08DB	–	–
	+150°	07DB	–	–
	+90°	04B0	04B0	–
	+30°	0190	0190	0190
	+10°	0085	0085	0085
	+0.075°	0001	0001	0001
	0°	0000	0000	0000
	-0.075°	FFFF	FFFF	FFFF
	-10°	FF7B	FF7B	FF7B
	-30°	FE70	FE70	FE70
	-90°	FB50	–	FB50
	-150°	F830	–	–
	-170°	F725	–	–

A + indicates a pan to the right and a tilt upward on the monitor screen.
 A – indicates a pan to the left and a tilt downward on the monitor screen.

Figura 10 Valores Hexadecimales de la posición de la cámara. Tomado de [11]

4.2. Software EVI Series

Una parte muy importante de este proyecto de grado es el control de la cámara que se va a utilizar. El algoritmo de autocalibración tiene como entrada un grupo de imágenes con distintas orientaciones, para capturar estas imágenes es necesario controlar de una manera eficaz la cámara, esto es posible con el EVI Series Software de Sony.

Este software fue desarrollado por Sony para el control de este tipo de cámaras. El control de este tipo de cámaras se basa en el lenguaje VISCA y el propósito principal es el control de dispositivos en una única red. Mediante el protocolo de VISCA, hasta siete dispositivos pueden compartir la misma red y pueden tratarse individualmente. VISCA es flexible y puede ampliarse para una amplia variedad de equipos como monitores, cámaras de vídeo. El protocolo es abierto y no requiere licencias.

El programa utilizado es un demo ofrecido por Sony donde las funciones principales y las necesitadas en este trabajo están activas y en consecuencia pueden ser utilizadas. El programa tiene una interfaz gráfica donde se puede controlar la cámara de una manera fácil y se pueden variar los parámetros sin

complicaciones utilizando los comandos respectivos. Para fines del trabajo de grado se utilizó el control gráfico de la cámara, la información de la posición de la cámara y también la posición de zoom.

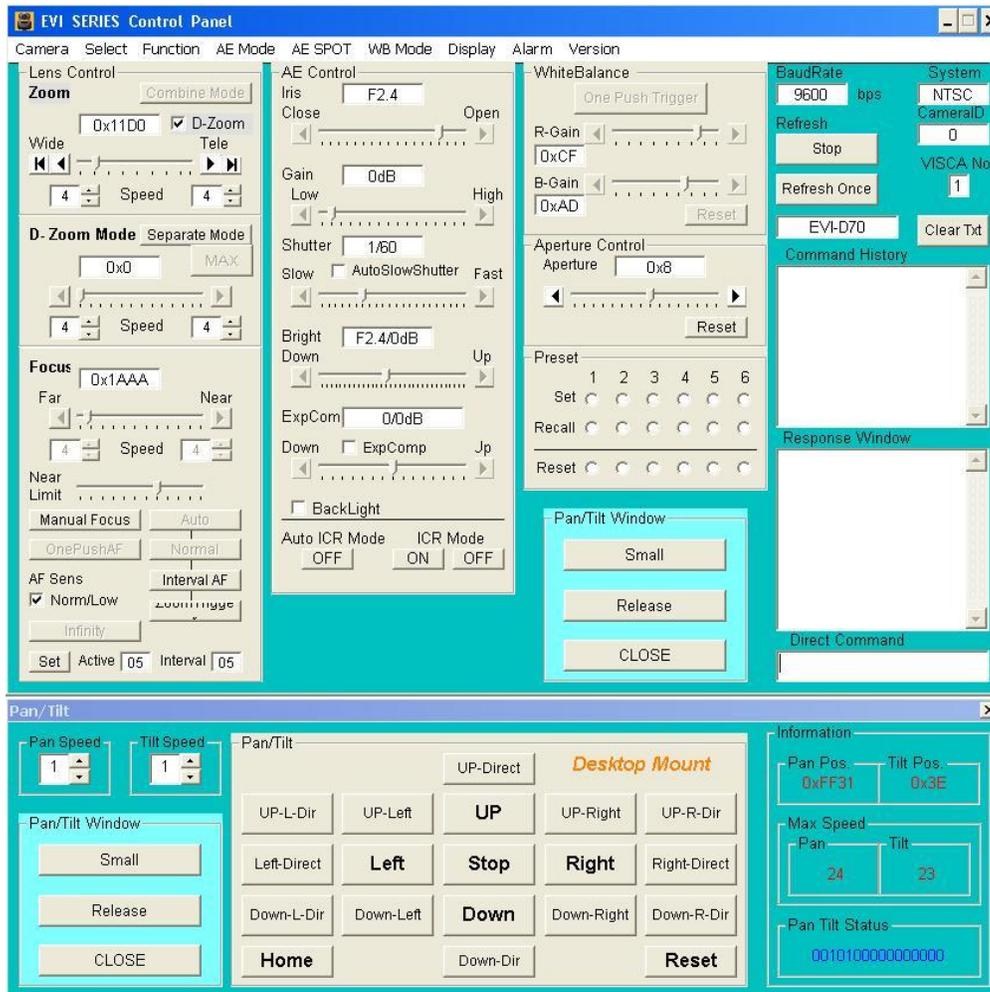


Figura 11 Interfaz gráfica del Software de Sony.

4.3. Diagrama de bloques

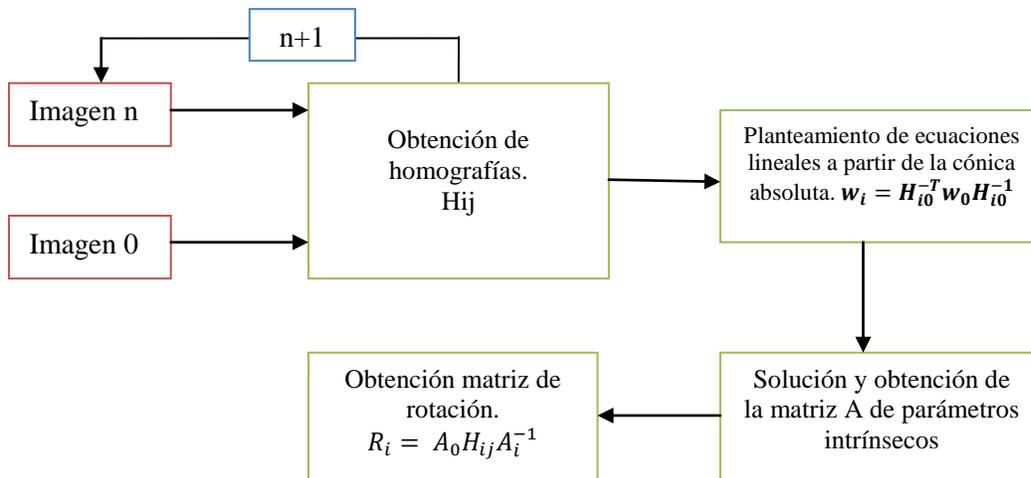


Figura 12 Diagrama de Bloques.

El diagrama de bloques muestra de una manera general cómo funciona el algoritmo de autocalibración. El algoritmo tiene como entradas la imagen de referencia I_0 y la n ésima imagen tomada de la secuencia. Sus salidas son la matriz de calibración que contiene los parámetros intrínsecos \mathbf{A} y la matriz de rotación \mathbf{R} . La matriz de calibración es obtenida para cada homografía obtenida entre la imagen de referencia y la n ésima imagen, por lo que se tendrán N matrices de calibración. Cada una de estas matrices da la información de los parámetros internos de la n ésima imagen. Por esto mismo se tendrá un igual número de matrices de Rotación donde se tiene la información de la rotación ocurrida entre la n ésima imagen y la imagen de referencia. Como el barrido se realiza sin variar los parámetros internos de la cámara, se sabe entonces que la matriz \mathbf{A} no cambiará, en consecuencia se hará un promedio con los valores obtenidos de todas las matrices de calibración \mathbf{A} para obtener una matriz de calibración generalizada para el barrido de la escena realizado. La matriz de rotación si varía por lo que, como se dijo anteriormente, cada matriz de rotación \mathbf{R} dará el movimiento entre esa imagen y la de referencia.

4.3.1. Obtención de homografías H_{ij} .

Esta sección describe el bloque para encontrar automáticamente la homografía entre dos imágenes. Las entradas de este algoritmo son la imagen I_0 y la n ésima imagen, la salida es la homografía estimada entre ese par de imágenes. Una de las imágenes de entrada es la que se toma como referencia I_0 (generalmente es la primera imagen de la secuencia) y se mantiene fija mientras que la otra varía pasando por las n imágenes.

El primer paso de este proceso es computar los puntos de interés de cada imagen a partir del algoritmo detector de esquinas de Harris. Por esquinas no se hace referencia a la esquina física de la imagen sino que las esquinas son puntos que tienen una variación de intensidad en todas las direcciones, esta variación se obtiene luego de medirse los cambios sobre este punto cuando deslizamos una ventana alrededor del mismo. Luego de tener estos puntos característicos de cada imagen se encontrará la correspondencia entre puntos de las imágenes.

Para cada esquina o punto localizado en (x, y) en la imagen I_0 se busca la correspondencia en la imagen n a partir del mayor valor de correlación cruzada encontrado en la región de búsqueda centrada en (x, y) . Igualmente para cada esquina o punto en la imagen n se buscará una correspondencia en la imagen I_0 . Con estas correspondencias se forma un grupo de puntos que serán la entrada para el algoritmo de RANSAC. Por último se implementa RANSAC obteniendo como salida la homografía H_{ij} .

4.3.2. Planteamiento de ecuaciones.

En el bloque de planteamiento de ecuaciones se utilizan las homografías encontradas en el bloque anterior para plantear el sistema de ecuaciones lineales. La ecuación que permite relacionar las homografías con la matriz de calibración es la siguiente $\mathbf{w}_i = \mathbf{H}_{i0}^{-T} \mathbf{w}_0 \mathbf{H}_{i0}^{-1}$ donde \mathbf{w} es la cónica absoluta y esta misma se relaciona con la matriz de calibración de la siguiente manera $\mathbf{w}_i = (\mathbf{A}_i^{-1})^T \cdot \mathbf{A}_i^{-1}$. La salida es un sistema de ecuaciones lineales sobre determinado. Estas ecuaciones son explicadas en forma completa en el desarrollo teórico, en donde se explica su origen y cómo son utilizadas para obtener los resultados deseados.

4.3.3. Solución y obtención de la matriz \mathbf{A} .

Para la solución de las ecuaciones lineales se utiliza el *Singular Value Decomposition* dado que es un método que se utiliza frecuentemente para la solución de sistemas sobre determinados, este es el caso del sistema de ecuaciones planteado en el bloque anterior. Con el sistema de ecuaciones resuelto se obtiene la cónica absoluta, esta cónica absoluta se factoriza utilizando la factorización de Cholesky y así se encuentra la matriz de calibración \mathbf{A} .

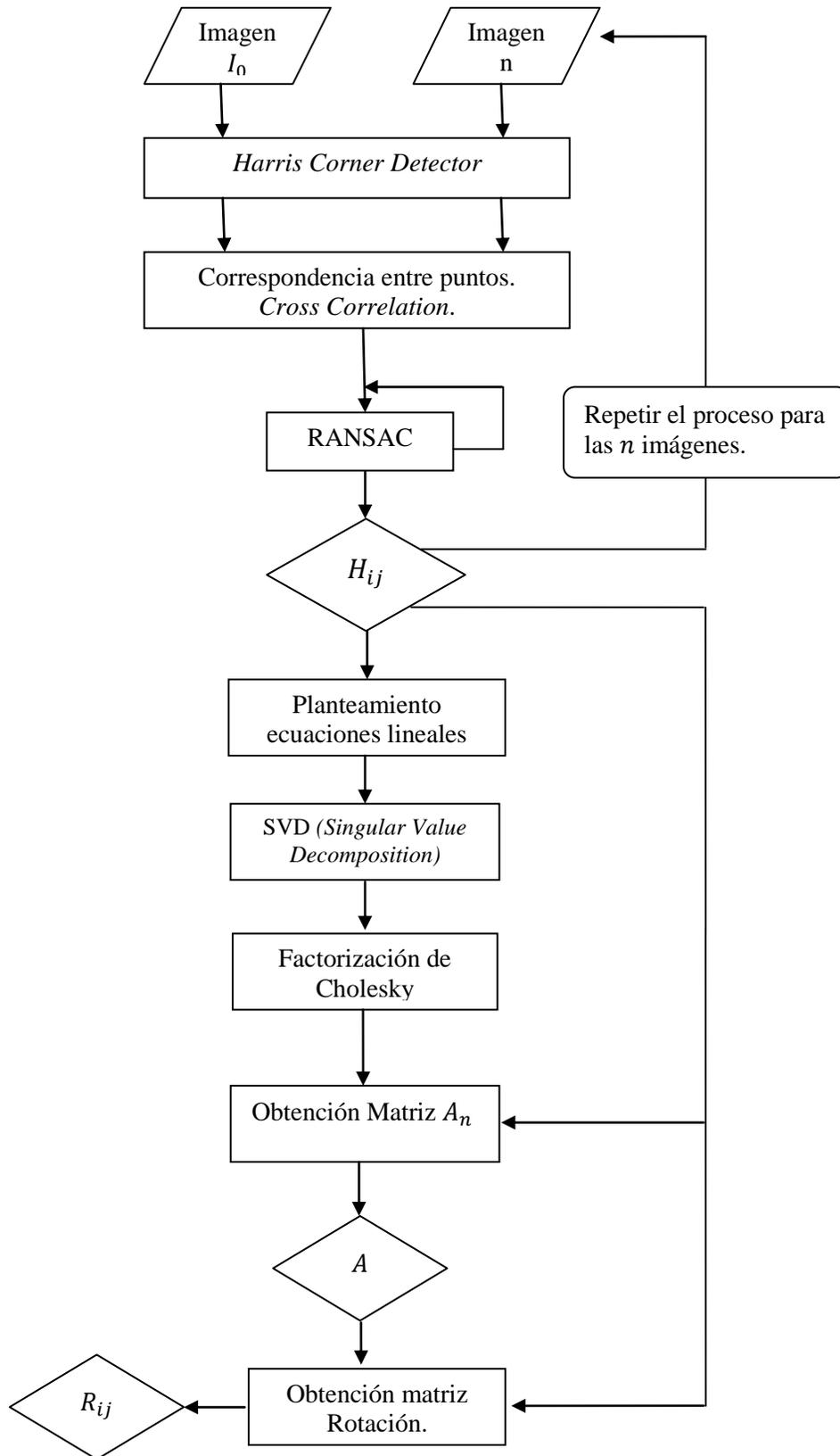
4.3.4. Obtención de la matriz de Rotación.

Para obtener la matriz de rotación de la cámara se resuelve la ecuación $R_i = A_0 H_{ij} A_i^{-1}$, donde la matriz de parámetros intrínsecos para la imagen de referencia es A_0 y A_i es la matriz de parámetros intrínsecos de la i ésima imagen. las homografías H_{ij} son obtenidas de los bloques anteriores.

Con este último bloque queda realizado todo el proceso de autocalibración. Como resultado se tiene una matriz promedio de calibración \mathbf{A} que contiene todos los parámetros intrínsecos de la cámara y las matrices de rotación \mathbf{R}_{ij} que contienen los parámetros extrínsecos. Para evaluar los resultados obtenidos se utiliza un *toolbox* para Matlab que implementa un método clásico de calibración para encontrar los parámetros internos de la cámara. En el punto 4.5 se explica el funcionamiento del Toolbox, mientras que en la sección 5.1.1 se explican las bases de este método.

A continuación se muestra el diagrama de flujo de este algoritmo, sus entradas son las imágenes tomadas del barrido de la escena y sus salidas son la matriz de calibración y las respectivas matrices de rotación.

4.4. Diagrama de flujo



4.5. Implementación del método de Zhang. Toolbox de calibración.

Con los resultados obtenidos del proceso de autocalibración de Agapito se prosigue a realizar una comparación entre los resultados de la autocalibración y los resultados obtenidos del toolbox. Este toolbox de calibración es desarrollado por Jean-Yves Bouguet y se basa en el método de calibración clásico de Zhang, las bases de este método se explican en la sección 5.1.1. A continuación se explica su funcionamiento:

Para realizar la calibración de la cámara utilizando este toolbox el usuario debe conocer e introducir la siguiente información:

- Número de imágenes: el número de imágenes que se van a considerar para la calibración, estas imágenes contienen el patrón o plantilla de calibración en distintas posiciones.
- Número de cuadrados por lado: el número de cuadrados que componen cada lado del plano seleccionado en el patrón.
- Longitud del lado de los cuadrados (expresada en mm) que componen la plantilla utilizada.
- Nombre con el que se han guardado las imágenes: el programa muestra cada imagen j al usuario y le pide que marque los puntos correspondientes a las esquinas del plano.

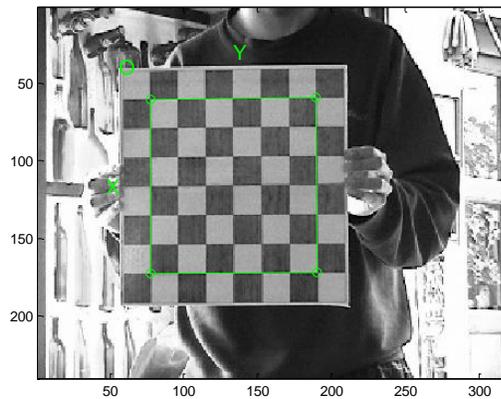


Figura 14 Puntos que hay que marcar (las esquinas del tablero)

Con esta información se calculan todos los datos de entrada necesarios para la ejecución del método de Zhang. Los datos de salida son los siguientes:

- **Distancia Focal:** La distancia focal dada en píxeles y está guardada en la variable **fc**.
- **Punto Principal:** Las coordenadas del punto principal son guardadas en el vector 2×1 **cc**.
- **El Skew coefficient:** El *skew coefficient* que define el ángulo entre los píxeles x y y son guardados en **alpha_c**.
- **Distorsiones:** Los coeficientes de distorsión de la imagen (radial y tangencial) son almacenados en el vector 5×1 **kc**.

La matriz de los parámetros intrínsecos de la cámara en este toolbox está compuesta de la siguiente manera utilizando los datos de salida anteriores:

$$KK = \begin{bmatrix} \mathbf{fc}(1) & \alpha_c * \mathbf{fc}(1) & \mathbf{cc}(1) \\ 0 & \mathbf{fc}(2) & \mathbf{cc}(2) \\ 0 & 0 & 1 \end{bmatrix}$$

5. Desarrollo teórico

5.1. Métodos de calibración utilizados.

Se utilizaron dos tipos de calibración en este proyecto, la calibración clásica y la autocalibración. Los métodos de calibración clásicos parten de una escena conocida en el cual 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. Existen varios toolbox que implementan la calibración clásica de 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. 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 [12].

Por otro lado, el método de autocalibración de Agapito es un método que permite la calibración automática de cámaras cuyo movimiento se limite a rotaciones, aunque pueden variar los parámetros intrínsecos durante la adquisición de las imágenes necesarias para realizar este proceso. El método es aplicable bajo una serie de restricciones: se supone que los ejes del plano imagen son perfectamente ortogonales, y que las imágenes no están afectadas por distorsión. Además, en el caso de que solo exista un movimiento de *Pan-tilt*, o sea, que no exista ciclotorsión (*roll*), debe añadirse, al menos, una restricción más. Este es el caso del sistema planteado en este trabajo de grado donde sólo se realizan movimientos *Pan* y *Tilt*, en consecuencia es necesaria la suposición de que los píxeles sean cuadrados, o que al menos se conozca la relación entre sus lados [2]. Generalmente las cámaras tienen píxeles cuadrados por lo que esta restricción no afecta la flexibilidad del sistema propuesto en este trabajo de grado.

5.1.1. Calibración clásica: método de Zhang.

Zhang propone una técnica de calibración basada en la observación de una plantilla plana como la de la Figura 15 desde varias posiciones. El usuario tiene la libertad de elegir si mueve la posición de la cámara o la superficie de calibración, estos movimientos no tienen que ser conocidos.



Figura 15 Plantilla utilizada para el método clásico

Como se dijo anteriormente, la ventaja de este método de calibración con respecto al resto de los métodos clásicos es que permite obtener los parámetros de la cámara fácilmente a partir de una plantilla plana sin necesidad de conocer la posición de los puntos 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 el método sea muy flexible aunque utilice un patrón de calibración.

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.

5.1.2. Autocalibración: método lineal de Agapito

El objetivo del método lineal de Agapito es hallar los parámetros intrínsecos y extrínsecos de la cámara. A diferencia del método de Zhang, lo único necesario para realizar la calibración es disponer de una serie de imágenes que capturen una misma escena y que hayan sido tomadas con variaciones de rotación (movimientos *Pan-Tilt*). Como en toda técnica de autocalibración, no es necesario conocer ninguna información sobre la escena [12].

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 [2]. 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.

Se establecen las siguientes restricciones sobre las imágenes a trabajar:

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 $f_x = f_y$

Los pasos que hay que realizar son los siguientes:

5.1.2.1. Obtención de la homografía (teoría)

El método de Agapito requiere la obtención de la homografía de cada imagen con respecto a una que se toma como referencia. En la i ésima imagen un punto 2D, \mathbf{m} , corresponde a un punto 3D \mathbf{M} de la escena a partir de la ecuación

$$\mathbf{M} = P_i^{-1} \cdot \mathbf{m} \quad (1.8)$$

donde P_i es la matriz de proyección del modelo *pinhole* en la i ésima escena definida en (1.7). El mismo punto \mathbf{M} de la escena corresponde a otro punto 2D de la j ésima imagen, \mathbf{m}' mediante la matriz de proyección P_j como

$$\mathbf{m}' = P_j \cdot \mathbf{M} \quad (1.9)$$

donde P_j es la matriz de proyección del modelo *pinhole* en la j ésima escena definida en (1.7). Por lo tanto, combinando las ecuaciones (1.8) y (1.9) se puede determinar que los puntos \mathbf{m} y \mathbf{m}' (correspondientes al mismo punto 3D de la escena) están relacionados de la siguiente forma:

$$\mathbf{m}' = P_j \cdot P_i^{-1} \cdot \mathbf{m} \quad (1.10)$$

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

$$H_{ij} = P_j \cdot P_i^{-1} \quad (1.11)$$

Para calcular la homografía utilizando (1.11) es necesario conocer los parámetros intrínsecos y extrínsecos de la cámara. Estos valores de los parámetros no se conocen y en consecuencia hay que encontrar la homografía con otro método. En [3] se indica un método práctico para obtener la homografía a partir de las coordenadas de cuatro puntos de cada imagen. A continuación se hace un resumen del algoritmo para encontrar la homografía a partir de la correspondencia de n puntos $\mathbf{x}_i' \leftrightarrow \mathbf{x}_i$. Donde $\mathbf{x}_i' = (x_i', y_i', w_i')$ y $\mathbf{x}_i = (x_i, y_i, w_i)$ son coordenadas homogéneas.

Lo primero que hay que realizar para cada correspondencia de puntos es computar la matriz N_i la cual viene dada como:

$$N_i = \begin{bmatrix} \mathbf{0}^T & -w'_i \mathbf{x}'_i{}^T & y'_i \mathbf{x}'_i{}^T \\ w'_i \mathbf{x}'_i{}^T & \mathbf{0}^T & -x'_i \mathbf{x}'_i{}^T \\ -y'_i \mathbf{x}'_i{}^T & x'_i \mathbf{x}'_i{}^T & \mathbf{0}^T \end{bmatrix}$$

De esta matriz sólo se utilizan las dos primeras filas, la razón es porque al plantear el sistema de ecuaciones las primeras dos filas forman ecuaciones linealmente independientes. El sistema de ecuaciones queda planteado de la siguiente manera para cada par de puntos:

$$\underbrace{\begin{bmatrix} \mathbf{0}^T & -w'_i \mathbf{x}'_i{}^T & y'_i \mathbf{x}'_i{}^T \\ w'_i \mathbf{x}'_i{}^T & \mathbf{0}^T & -x'_i \mathbf{x}'_i{}^T \end{bmatrix}}_{N_i} \begin{pmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{pmatrix} = 0 \quad (1.12)$$

donde

$$\mathbf{h} = \begin{pmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{pmatrix}, H = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}$$

siendo \mathbf{h} un vector de longitud 9 conformado por las componentes de H , H es la homografía correspondiente entre un par de imágenes. Utilizando (1.12) se tiene una matriz de tamaño 2×9 y con n puntos se ensambla una matriz N de tamaño $2n \times 9$. Con este sistema de ecuaciones se obtiene una solución a partir de SVD. Utilizando esta solución se arma la matriz de la homografía H .

5.1.2.2. Obtención de la homografía (implementación)

El objetivo de esta etapa es encontrar la homografía entre dos imágenes usando un algoritmo robusto como lo es RANSAC. Este algoritmo puede ser aplicado por ejemplo a dos imágenes que han sido adquiridas por una cámara que ha rotado alrededor de su centro, este es el caso de las cámaras PTZ

La entrada de esta etapa son las dos imágenes que se van a utilizar para crear la homografía. Una de estas imágenes es la que se usa de referencia (I_0) para todo el proceso mientras que la otra imagen (I_i) varía, I_i varía para las n imágenes obtenidas en el recorrido. La salida es la homografía estimada entre cada par de imágenes, en total se tendrán n homografías para las n imágenes utilizadas.

El primer paso para hallar la homografía es computar los puntos de interés de cada imagen (I_0, I_i), estos puntos de interés se encuentran utilizando el detector de esquinas de Harris. Este algoritmo es bastante popular debido a que trabaja muy bien bajo las siguientes variables: rotación, variación de iluminación y el ruido en la imagen [5]. El detector de esquinas está basado en la función de auto-correlación de una señal; donde esta función de auto correlación mide los cambios locales de la señal aplicando un barrido con una ventana alrededor del punto de interés. En la Figura 16 se muestra un ejemplo de un barrido donde se encuentran cambios en ambas direcciones.

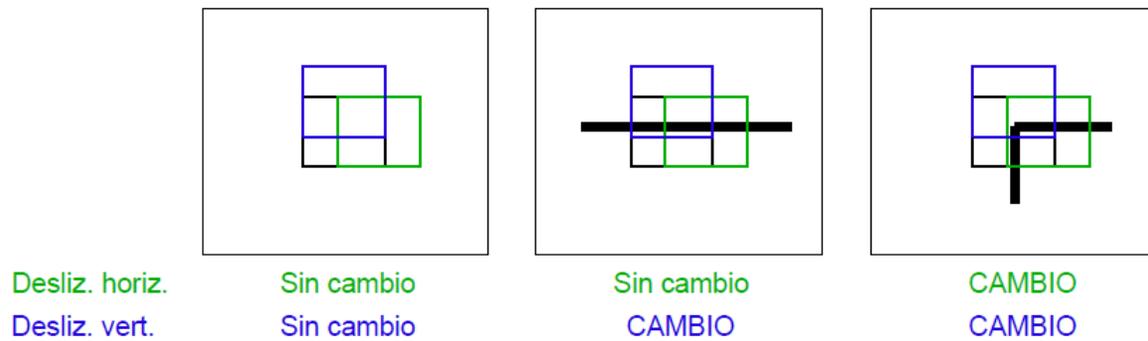


Figura 16 Barrido y cambios sobre un punto de interés.

En la Figura 17 se presentan los resultados del algoritmo de Harris sobre un par de imágenes (I_0, I_i) utilizadas en el desarrollo del proyecto.

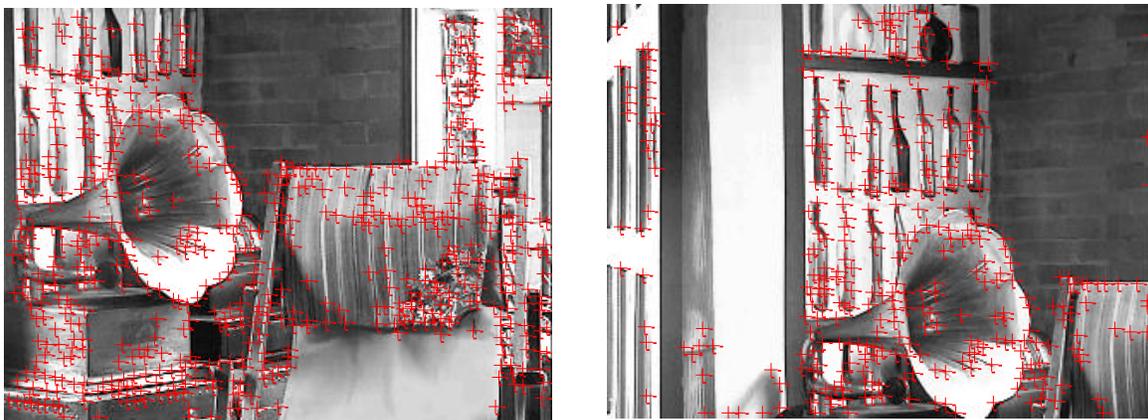


Figura 17 El algoritmo detector de esquinas de Harris implementado para I_0, I_j

El grupo de puntos detectados para I_0 y para I_i se utiliza como entrada para encontrar las correspondencias de puntos entre cada par de imágenes. Esta correspondencia se realiza por medio de la correlación cruzada. Para cada punto s_i ubicado en las coordenadas (x, y) de la imagen I_0 se encontrará un punto en la imagen I_i que tenga el valor más alto de correlación cruzada. La búsqueda de este valor se centra en la vecindad de la coordenada (x, y) del punto s_i . El mismo proceso se realiza para los puntos de la imagen I_i , en caso de haber más de un punto en la correspondencia se elige el que tenga el valor más alto de correlación cruzada.

RANSAC es aplicado utilizando el grupo de puntos considerados como correspondencias que fueron encontrados en el paso anterior, como salida se obtiene el mejor modelo de homografía para las imágenes I_0 y I_i . La idea general de RANSAC es encajar un modelo sobre un grupo de puntos a partir de varias iteraciones. En resumen el algoritmo RANSAC hace lo siguiente:

1. Se selecciona aleatoriamente una muestra de s puntos a partir de S (grupo de correspondencias entre puntos) y se calcula el modelo para este subconjunto.
2. Se determina el conjunto de puntos S_i que están dentro de una distancia umbral t del modelo. El conjunto S_i es el consenso de la muestra y define los *inliers* de S .
3. Si el tamaño de S_i (número de inliers) es más grande que un umbral T , se recalcula el modelo utilizando todos sus puntos y finaliza el algoritmo.
4. Si el tamaño de S_i es más pequeño que el umbral T , se selecciona un nuevo subconjunto y se repiten los pasos anteriores.

- Después de N intentos el consenso S_i mayor se selecciona y el modelo se recalcula utilizando todos los puntos en el subconjunto S_i .

Este algoritmo está totalmente explicado en [6].

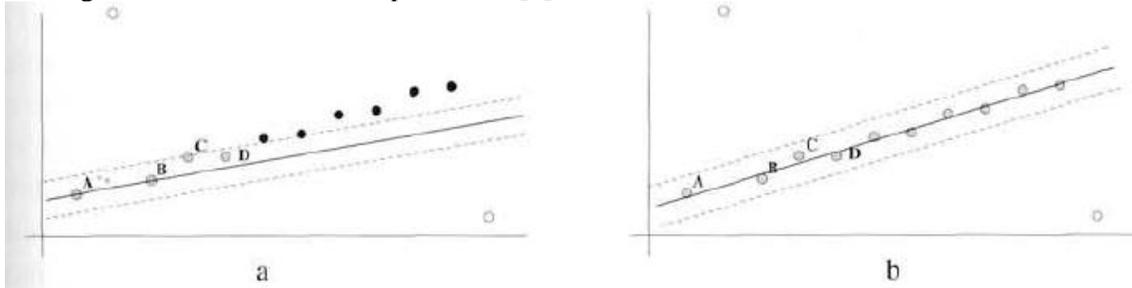


Figura 18 Ejemplo de ajuste de un modelo sobre un conjunto de puntos. En la figura a solo cuatro puntos son considerados como *inliers*, mientras que en la figura B la mayoría de puntos son *inliers*. Las líneas punteadas son el *threshold* y la distancia d es la distancia que hay entre la línea negreada hasta al punto. Tomado de [6]

Después de la aplicación de RANSAC finaliza el bloque de obtención de la homografía entre dos imágenes. La salida de este bloque es la homografía correspondiente entre las dos imágenes de la entrada (I_0, I_i) .

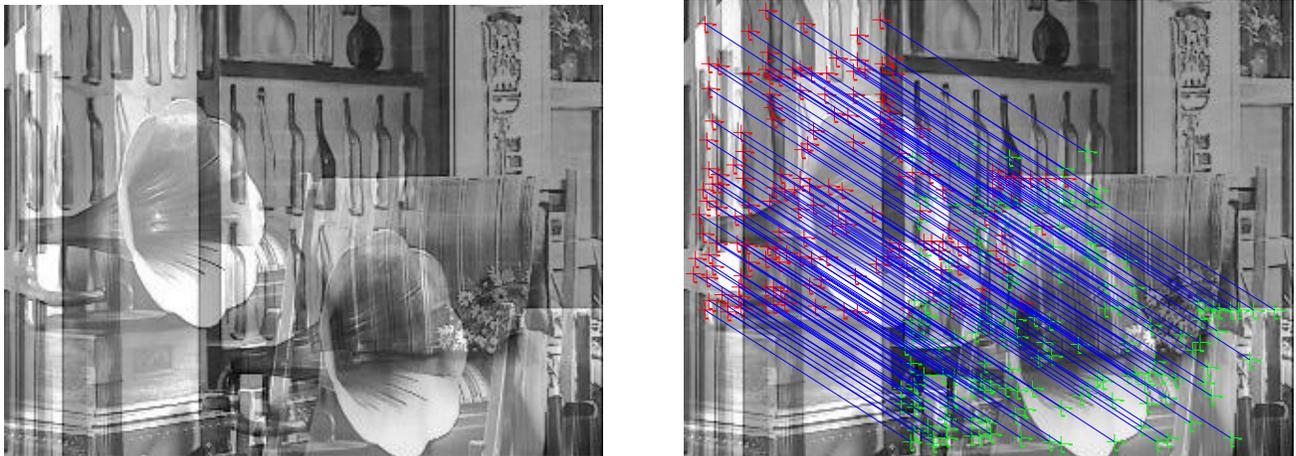


Figura 19 Resultado de RANSAC, donde se superponen las imágenes y se grafican los puntos denominados *inliers*

El proceso anterior se repite para las n imágenes obtenidas y así al final se tiene un conjunto de n homografías. Se continúa con el planteamiento del sistema de ecuaciones lineales. Este planteamiento se basa en la definición de cónica absoluta.

5.1.2.3. Obtención de la cónica absoluta.

La cónica absoluta, w , se define a partir de la matriz de parámetros intrínsecos, A (definida en la ecuación 1.5):

$$w = (A^{-1})^T \cdot A^{-1} \quad (1.13)$$

Para la j ésima imagen se tiene $\mathbf{w}_j = (\mathbf{A}_j^{-1})^T \cdot \mathbf{A}_j^{-1}$ y para la i ésima imagen se tiene $\mathbf{w}_i = (\mathbf{A}_i^{-1})^T \cdot \mathbf{A}_i^{-1}$.

Si se descompone la matriz de proyección \mathbf{P} 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 (1.11) puede escribirse como

$$\mathbf{H}_{ij} = \mathbf{A}_j \mathbf{R}_j \mathbf{R}_i^{-1} \mathbf{A}_i^{-1} = \mathbf{A}_j \mathbf{R}_{ij} \mathbf{A}_i^{-1} \quad (1.14)$$

donde \mathbf{R}_{ij} es la rotación de j con respecto a i (según se ha visto en el estudio del método en [6]).

Como $\mathbf{R}_{ij} \mathbf{R}_{ij}^T = \mathbf{I}$, la ecuación (1.14) queda:

$$\mathbf{H}_{ij} \mathbf{w}_i \mathbf{H}_{ij}^T = \mathbf{w}_j \quad (1.15)$$

Para expresar la ecuación (1.15) en función de la cónica absoluta de cada imagen se aplica la inversa a la ecuación (1.14) y se obtiene:

$$\mathbf{H}_{ij}^{-T} \mathbf{w}_i \mathbf{H}_{ij}^{-1} = \mathbf{w}_j \quad (1.16)$$

Por lo que se obtiene una ecuación que permite relacionar la cónica de la imagen i con la cónica de la imagen j . A partir de la ecuación (1.16) y si se toma una de las imágenes como referencia (cuya cónica es \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:

$$\mathbf{w}_i = \mathbf{H}_{i0}^{-T} \mathbf{w}_0 \mathbf{H}_{i0}^{-1} \quad (1.17)$$

La cónica absoluta expresada en función de los parámetros intrínsecos es:

$$\mathbf{w} = \begin{bmatrix} \frac{1}{\alpha^2} & 0 & -\frac{P_x}{\alpha^2} \\ 0 & \frac{1}{\alpha^2} & -\frac{P_y}{\alpha^2} \\ -\frac{P_x}{\alpha^2} & -\frac{P_y}{\alpha^2} & 1 + \frac{P_x^2 + P_y^2}{\alpha^2} \end{bmatrix}$$

Utilizando la equivalencia de la cónica absoluta y con las restricciones de cero skew y pixeles cuadrados nombradas en 5.1.2 se pueden deducir dos ecuaciones para cada par de imágenes:

$$\mathbf{w}(1,2) = 0 \quad \mathbf{w}(1,1) = \mathbf{w}(2,2) \quad (1.18)$$

Utilizando la ecuación (1.17) y las ecuaciones (1.18) se plantea el sistema de ecuaciones lineales con el objetivo de encontrar la cónica absoluta \mathbf{w}_i . Los datos conocidos que se tienen de la ecuación (1.17) son \mathbf{H}_{i0}^{-T} y \mathbf{H}_{i0}^{-1} , el planteamiento de las ecuaciones se explica a continuación.

Si se escribe la ecuación (1.17) en forma matricial y se parametrizan los elementos de las cónicas absolutas $\mathbf{w}_i, \mathbf{w}_0$ en seis variables ya que los elementos de la diagonal superior son iguales a los de la diagonal inferior se tiene que:

$$\underbrace{\begin{bmatrix} a_{i1} & a_{i2} & a_{i3} \\ a_{i2} & a_{i4} & a_{i5} \\ a_{i3} & a_{i5} & a_{i6} \end{bmatrix}}_{w_i} = \underbrace{\begin{bmatrix} b_1 & b_4 & b_7 \\ b_2 & b_5 & b_8 \\ b_3 & b_6 & b_9 \end{bmatrix}}_{H_{i0}^{-T}} \underbrace{\begin{bmatrix} a_1 & a_2 & a_3 \\ a_2 & a_4 & a_5 \\ a_3 & a_5 & a_6 \end{bmatrix}}_{w_0} \underbrace{\begin{bmatrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_7 & b_8 & b_9 \end{bmatrix}}_{H_{i0}^{-1}} \quad (1.19)$$

Resolviendo (1.19) se destacan las siguientes tres ecuaciones, se destacan porque estas son las que se van a utilizar para formar el sistema de ecuaciones aplicando (1.18).

$$a_{i1} = b_1(a_1b_1 + a_2b_4 + a_3b_7) + b_4(a_2b_1 + a_4b_4 + a_5b_7) + b_7(a_3b_1 + a_5b_4 + a_6b_7) \quad (1.20)$$

$$a_{i2} = b_1(a_1b_2 + a_2b_5 + a_3b_8) + b_4(a_2b_2 + a_4b_5 + a_5b_8) + b_7(a_3b_2 + a_5b_5 + a_6b_8) \quad (1.21)$$

$$a_{i4} = b_2(a_1b_2 + a_2b_5 + a_3b_8) + b_5(a_2b_2 + a_4b_5 + a_5b_8) + b_8(a_3b_2 + a_5b_5 + a_6b_8) \quad (1.22)$$

Con las igualdades de (1.18) y utilizando (1.20) (1.21) (1.22) se forma un sistema de ecuaciones de la forma $\mathbf{E} \cdot \mathbf{a} = \mathbf{0}$, donde $\mathbf{a} = (a_1, a_2, \dots, a_6)^T$ y cada fila de \mathbf{E} corresponde a una ecuación obtenida de (1.18). Estas ecuaciones quedarán de la siguiente manera para cada homografía:

$$\frac{a_{i1}}{w_{i(1,1)}} = \frac{a_{i4}}{w_{i(2,2)}} \quad (1.22)$$

$$\frac{a_{i2}}{w_{i(1,2)}} = 0 \quad (1.23)$$

Reemplazando los elementos de (1.22) por (1.20) y (1.21) e igualando (1.22) a cero se tiene que:

$$\begin{aligned} & b_1(a_1b_1 + a_2b_4 + a_3b_7) + b_4(a_2b_1 + a_4b_4 + a_5b_7) + b_7(a_3b_1 + a_5b_4 + a_6b_7) \\ & = b_2(a_1b_2 + a_2b_5 + a_3b_8) + b_5(a_2b_2 + a_4b_5 + a_5b_8) + b_8(a_3b_2 + a_5b_5 + a_6b_8) \end{aligned}$$

$$b_1(a_1b_2 + a_2b_5 + a_3b_8) + b_4(a_2b_2 + a_4b_5 + a_5b_8) + b_7(a_3b_2 + a_5b_5 + a_6b_8) = 0$$

Organizando y reescribiendo matricialmente estas ecuaciones quedan de la siguiente manera:

$$\underbrace{\begin{bmatrix} b_1^2 - b_2^2 & 2b_1b_4 - 2b_2b_5 & 2b_1b_7 - 2b_2b_8 & b_4^2 - b_5^2 & 2b_4b_7 - 2b_5b_8 & b_7^2 - b_8^2 \\ b_1b_2 & b_1b_5 + b_2b_4 & b_1b_8 + b_2b_7 & b_4b_5 & b_4b_8 + b_5b_7 & b_7b_8 \end{bmatrix}}_{\mathbf{E}} \cdot \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix}}_{\mathbf{a}} = 0 \quad (1.25)$$

Para cada par de imágenes se generan dos ecuaciones lineales iguales que (1.25). Para resolverlo, basta con encontrar el vector \mathbf{a} que minimice $\|\mathbf{Ea}\|$ tal que $\|\mathbf{a}\| = 1$. Esto se consigue descomponiendo en valores singulares la matriz \mathbf{E} y tomando como solución el autovector correspondiente al autovalor más pequeño.

Para extraer los parámetros intrínsecos de cada imagen, matriz A_i , a partir de w_i se aplica la factorización de Cholesky sobre w_i y se invierte el resultado, esto basado en la ecuación (1.13). Hay que tener en cuenta que para poder realizar esta operación, la matriz que contiene la cónica absoluta de cada imagen debe ser

definida positiva [6]. Por último, los parámetros extrínsecos, es decir, la matriz de rotación que relaciona cada imagen respecto a la utilizada como referencia, se pueden obtener de la siguiente ecuación:

$$R_i = A_0 H_{ij} A_i^{-1}$$

5.2. Estimación de distancia.

Una de las aplicaciones utilizando la información de los parámetros intrínsecos y extrínsecos de la cámara es la extracción de información métrica de la escena. Para lograr este objetivo se partió de la información que se tiene del ángulo recorrido por la cámara entre dos vistas, esta información se obtiene de la matriz de rotación entre dos imágenes. En la Figura 20 que se muestra a continuación se observa cómo se puede relacionar el ángulo recorrido por la cámara con ciertas distancias de la escena, la ecuación resultante es $\tan(\theta) = \frac{y}{x}$ (1.26). Hay que aclarar que se necesita tener previamente alguna información métrica de la escena, ya sea la distancia x o la distancia y .

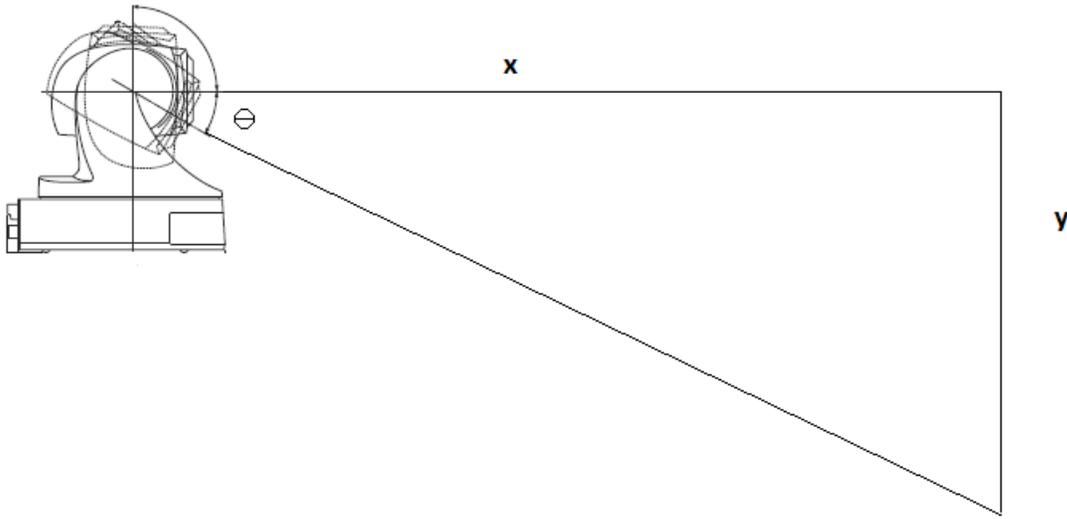


Figura 20 Relación de distancias con el ángulo recorrido por la cámara.

Utilizando la trigonometría se puede obtener la distancia x o la distancia y , esto depende de la información que se tenga previamente. Si no se tiene información métrica alguna, no se puede realizar ningún tipo de estimación de distancia en la escena.

6. Resultados algoritmo de autocalibración.

6.1. Protocolos de pruebas.

Para probar el desempeño del algoritmo de autocalibración se tomaron varios videos con distintas configuraciones de la cámara PTZ, la idea general de las pruebas es observar y comparar el comportamiento del algoritmo frente a distintos niveles de zoom y mirar que variables pueden afectar sus resultados. Los videos se grabaron con el mayor nivel posible de luz y con el enfoque automático prendido.

En el primer grupo de videos se analizan los resultados obtenidos de los parámetros intrínsecos de la cámara para distintos valores de zoom. Los valores de rotación obtenidos de la cámara serán mostrados pero no comparados con los datos reales. El objetivo inicial es observar los resultados de los parámetros intrínsecos. En cada video utilizado se muestra el barrido de la escena realizado por la cámara, sin embargo no se muestran todas las imágenes si no que se muestran las imágenes donde la cámara cambia de dirección durante el barrido. Luego se grafican los resultados obtenidos para los valores intrínsecos de distancia focal y punto principal además de mostrar el recorrido estimado por el algoritmo. Estos resultados son comparados con los valores obtenidos por el algoritmo de calibración clásico desarrollado por Zhang e implementado en el toolbox de Jean Ives Bouquet. La comparación de parámetros se realiza gráficamente y porcentualmente.

En el siguiente grupo de videos se analizan los parámetros extrínsecos de la cámara. Los resultados se comparan con la información de la posición de la cámara registrada por el software de Sony. Para finalizar se implementa la aplicación de estimación de distancia y se evalúa su eficacia comparando las medidas reales con las medidas estimadas.

Como parte de las pruebas se hace un análisis de cómo el número de imágenes utilizadas en el método afectan los resultados.

6.2. Parámetros intrínsecos.

6.2.1. Video número uno.

La Figura número 21 muestra el recorrido de la escena que realizó la cámara PTZ en la primera prueba.

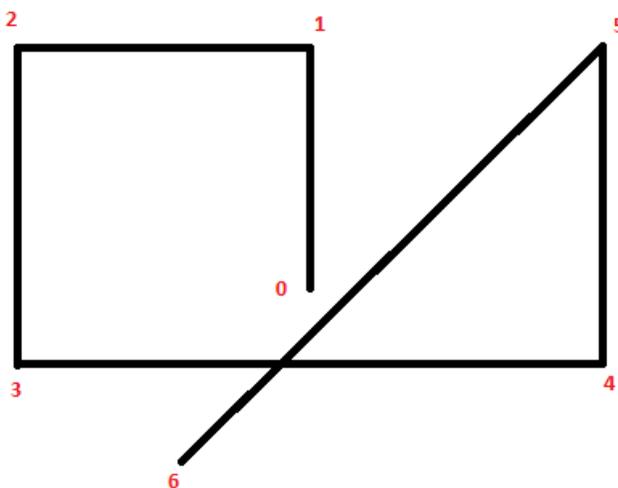


Figura 21 Recorrido realizado por la cámara PTZ

Durante el recorrido mostrado en la Figura 21 se grabó un video, a partir de este video se capturaron varias imágenes que fueron utilizadas para la autocalibración. Estas imágenes fueron escogidas

aleatoriamente y simplemente son frames resultantes de la descomposición del video del recorrido de la cámara. En la Figura 22 se muestran las imágenes en distintos puntos del recorrido.



Figura 22 Recorrido de la escena.

Con la secuencia de imágenes se realizó el proceso de autocalibración. El número de imágenes utilizadas para realizar la autocalibración fue de 40, el nivel de zoom fue escogido aleatoriamente y los resultados obtenidos son mostrados en las siguientes gráficas. Las figuras 24 y 23 muestran los resultados de la distancia focal y el punto principal respectivamente, estos valores están dados para cada imagen utilizada en el proceso de autocalibración y se comparan con los datos obtenidos utilizando el Toolbox. La Figura 25 ilustra el recorrido estimado de la cámara.

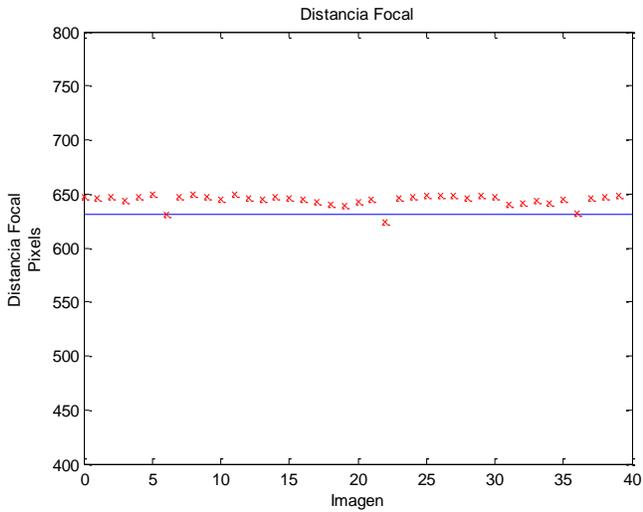


Figura 23 Comparación de la distancia focal obtenida para cada Homografía (puntos rojos) y la referencia (línea azul)

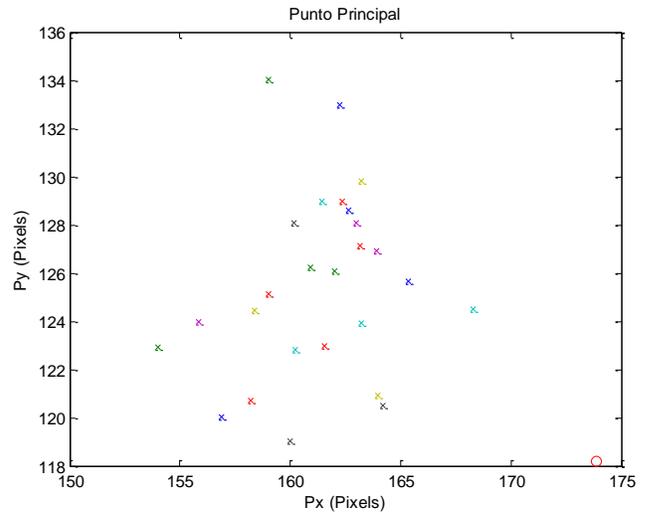


Figura 24 Valores del punto principal obtenidos para las distintas homografías. El círculo rojo indica el valor obtenido del Toolbox.

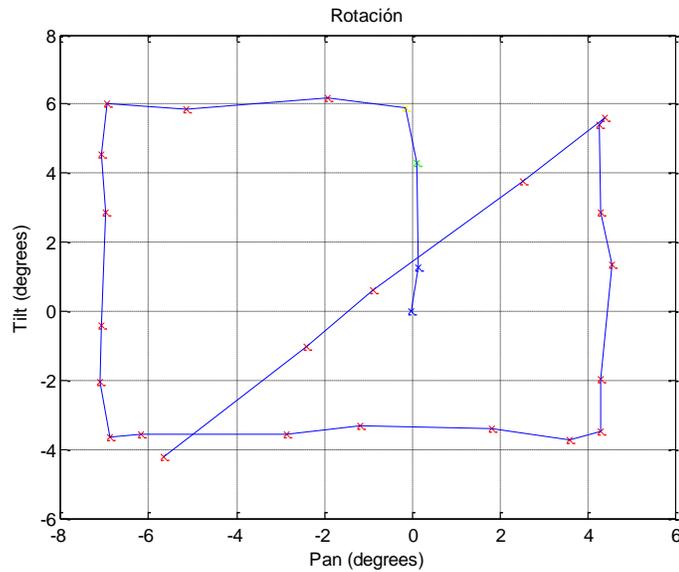


Figura 25 Recorrido estimado de la cámara

Los valores promedio de los parámetros intrínsecos obtenidos para esta secuencia de imágenes fueron los siguientes:

Distancia focal:

$$(f_x, f_y) = (644.4609; 644.4715)$$

Punto Principal:

$$(P_x, P_y) = (159.6106; 126.2642)$$

La desviación estándar de los resultados es ilustrada en la Tabla 1.

Tabla 1 Desviación Estándar

Parámetro	Desviación Estándar (pixeles)
fx	5,80012461
fy	5,784458966
Px	3,047920965
Py	3,760839288

Utilizando el método clásico de calibración los resultados obtenidos fueron los siguientes:

Distancia Focal:

$(f_x, f_y) = (630.777293069530290 ; 629.069712081724790)$

$fc_error = (12.849838658439788 ; 12.883189584045441)$

Punto Principal:

$(P_x, P_y) = (173.869201936033050 ; 118.198297077601840)$

$cc_error = (3.442180512037617 ; 5.168400501270845)$

Comparando los resultados del método clásico con el método de autocalibración se obtuvieron las gráficas de las Figuras 26 y 27 correspondientes a los errores para la distancia focal y el punto principal, respectivamente.

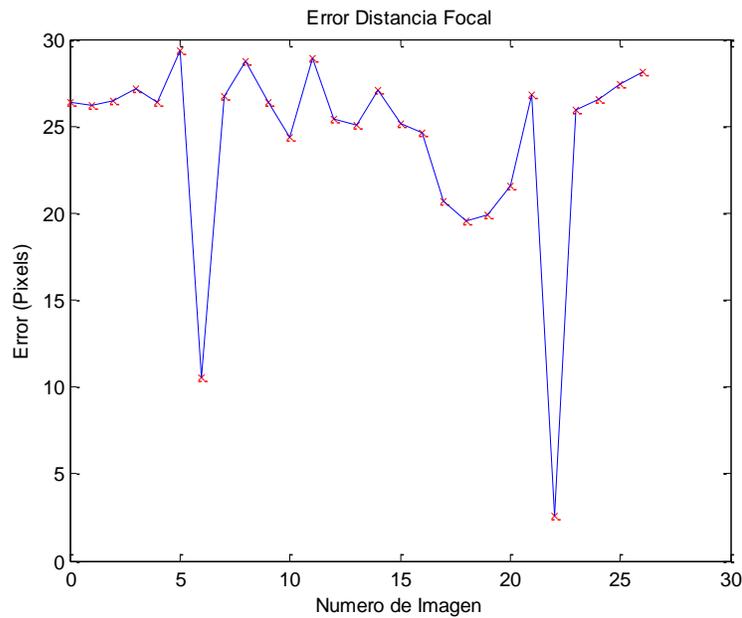


Figura 26 Error de la distancia focal obtenida para cada imagen.

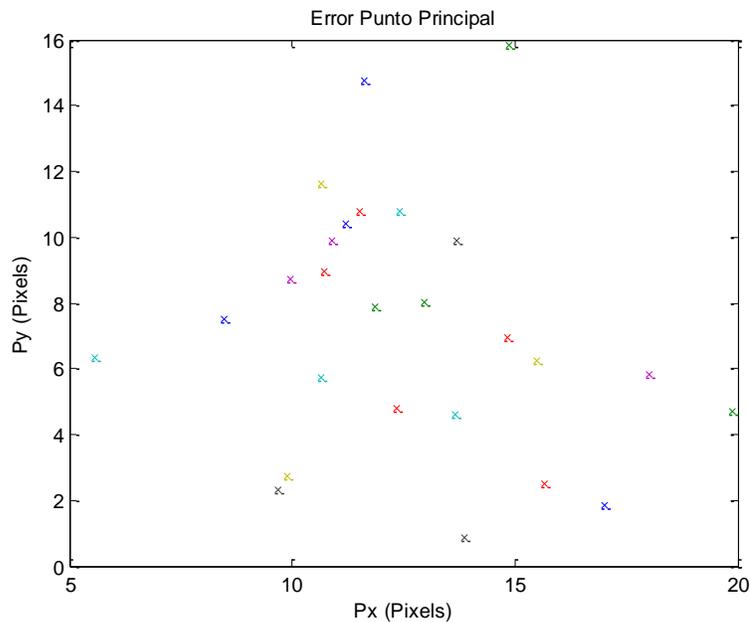


Figura 27 Error del Punto Principal en pixels.

Para este primer video se puede concluir que los resultados obtenidos por la autocalibración se acercan a los resultados obtenidos con la calibración clásica. En cuanto a la distancia focal se encuentra un error promedio de 13.7 pixeles, respecto al punto principal el error fue de 14 pixeles en P_x y de 8 pixeles para P_y . Esto da un porcentaje de error del dos por ciento para la distancia focal y para el punto principal se tiene un error del ocho por ciento para P_x y del 6.7 por ciento para P_y .

6.2.2. Video número dos.

Utilizando la misma escena se aumentó el nivel de zoom a una magnitud de x2. La cámara se mantuvo en la misma posición como en el primer video así que los objetos mantienen la misma distancia hasta la cámara. El número de imágenes utilizadas fue de 29 y el recorrido realizado fue igual al del video numero uno. Los objetos se encuentran a una distancia de 4 metros aproximadamente. En la Figura 28 se muestran imágenes que hacen parte del recorrido realizado por la cámara.

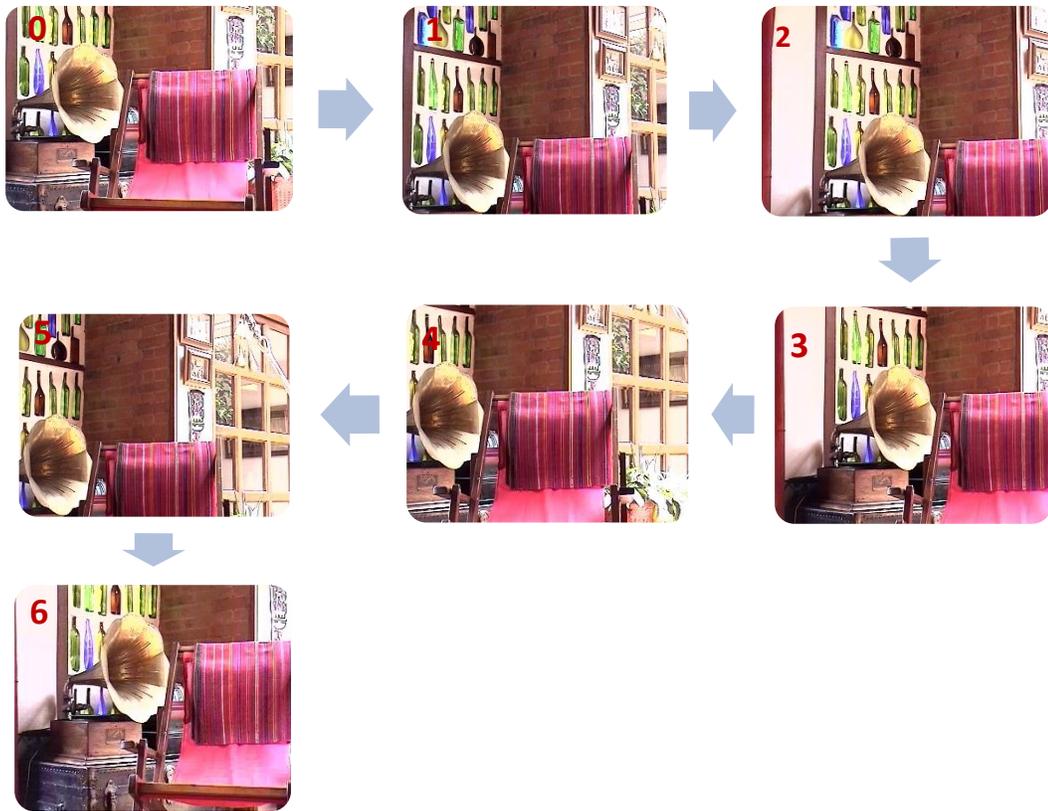


Figura 28 Recorrido de la cámara.

Las figuras 29, 30 y 31 ilustran los resultados obtenidos para el punto principal, estimación de rotación y distancia focal respectivamente.

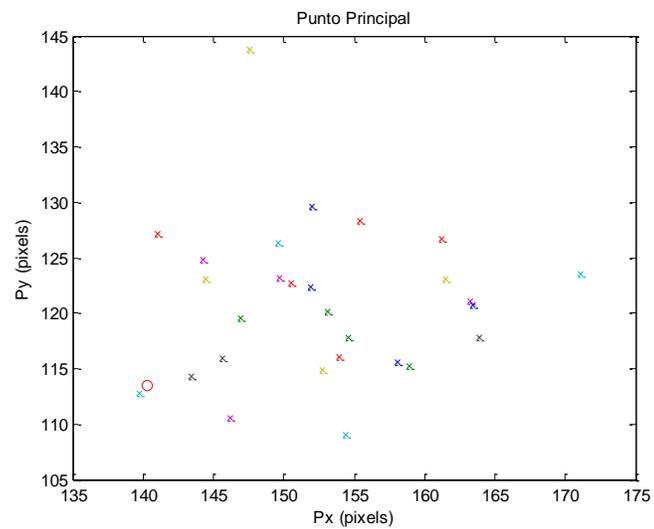


Figura 29 Resultados del punto principal, el círculo rojo es el valor obtenido utilizando el Toolbox.

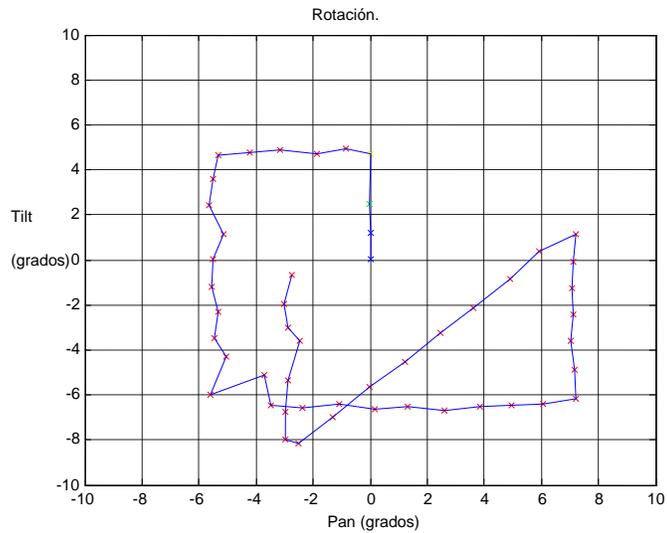


Figura 30 Rotación estimada.

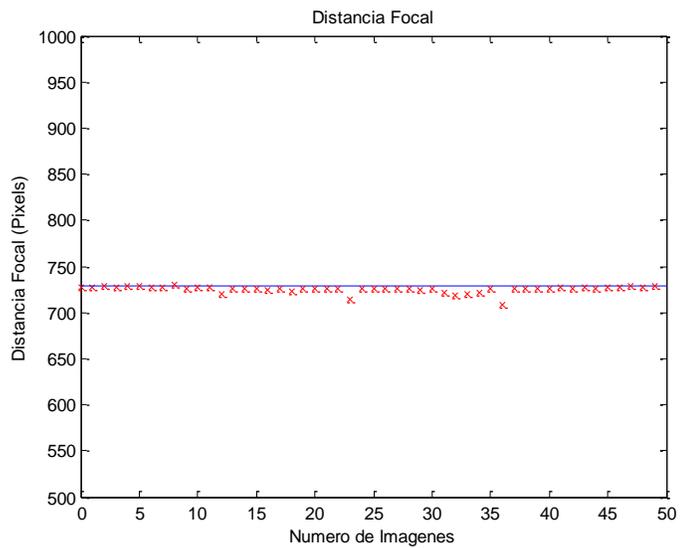


Figura 31 Distancia Focal estimada (puntos rojos), distancia focal de referencia (línea azul).

Los promedios para los valores intrínsecos de la cámara son los siguientes.

Distancia Focal:

$$(f_x, f_y) = (725.3537; 725.3629)$$

Punto principal:

$$(P_x, P_y) = (156.7271; 126.3976)$$

La desviación estándar se ilustra en la Tabla 2. Los resultados son buenos, la mayor desviación se encuentra en P_y donde fue de 11.17 píxeles.

Tabla 2 Desviación Estándar

Parámetros	Desviación Estándar (pixeles)
fx	3,765050155
fy	3,801594555
Px	4,796512727
Py	11,17196439

Los resultados obtenidos con el método clásico fueron los siguientes.

Distancia Focal:

$$(f_x, f_y) = (728.59241 , 728.66908)$$

$$fc_error = (8.92043 \quad 8.50147)$$

Punto Principal:

$$(P_x, P_y) = (140.26264, 113.49208)$$

$$cc_error = (16.44374 \quad 15.39861)$$

El error para la distancia focal fue de 3 pixeles, esto es aproximadamente un error del 0.5%. El punto principal tuvo un error de 16 pixeles para el punto principal en el eje X y de 13 pixeles en el punto principal en el eje Y dando un porcentaje de error del 11% para P_x y del 11.5% para P_y .

6.2.3. Video número tres.

Se realizó un recorrido parecido al descrito por la Figura 21, la escena fue la misma y el número de imágenes que se utilizaron fue de 47, se aumentó el nivel de zoom y una muestra de las imágenes de la escena se ilustra en la Figura 32:

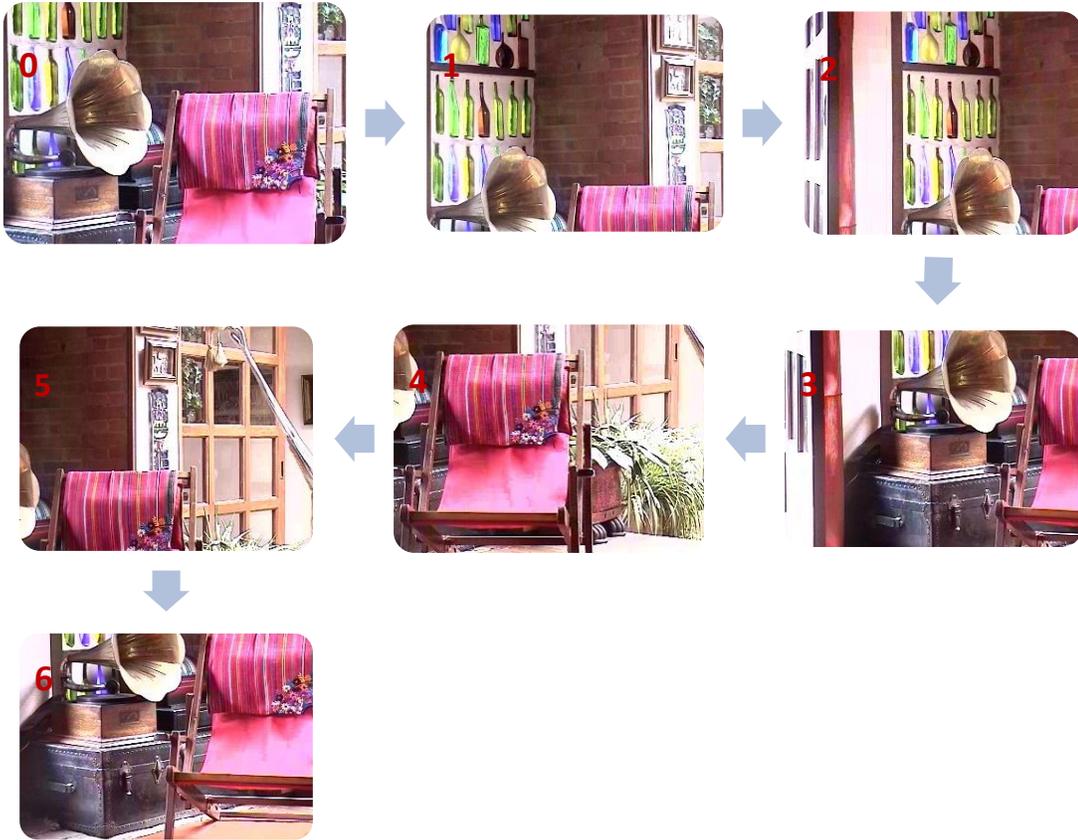


Figura 32 Imágenes del recorrido realizado por la cámara.

Al igual que en las demás pruebas se graficaron la distancia focal y el punto principal obtenido para cada imagen, los resultados obtenidos fueron los siguientes:

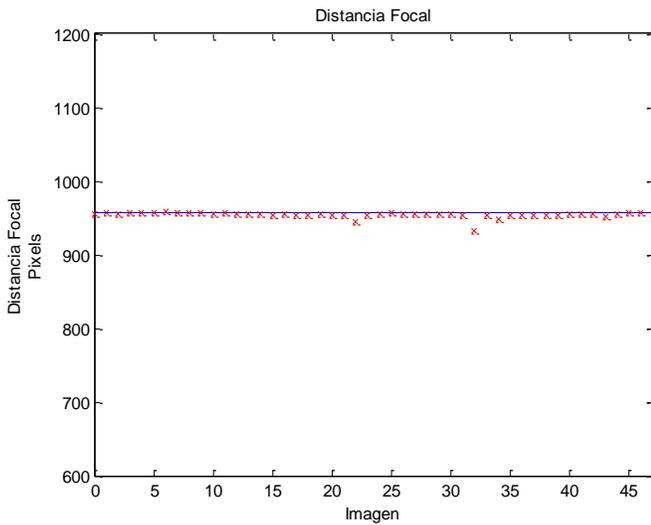


Figura 34 Distancia Focal estimada (puntos rojos) y referencial (línea azul).

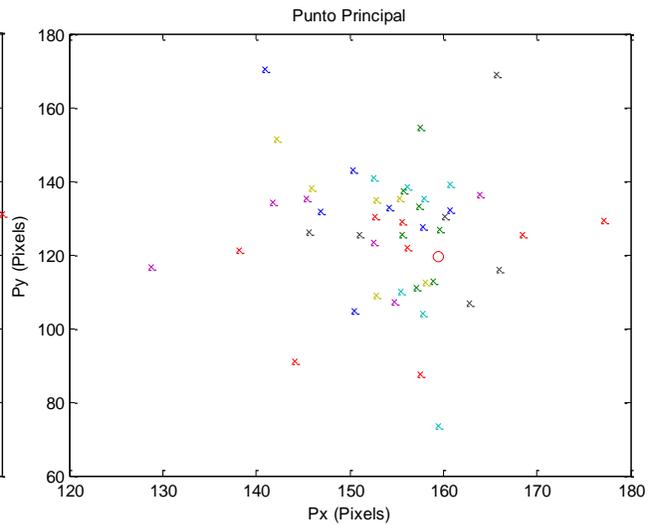


Figura 33 Puntos Principales estimados para cada homografía y el referencial (círculo rojo).

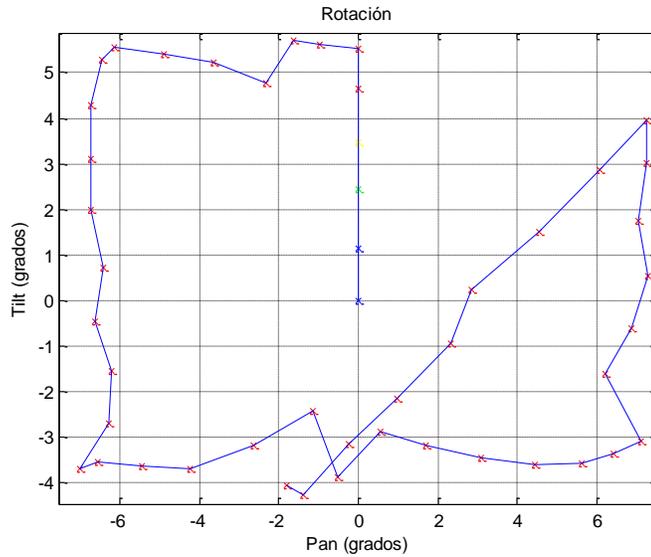


Figura 35 Rotación estimada.

Los valores promedio de los parámetros intrínsecos de la cámara son los siguientes:

Distancia Focal:

$$(f_x, f_y) = (954.3014; 954.2964)$$

Punto Principal:

$$(P_x, P_y) = (154.5055; 126.2031)$$

La desviación estándar se muestra en la Tabla 3. Al igual que en las pruebas anteriores, la mayor desviación estándar está en P_y .

Tabla 3 Desviación Estándar

Parámetro	Desviación Estándar (píxeles)
fx	3,938160629
fy	3,787248897
Px	8,351066462
Py	18,19062761

Realizando la calibración por el método clásico se obtuvieron los siguientes valores de los parámetros intrínsecos.

Distancia Focal:

$$(f_x, f_y) = (956.168196864726270 ; 953.317981606916760)$$

$$\text{error} = (34.929646807630476 ; 34.815857818210496)$$

Punto Principal:

$$(P_x, P_y) = (159.5, 119.5)$$

Los errores para la distancia focal y el punto principal estimado son graficados en la Figuras 36 y 37 respectivamente.

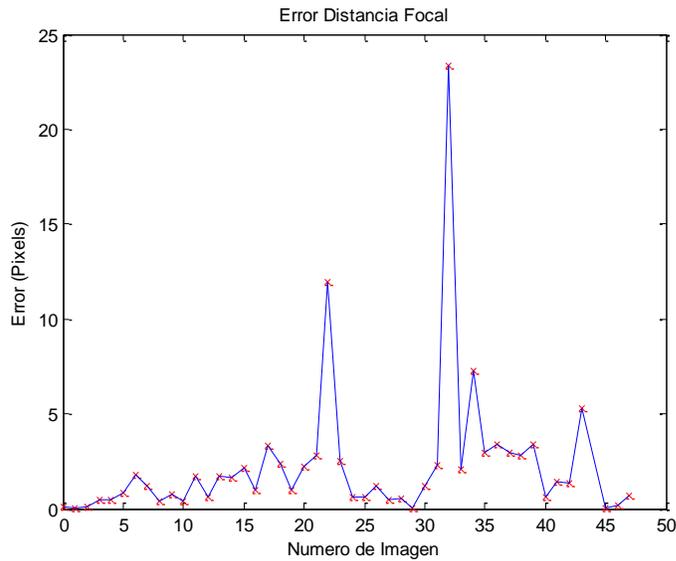


Figura 36 Error de la Distancia Focal dado en pixeles.

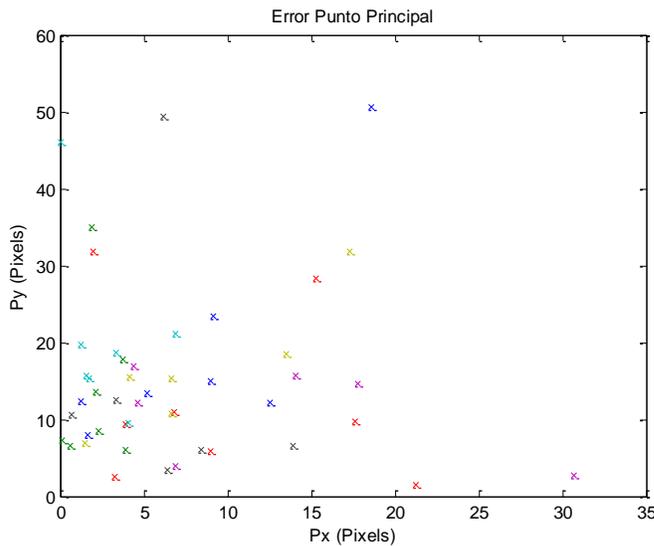


Figura 37 Error en el punto principal.

El error promedio para la distancia focal fue de 2 pixeles, para el punto principal el error fue de 5 pixeles para P_x y 7 pixeles para P_y . El porcentaje de error para la distancia focal fue menor al 1 por ciento mientras que el porcentaje de error para el punto principal fue del tres por ciento en P_x y de 5.83 por ciento para P_y . Una razón para tener mayor precisión en el resultado es el del número de imágenes procesadas para realizar el algoritmo y otra puede ser el nivel del zoom que favorece el algoritmo. Más adelante se evaluará la influencia del número de imágenes utilizadas en la autocalibración sobre los resultados.

6.2.4. Mínimo nivel de zoom.

Para este video el nivel del zoom utilizado fue el mínimo. La escena utilizada fue la misma a los puntos anteriores, el recorrido realizado por la cámara PTZ se ilustra en la Figura 21 y en la Figura 38 se ilustra el recorrido en imágenes:

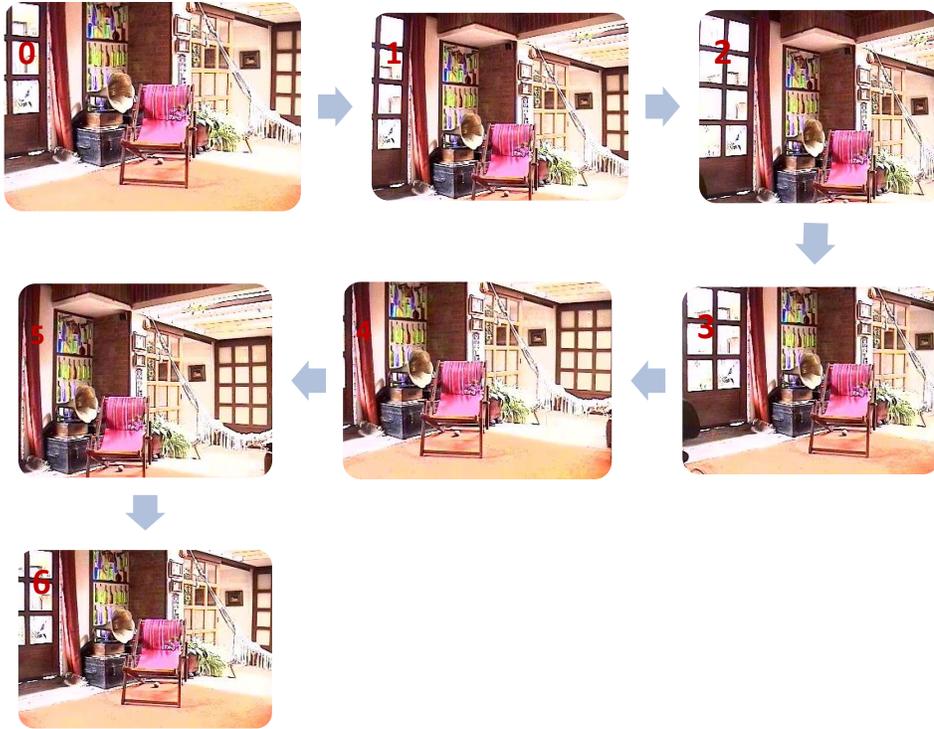


Figura 38 Imágenes del recorrido realizado.

Los resultados obtenidos para la distancia focal, el punto principal y el recorrido estimado se exponen en las Figuras 39,40 y 41 respectivamente.

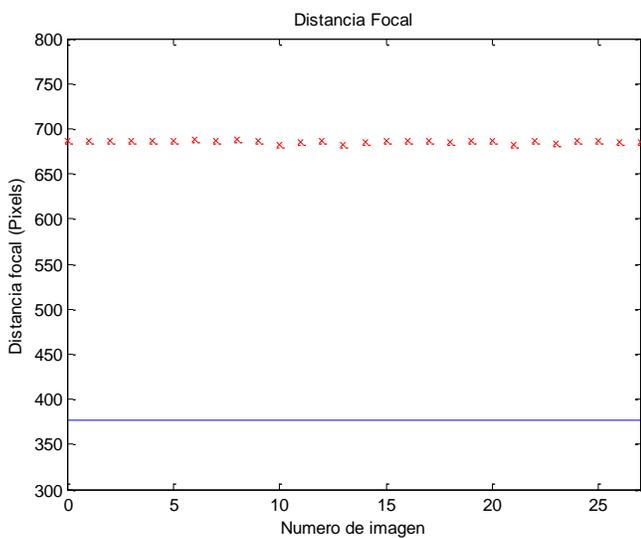


Figura 39 Distancia Focal estimada (puntos rojos) y la distancia focal de referencia (línea azul)

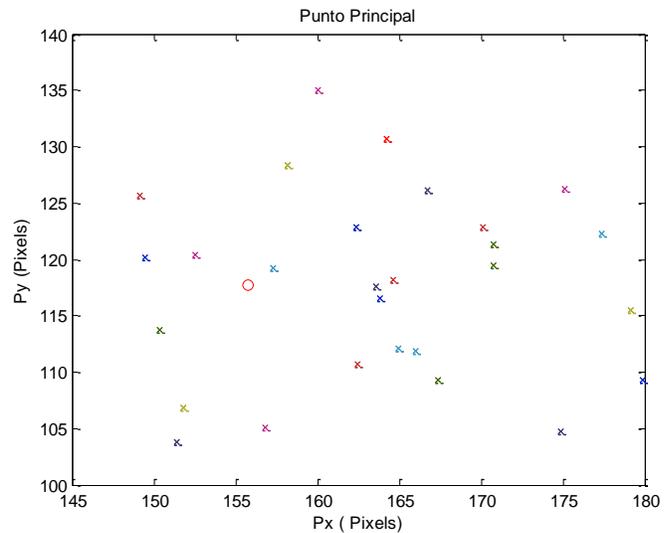


Figura 40 Puntos Principales estimados y punto principal referencial (circulo rojo)

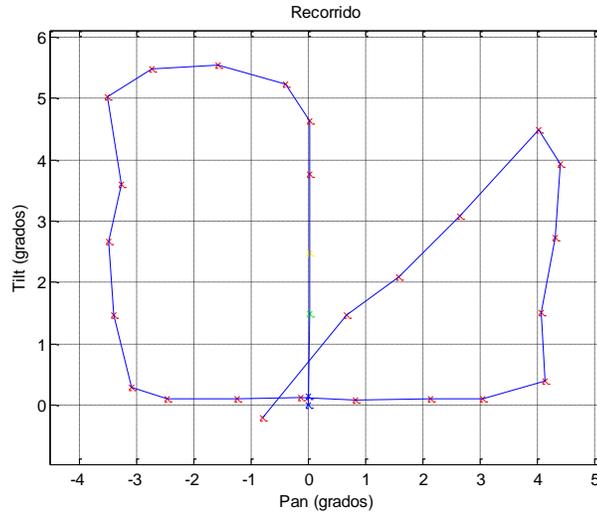


Figura 41 Rotación de la cámara estimada.

Los valores promedio de los parámetros intrínsecos de la cámara son los siguientes.

Distancia Focal:

$$(f_x, f_y) = (686.1853; 686.2039)$$

Punto Principal:

$$(P_x, P_y) = (163.5696; 118.1234)$$

A continuación se ilustran los resultados para la desviación estándar, la mayor variación se encuentra en P_x y P_y .

Tabla 4 Desviación Estándar.

Parámetro	Desviación Estándar
fx	1,592184688
fy	1,296530999
Px	8,990688919
Py	8,330203368

Con el método de calibración clásico se obtuvieron los siguientes resultados.

Distancia Focal:

$$(f_x, f_y) = (377.232927897205120 ; 375.519934950102250)$$

$$\text{error} = (5.796004649911210 ; 5.923162668740111)$$

Punto Principal:

$$(P_x, P_y) = (155.734844478920960 ; 117.706880630976060)$$

$$\text{error} = (6.739510071766659 ; 8.762542200683646)$$

El error para la Distancia Focal y el punto principal es graficado en las Figuras 42 y 43:

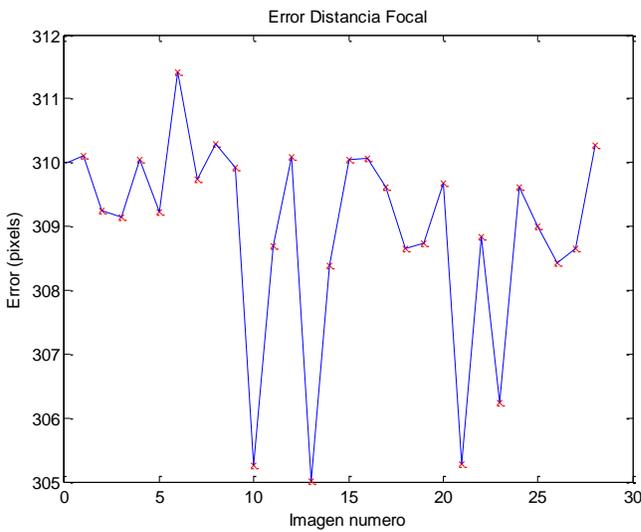


Figura 42 Error Distancia Focal.

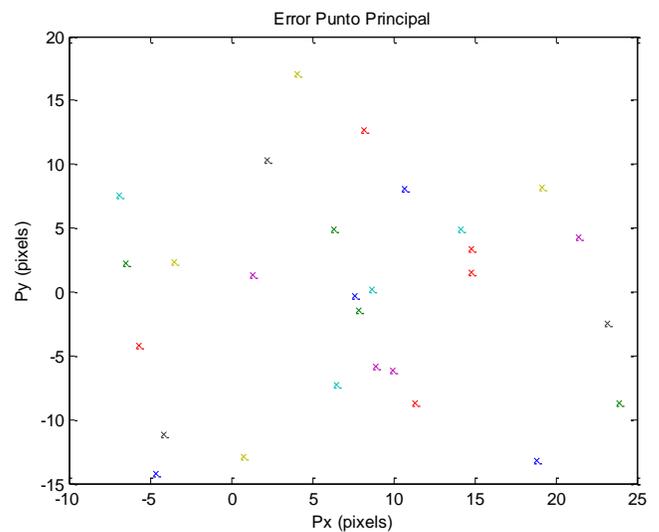


Figura 43 Error del punto principal.

Como se puede observar en las graficas anteriores, el error para la distancia focal es muy alto. El error supera los 300 pixeles para la distancia focal, mientras que el error del punto principal oscila entre los 20 pixeles.

6.2.5. Máximo nivel de zoom.

Se realizó una prueba con el máximo zoom de la cámara. La escena escogida fue una donde la distancia fuera larga desde la escena hasta la cámara, esto para que al realizar el barrido de la escena de la cámara no exista un gran cambio de los objetos en la escena y así encontrar una mayor correspondencia de puntos entre las imágenes. Un muestreo del barrido de la escena se observa en la Figura 43.

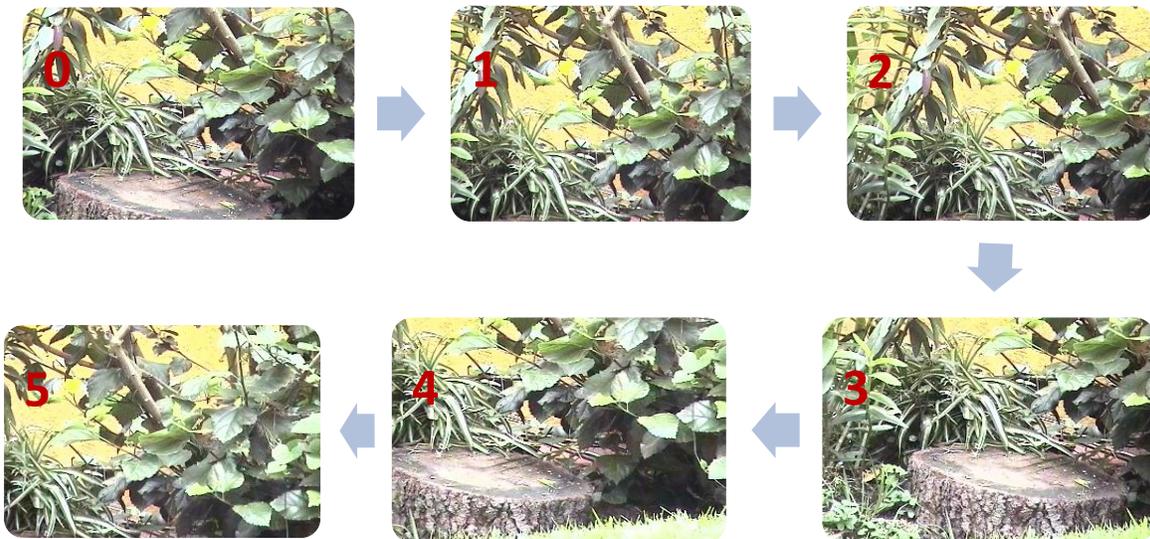


Figura 44 Imágenes del barrido de la escena.

El número de imágenes utilizadas para realizar la autocalibración fue de 28 imágenes. Los resultados obtenidos para la distancia focal se ilustran en la Figura 45.

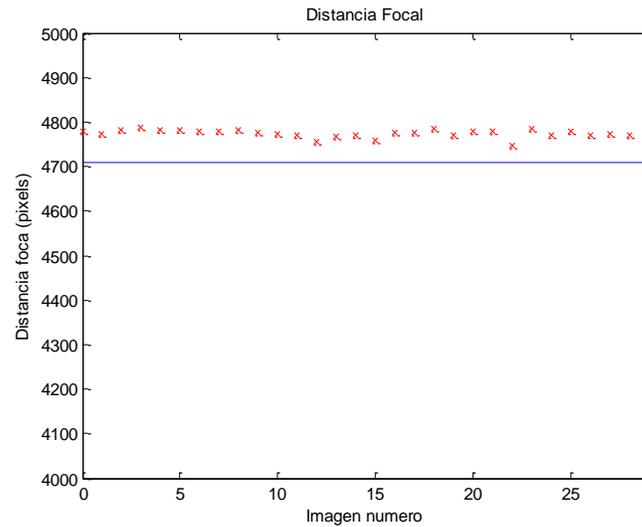


Figura 45 Distancia Focal estimada (Puntos rojos) y distancia focal de referencia (línea azul)

Los valores obtenidos para los parámetros intrínsecos de la cámara son los siguientes.

Distancia Focal:

$$(f_x, f_y) = (4.7747e+003; 4.7749e+003)$$

Punto Principal:

$$(P_x, P_y) = (325.7834; 57.3071)$$

Los resultados obtenidos con el algoritmo de calibración clásico son los siguientes:

Distancia Focal:

$$(f_x, f_y) = (4710.180704386500100 ; 4715.349800806342500)$$

$$\text{error} = (675.491637258006220 ; 676.337604161930360)$$

Punto Principal :

$$(P_x, P_y) = (181.871171526802610 ; 125.674194552943590)$$

$$\text{error} = (15.218935648961130 ; 10.020633985446313)$$

Al igual que con el nivel mínimo de zoom, los resultados no son los mejores ni para el punto principal ni para la distancia focal.

A continuación se muestra la Tabla 5 con el resumen de todos los parámetros intrínsecos obtenidos en los videos evaluados, así como el error en píxeles y el correspondiente porcentaje del error obtenido.

Tabla 5 Resumen de resultados con todos los videos utilizados.

	Parámetros	Estimados	Reales	Error	Porcentaje
VIDEO 1	Fx	644,4609	630,7772	13,6837	2,1693397
	Fy	644,4715	629,0697	15,4018	2,4483455
	Px	159,6106	173,8692	14,2586	8,2007624
	Py	126,2642	118,1982	8,0660	6,824131
VIDEO 2	Fx	725,3537	728,5924	3,2387	0,444516
	Fy	725,3629	728,6690	3,3061	0,4537177
	Px	156,7271	140,2626	16,4645	11,738339
	Py	126,3976	113,4929	12,9047	11,370491
VIDEO 3	Fx	954,3014	956,1681	1,8667	0,1952272
	Fy	954,2964	953,3179	0,9785	0,1026415
	Px	154,5055	159,5000	4,9945	3,131348
	Py	126,2031	119,5000	6,7031	5,6092887
VIDEO 4	Fx	686,1853	377,2329	308,9524	81,899643
	Fy	686,2039	375,5199	310,6840	82,734364
	Px	163,5696	155,7348	7,8348	5,0308602
	Py	118,1234	117,7068	0,4166	0,3539303
VIDEO 5	Fx	1.082,8000	1.080,6375	2,1625	0,2001134
	Fy	1.082,8000	1.077,2631	5,5369	0,5139784
	Px	152,8481	190,6032	37,7551	19,808219
	Py	119,9151	128,0204	8,1053	6,3312566
VIDEO 6	Fx	4.774,7000	4.710,1807	64,5193	1,369784
	Fy	4.774,9000	4.715,3498	59,5502	1,262901
	Px	325,7834	181,8711	143,9123	79,128735
	Py	57,3071	125,6741	68,3670	54,40023
Error Promedio					16,0718

En conclusión se obtuvieron buenos resultados con el algoritmo de autocalibración. El mayor error que se puede observar fue el que se obtuvo con el menor nivel de zoom de la cámara y cuando se utilizó el mayor nivel de zoom. En la Figura 46 se observa una tendencia del error porcentual frente a cada nivel de zoom, el error más alto está en la distancia focal mínima y se va reduciendo a medida que se aumenta el zoom. Cuando la distancia focal es la máxima el error aumenta levemente.

Distancia focal vs Error porcentual

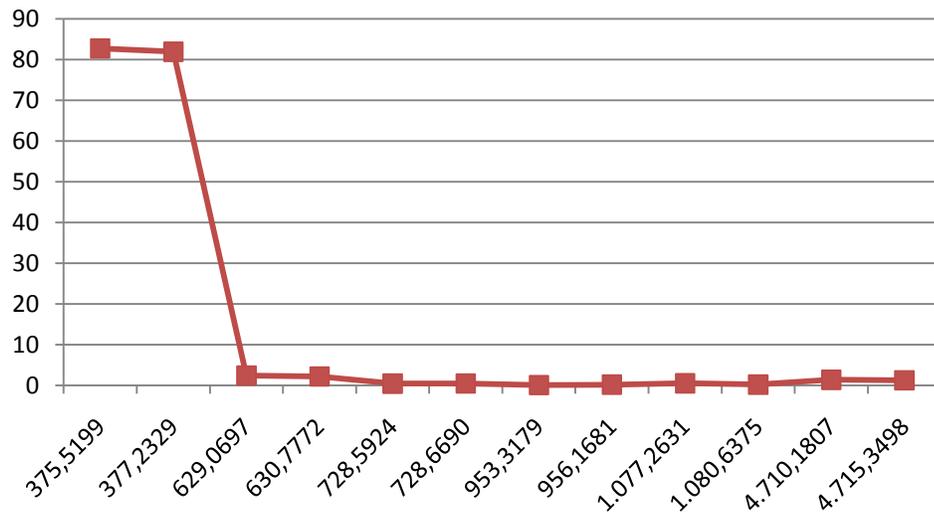


Figura 46 Gráfica de Distancia focal vs el Error porcentual.

En la Figura 47 se muestra una grafica más detallada del rango de distancia focales entre 629 pixeles y 4715. Los resultados muestran una tendencia respecto al error y la distancia focal, con el mayor y nivel menor de zoom se obtiene el mayor error.

Distancia focal vs Error porcentual

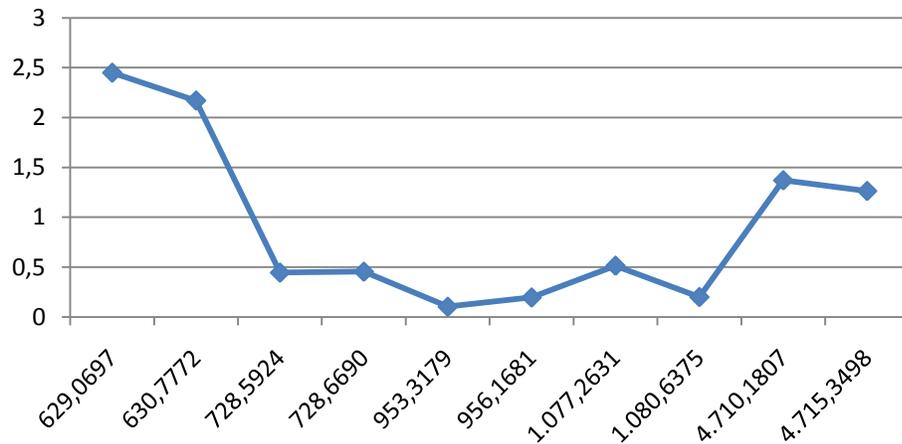


Figura 47 Distancia vs Error porcentual.

6.2.6. Variación del número de imágenes.

En esta sección se realiza el análisis de la influencia del número de imágenes sobre la precisión de los resultados de la calibración. Para comprobar esta influencia se ha realizado un experimento donde el número de imágenes se ha variado para realizar el proceso de calibración. Durante toda la secuencia de

imágenes utilizada para realizar la prueba, los parámetros intrínsecos se han mantenido fijos. El rango de variación del número de imágenes en todos los casos ha sido entre 5 y 52. Los resultados obtenidos en este experimento se muestran en la figura 48, donde se muestra el resultado de la distancia focal utilizando un distinto número de imágenes.

En general, se observa que este método presenta errores muy grandes en la estimación de los parámetros si el número de imágenes procesadas es inferior a 40. Este número de imágenes puede variar dependiendo de la calidad de las homografías. Si se tiene un grupo de homografías muy precisas, este valor es inferior.

Distancia Focal

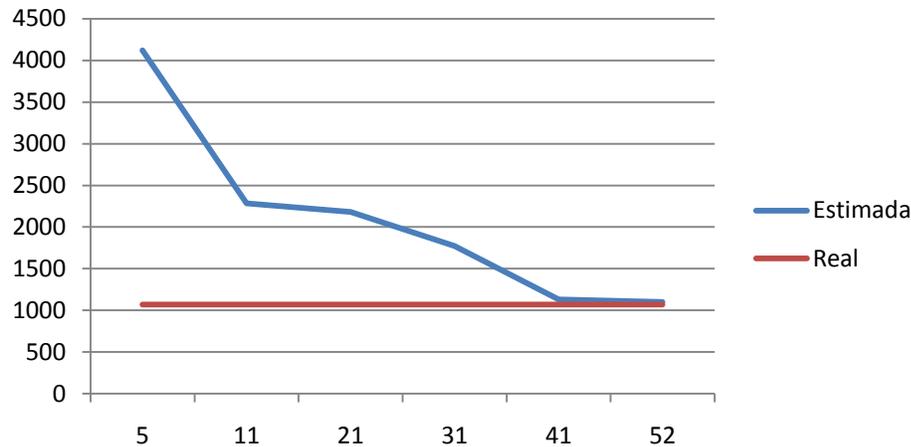


Figura 48 Valores de la Distancia Focal Respecto al número de imágenes utilizadas.

En la Tabla 6 se muestran los resultados obtenidos de la distancia focal utilizando un distinto número de imágenes para realizar el proceso de autocalibración.

Tabla 6 Tabla de valores de la distancia focal respecto al número de imágenes.

N° Imágenes	Distancia Focal Estimada (Píxeles)	Distancia Focal de referencia (Píxeles)
52	1100,2	1100,3
41	1129,7	1129,7
31	1773,6	1773,6
21	2180,7	2181,2
11	2286,1	2286,3
5	4122,5	4122,9

Error Porcentual

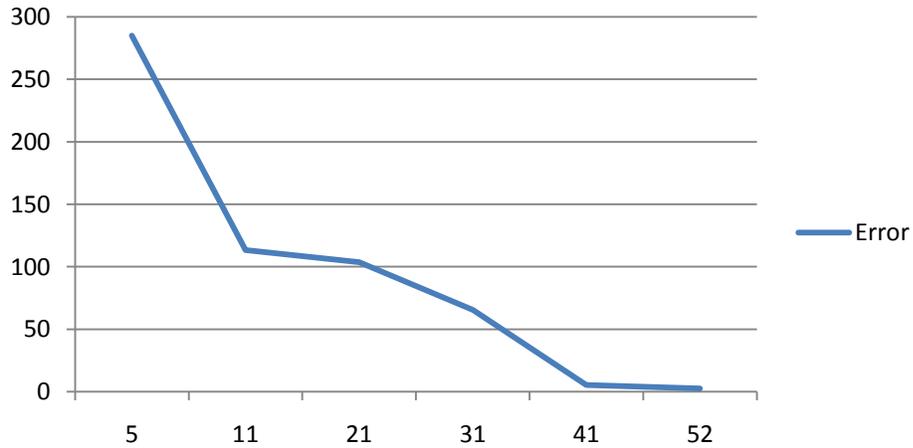


Figura 49 Error Porcentual de la distancia focal respecto al número de imágenes utilizadas.

Como conclusión se puede decir que es necesario un buen número de imágenes para que el método converja y dé buenos resultados. El número de imágenes para que el método converja puede variar ya que depende de la calidad de las homografías y de las imágenes utilizadas.

6.3. Pruebas de evaluación de rotación.

Para estas pruebas se compara directamente el recorrido real de la cámara con el estimado por el algoritmo, como se ha dicho anteriormente la posición real de la cámara es obtenida por el programa de Sony. El zoom fue escogido aleatoriamente y los resultados de las rotaciones son los siguientes.

Video Uno

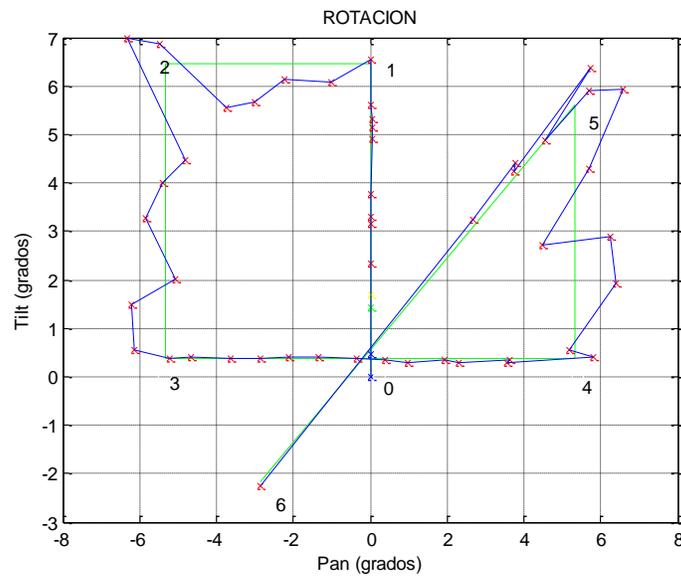


Figura 50 Comparación de la rotación estimada (línea azul) y la real (línea verde).

En la Figura 50 se ilustra el resultado de la rotación estimada y se compara con el recorrido real. Los resultados son buenos, dando un error menor a 1 grado. En la tabla 7 se resumen los resultados y errores obtenidos.

Tabla 7 Evaluación de los datos de rotación.

Posición Predeterminada Número	Pan (grados)			Tilt (grados)		
	Real	Estimado	Error (°)	Real	Estimado	Error (°)
1	0	0.02	0.02	6.45	6.52	0.07
2	-5.325	-5.27	0.055	6.45	6.6	0.15
3	-5.325	-6.03	0.705	0.375	0.53	0.155
4	5.325	5.88	0.555	0.375	0.413	0.038
5	5.325	5.707	0.382	5.625	6.33	0.705
6	-2.85	-2.659	0.191	-2.175	-2.081	0.094

Video dos

La figura 50 ilustra los resultados de la rotación estimada para otro video realizado.

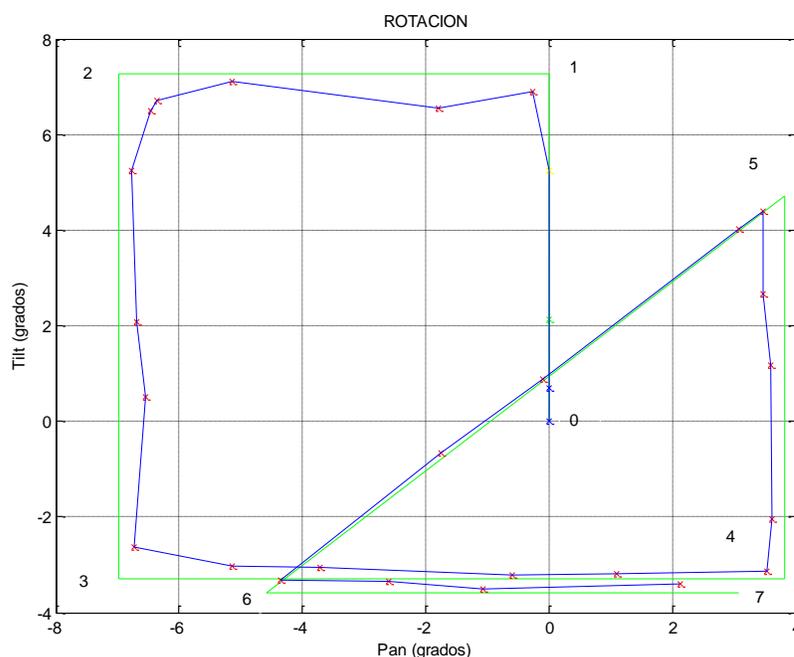


Figura 51 Recorrido real (línea verde) y estimado de la cámara (línea azul).

Los resultados se resumen en la Tabla 8 y se calcula el error de cada medición. Los resultados son buenos, dando un error menor a 1 grado.

Tabla 8 Evaluación de los datos de rotación.

Posición Predeterminada Número	Pan (grados)			Tilt (grados)		
	Estimado	Real	Error	Estimado	Real	Error
1	-0.2	0	0.2	6.89	7.275	0.385
2	-6.367	-6.975	0.608	6.715	7.275	0.56
3	-6.733	-6.975	0.242	-2.632	-3.3	0.668
4	3.525	3.825	0.3	-3.13	-3.3	0.17
5	3.469	3.825	0.356	4.402	4.725	0.323
6	-4.363	-4.575	0.212	-3.327	-3.6	0.273
7	2.141	3.075	0.934	-3.395	-3.6	0.205

6.4. Prueba con otro tipo de escena

Para concluir con las pruebas de los parámetros intrínsecos y extrínsecos se cambió la escena a trabajar para probar el algoritmo con otro tipo de escena, esto se realizó porque se sabe que el número de puntos encontrados por el detector de esquinas Harris es importante para la calidad de las homografías y por consiguiente para la calidad de los resultados [4]. Se escogió una escena donde existiera un mayor número de cambios de intensidad en los pixeles.

La escena que se usó y el recorrido de la cámara son mostrados en las figuras 52 y 53.

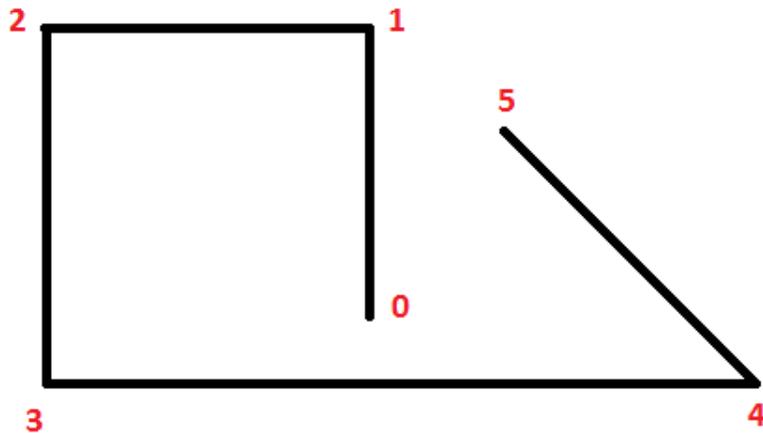


Figura 52 Recorrido de la cámara PTZ

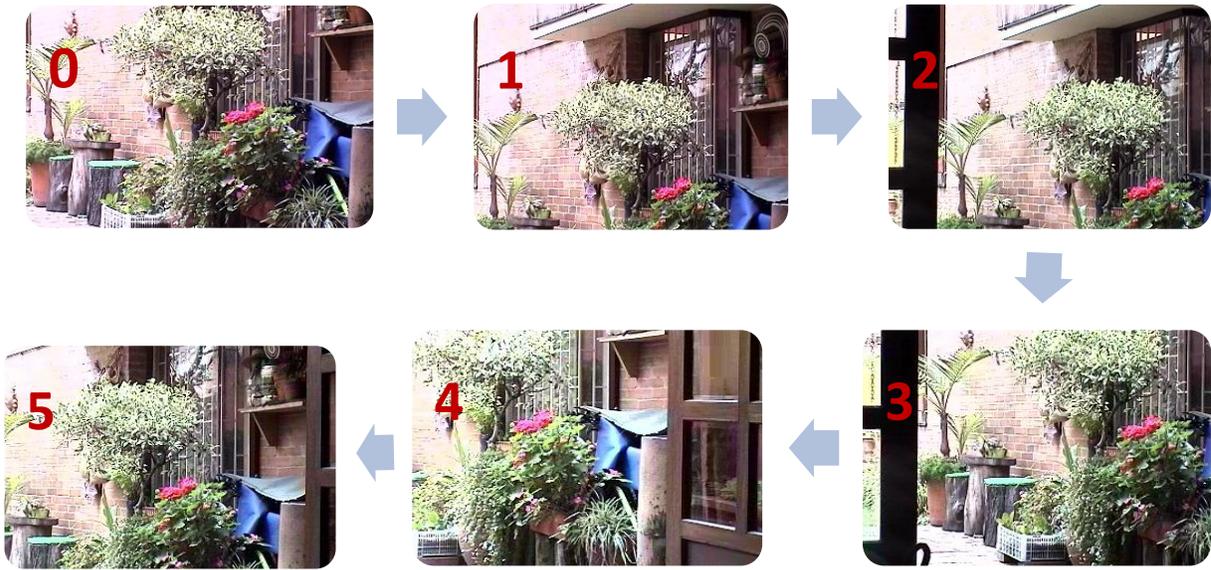


Figura 53 Imágenes del recorrido realizado.

Resultados:

Al igual que en las pruebas realizadas en la sección 6.2, se evaluaron los resultados de la distancia focal y el punto principal. Para este último video se agrega la evaluación de la rotación de la cámara. Las figuras 55, 54 y 56 ilustran estos resultados respectivamente.

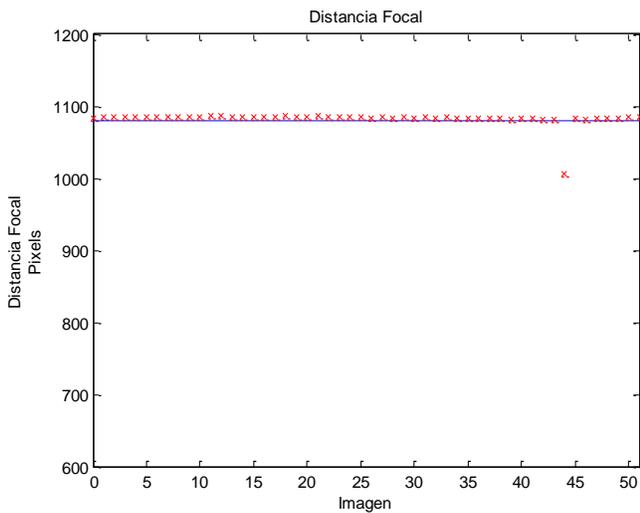


Figura 55 Distancia Focal estimada (puntos rojos) y la distancia focal referencia (línea azul)

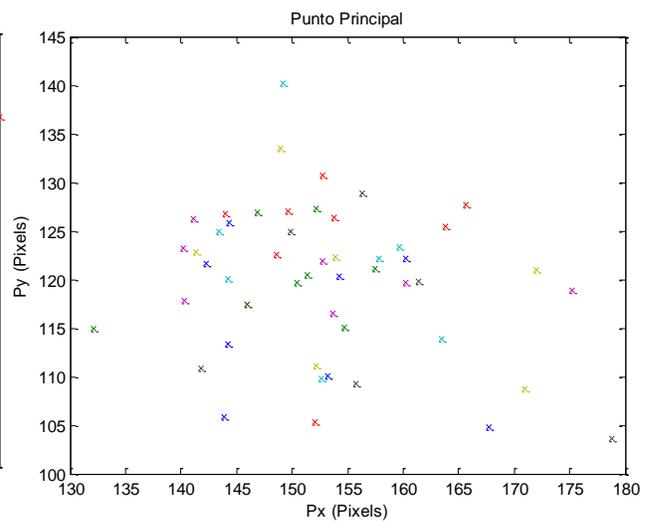


Figura 54 Punto Principal estimado.

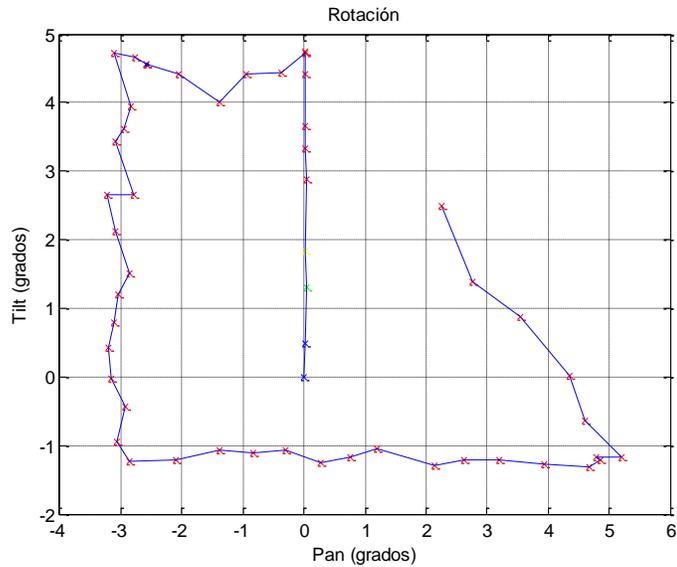


Figura 56 Rotación Estimada.

Los resultados promedio de los parámetros intrínsecos de la matriz de calibración son:

Distancia Focal:

$$(f_x, f_y) = (1.0828e+003; 1.0828e+003)$$

Punto Principal:

$$(P_x, P_y) = (152.8481; 119.9151)$$

Los parámetros encontrados por medio del método clásico son los siguientes:

Distancia Focal:

$$(f_x, f_y) = (1080.637525790507400 ; 1077.263146223431000)$$

$$\text{error} = (24.641868814236325 ; 24.575748085893608)$$

Punto Principal :

$$(P_x, P_y) = (190.603201009570280 ; 128.020446099257360)$$

$$\text{error} = (9.552701634959803 ; 9.391487710069471)$$

Los errores resultantes se ilustran en las gráficas 57 y 58 para la distancia focal y el punto principal respectivamente.

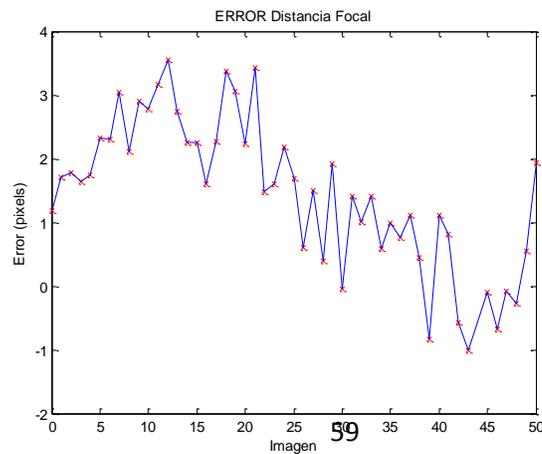


Figura 57 Error en la distancia focal estimada.

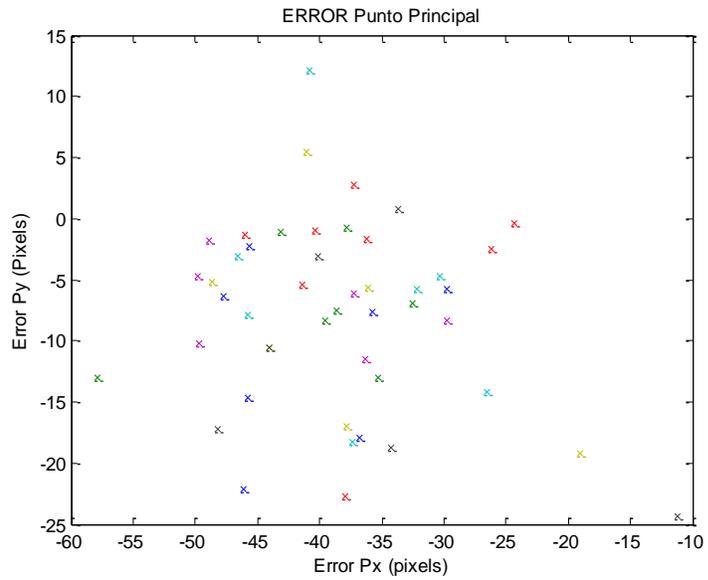


Figura 58 Error en el punto principal estimado

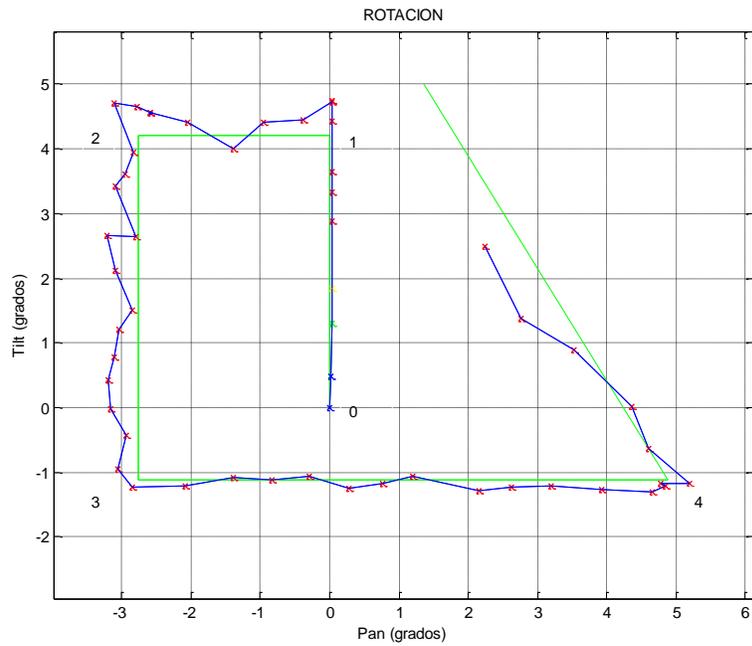


Figura 59 Rotación estimada (línea azul) y rotación real (línea verde)

A continuación se muestra una tabla donde se muestran los valores de rotación estimados y los valores reales.

Tabla 9 Tabla comparativa de la rotación de la cámara.

Posición Predeterminada Número	Pan (grados)			Tilt (grados)		
	Real	Estimado	Error	Real	Estimado	Error
1	0	0.02	0.02	4.2	4.73	0.53
2	-2.77	-3.11	0.34	4.2	4.71	0.51
3	-2.77	-2.85	0.08	-1.125	-1.23	0.105
4	4.875	5.19	0.315	-1.125	-1.17	0.045

Se puede observar que los valores de la distancia focal y la rotación estimada son muy buenos, en cuanto a el punto principal los resultados llegan a un error del 20% para P_x y de un 7% para P_y . En la tabla se comparan los resultados de las rotaciones de la cámara, el resultado es muy bueno ya que el error en grados que cada estimación es menor a un grado.

6.5. Estimación de distancia.

A continuación se muestran los resultados para obtener mediciones de objetos o de la distancia de la cámara hasta cierto punto.

Para no realizar el proceso completo de autocalibración se utilizó la información obtenida de los parámetros intrínsecos de una de las pruebas realizadas anteriormente. Para esto se ajustó la cámara con el mismo nivel de zoom utilizado en el video uno, estos parámetros no cambian con la escena. Los parámetros intrínsecos obtenidos para el nivel de zoom 11D0 son los siguientes:

$$A = \begin{bmatrix} 644.4609 & 0.5571 & 159.6106 \\ 0 & 644.4715 & 126.2642 \\ 0 & 0 & 1.0486 \end{bmatrix}$$

Para poder obtener el ángulo recorrido por la cámara es necesario tener la matriz de parámetros intrínsecos y la homografía respectiva. Las imágenes que se utilizan para la homografía deben ser tomadas dependiendo de lo que se quiere medir o dependiendo de la información que se conoce. Para la primera prueba se conoce la medida de un lado del marco de una foto como se indica en la figura 60, este lado mide 15 cm. Utilizando esta información y con el ángulo de rotación se encuentra a qué distancia está la cámara del objeto. Para obtener el ángulo de rotación se toma como punto de referencia el lado derecho de la imagen capturada, con este punto de referencia se toman dos imágenes. Las imágenes se capturan en el momento que el punto de referencia coincide con el punto de origen y final de la medida conocida, en este caso los 15cm del marco de la foto. En consecuencia la primera imagen se captura cuando el borde derecho de la imagen capturada coincide con el punto de origen de los 15cm. Luego se mueve la cámara hacia la izquierda de forma que recorra los 15cm del marco y la otra imagen se captura en el momento que el lado de 15 cm desaparece de la imagen.



Figura 60 Imágenes A y B para encontrar la homografía correspondiente. La primera imagen se ajusta para que coincida el borde de la imagen capturada con el punto de origen de los 15cm. La segunda imagen se captura cuando el borde derecho de la imagen llega al punto final de los 15cm.

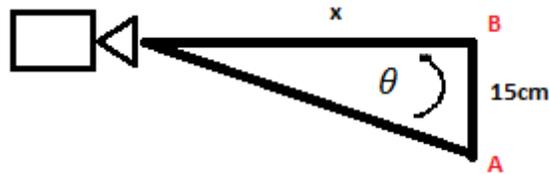


Figura 61 Vista superior marcando los puntos A y B correspondientes a las imágenes tomadas.

Con estas imágenes se obtuvo la homografía que se utilizó para encontrar el ángulo recorrido. La homografía obtenida se ilustra la Figura 62.

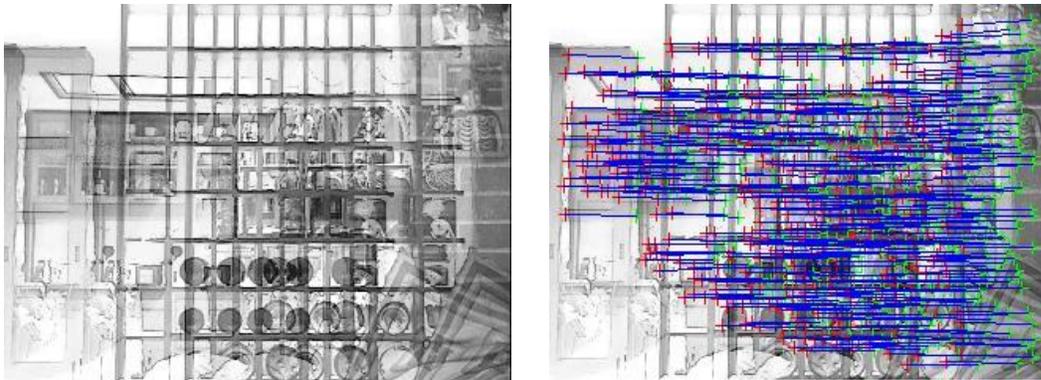


Figura 62 Imágenes superpuestas y la homografía resultante.

El ángulo recorrido por la cámara para tomar estas dos imágenes fue de 3.9442° , esta información se obtuvo a partir de la matriz de rotación R . Con estos datos se encontró la distancia de la cámara hasta la pared utilizando la ecuación (1.26) de la sección 5.2. El resultado obtenido fue de 217.55 cm mientras que el dato real fue de 217cm, dando un error menor a un centímetro.

Se realizó una nueva prueba con una escena diferente y manteniendo los mismos parámetros internos de la escena anterior, en la Figura 63 se ilustra las imágenes utilizadas y la información que se conoce de la escena. Se sabe que la parte inferior de la puerta de madera mide 43cm.



Figura 63 Imágenes utilizadas para la homografía

Con estas dos imágenes se encontró la homografía respectiva que se ilustra en la Figura 64.

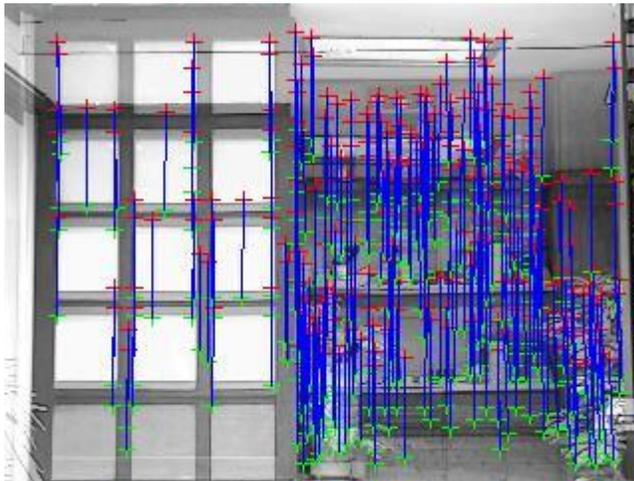


Figura 64 Superposición de las imágenes y la homografía resultante

Para encontrar el ángulo recorrido se realiza el mismo procedimiento explicado para la primera escena de este capítulo, pero se cambia el lado de referencia que en este caso es el inferior. El ángulo recorrido de la cámara que se encontró fue de 4.4418° hacia arriba. Con esto se obtiene que la distancia estimada de la cámara hasta la puerta era de 5.53 metros, la distancia real de la cámara hasta la puerta fue de 5.50 metros dando un error de 3cm.

Utilizando esta misma escena y sin trasladar la cámara se encontró la altura de la matera que se encuentra al lado de la puerta. Con la distancia de 5.53 metros encontrada en la prueba anterior se despejó la ecuación para poder encontrar la altura de la matera. En la Figura 65 se ilustran las imágenes utilizadas para encontrar la homografía y así encontrar el ángulo recorrido por la cámara.



Figura 65 Imágenes utilizadas para encontrar la altura de la materia que se encuentra en la mitad de la imagen.

En la Figura 66 se ilustra la homografía resultante para encontrar el ángulo θ de la imagen.

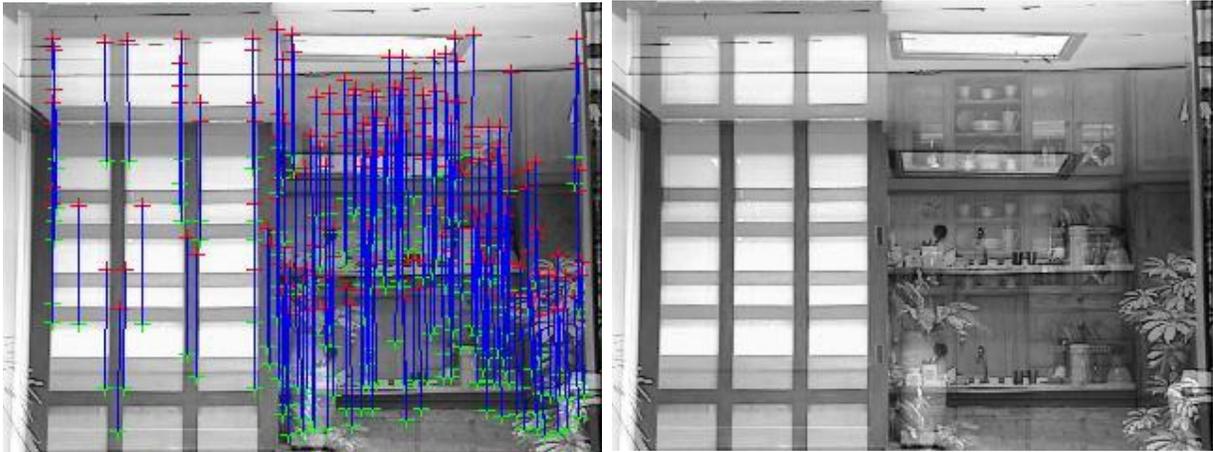


Figura 66 Homografía resultante

El ángulo de rotación de la cámara verticalmente fue de 5.6569° . Se obtuvo una altura estimada de la materia de 54.83cm mientras que la medida real es de 57 cm dando un error menor a 3cm. En la Tabla 11 se realiza un resumen de las pruebas de distancia realizadas con sus respectivos errores.

Tabla 10 Resumen de los distancias estimadas y sus respectivos errores.

Real (cm)	Estimada(cm)	Error(cm)	Porcentaje
217	217,55	0,55	0,2534562
550	553	3	0,5454545
57	54,83	2,17	3,8070175
	Promedio		1,5353094

7. Conclusiones

Existe una gran versatilidad utilizando el método de calibración propuesto por Agapito ya que no necesita de objetos especiales para realizar la calibración. Como no hay necesidad de utilizar objetos especiales se evita la necesidad de preparar la escena con anticipación, lo que le da un gran atractivo al método.

El algoritmo es bastante simple y no requiere de una estimación inicial de los parámetros. La mayoría de algoritmos de calibración son iterativos, por lo que necesitan estimaciones iniciales y esto genera problemas de convergencia.

Los métodos de autocalibración tienen la ventaja de poder realizar el proceso de forma autónoma. Además evitan una de las posibles fuentes de ruido que mayor incidencia puede tener sobre el resultado que es la imprecisión en la posición de los puntos en el patrón. También hay que decir que los métodos de autocalibración utilizan modelos más simplificados de cámara que los métodos clásicos. Esto produce que estos métodos sean menos robustos en un mayor número de circunstancias, aquellas en las que hay influencia de los parámetros que no modelan, en especial, la distorsión.

En cuanto al planteamiento de las ecuaciones y la solución del sistema se puede concluir que es bastante sencillo, este algoritmo es lineal y los resultados demuestran que es efectivo. Por ser este un algoritmo lineal se puede decir que la velocidad de ejecución es alta comparada a los métodos no lineales donde es necesario realizar iteraciones hasta encontrar una solución precisa. Este método es rápido a comparación de los métodos no lineales y podría ser utilizado en aplicaciones de tiempo real.

Los resultados obtenidos muestran que, aunque los parámetros intrínsecos han permanecido constantes a lo largo de las secuencias, la estimación del centro del eje óptico presenta una alta inestabilidad, todos los valores deberían permanecer constantes durante el proceso de autocalibración. Una de las razones es porque se utilizó la cámara con enfoque automático, este enfoque automático pudo afectar el punto principal entre cada imagen utilizada. Por otro lado, los resultados del factor de escalado son mucho más estables.

Se concluye que el método funciona para valores de zoom que son distintos al mínimo y al máximo. Como lo explica Agapito en su artículo *Self-calibration of rotating and zooming camera*, existe un parámetro que afecta mucho los resultados de la autocalibración cuando la distancia focal es muy corta, el parámetro el cual hace referencia es el de la distorsión radial. La distorsión es la distancia existente entre la posición ideal dada por el sistema pinhole y la posición real que ocupa cada pixel en la imagen. La distorsión radial se aprecia mucho más cuando la distancia focal es corta, este es el caso del video presentado en 6.2.4, donde la distancia focal de la cámara era la mínima. La solución para que la autocalibración funcione de manera óptima en todos los niveles de distancia focal es modelar la distorsión radial [2].

En el video con el nivel máximo de zoom también se observó un error significativo en los resultados. Una de las explicaciones es expuesta en [2] donde se habla de ciertas configuraciones y de ciertos movimientos realizados por la cámara donde la autocalibración no funciona. Este tipo de configuraciones son llamados *degenerate configurations*. Dentro de estas *degenerate configurations* se encuentra la *near ambiguity*, donde se explica que para rotaciones pequeñas existe una ambigüedad entre la rotación y la distancia focal, esto hace que sea complicado para el algoritmo distinguir entre una rotación pequeña con una distancia focal grande y una rotación grande con una pequeña distancia focal. Al igual que con la distorsión radial, este problema es explicado en detalle en [2] con base en el artículo de Kahls, Triggs y Arstrom.[7].

Se encontró que el número de imágenes utilizadas para realizar la autocalibración es un factor muy importante para los resultados. Para que el método converja es necesario un número mínimo de imágenes. Este número mínimo de imágenes puede variar dependiendo de la calidad de las homografías así que no se puede dar un número exacto de imágenes.

En cuanto a la estimación de la orientación relativa de la cámara se observa como las curvas de rotación describen los movimientos realizados por la cámara de una forma muy precisa. La dispersión en los resultados no parece muy alta y el error es mínimo.

Hay que evitar realizar movimientos ambiguos. Como Hartley recomienda en su libro, se ha visto que ocurren degeneraciones en la calibración si el movimiento de la cámara es muy restringido, por ejemplo el

movimiento sobre un solo eje. El movimiento de la cámara no debe ser muy pequeño, o abarcar muy poca visión de la escena.

La aplicación de estimación de la distancia y tamaño de objetos funcionó bastante bien. Esto se debe a que los datos más precisos del método de Agapito son los de la matriz de Rotación.

Aunque el método de autocalibración propuesto por Agapito tiene ciertas limitaciones, como la restricción de la translación de la cámara PTZ y el uso de un mínimo número de imágenes, los resultados obtenidos son buenos y como ya se mostró, puede ser implementado para obtener medidas reales de una escena. Se propone que para mejorar los resultados se deben modelar las distorsiones radiales y tangenciales además de mejorar la calidad de las imágenes utilizando un instrumento de captura más avanzado al utilizado en este trabajo de grado.

8. Bibliografía

- [1] Agapito, L., Hartley, R., & Hayman, E.. Linear calibration of a rotating and zooming camera. *IEEE Conference on Computer Vision and Pattern Recognition* , 15-21. (1999)
- [2] Agapito, L., Hartley, R., & Hayman, E. Self calibration of a rotating with varying intrinsic parameters. *British Machine Vision Conference* , 105-114. (1998)
- [3] Agapito, L., Reid, I., Hayman, E., & Hartley, R. Self-calibration of zooming cameras from a single viewpoint: application to panoramic mosaicing. *Panoramic Vision* . (2001)
- [4] García Ocón, J.. Autocalibración y sincronización de múltiples cámaras PTZ. (2007)
- [5] Harris, C., & Stephens, M.. A combined corner and edge detector. *Proceedings of the 4th Alvey Vision Conference* , 147--151. (1988)
- [6] Hartley, R., & Zisserman, A.. *Multiple View Geometry in Computer Vision*. Cambridge University Press. (2003)
- [7] Kahl, F., Triggs, B., & Astrom, K.. Critical Motions for Auto-Calibration When some Intrinsic Parameters can vary. *Mathematical Imaging and Vision* , 131-146. (2000)
- [8] Kovesi., P. D.. *MATLAB and Octave Functions for Computer Vision and Image Processing*. School of Computer Science & Software Engineering, The University of Western Australia: <http://www.csse.uwa.edu.au/~pk/research/matlabfns>
- [9] Mery, D.. *Vision por Computador*. Santiago de Chile, Chile. (2004)
- [10] Sanchez, J. (n.d.). *Visión Tridimensional*. Universidad de Las Palmas de Gran Canaria, Gran Canaria, España.
- [11] *Sony EVI D70 Technical Manual*.
- [12] Zhang, Z.. A Flexible New Technique for Camera Calibration. *Microsoft Research* , 22. (2008)

9. ANEXOS

A.1 Listado de funciones.

A.2 Secuencias de imágenes.

A.3 EVI Series Software.

A.4 Código Desarrollado.

A.5 Videos Utilizados.

A.1 Listado de funciones.

Listado de funciones principales utilizadas en el algoritmo de autocalibración.

Nombre Función	Entradas	Salidas	Descripción
Principal	-	-	Este código es el principal del algoritmo de autocalibración. Pide como entrada el número de imágenes que se van a utilizar para realizar este proceso y llama las funciones necesarias para realizar la autocalibración. Las funciones que llama son las siguientes: Ransachomografias Resolvermatrices Matrizcalibracion Matrizdefinitiva Graficarresultados.
rgb2gray	im1,im2	I1,I2	Esta función, como su nombre lo dice, convierte una imagen RGB a escala de grises. Esto se logra eliminando el hue y la saturación de la imagen mientras se mantiene la luminancia.
Ransachomografias	I1,I2	H(n), p	Realiza el proceso descrito en el marco conceptual para encontrar la homografía. Primero encuentra la correspondencia de puntos usando el harris corner detector. Luego encuentra la correspondencia de puntos por correlación y luego se implementa RANSAC para estimar la mejor homografía con sus respectivos <i>inliers</i> . H(n) es la homografía resultante para la <i>n</i> ésima imagen y p es el porcentaje de puntos <i>inliers</i> que se obtuvieron.
Resolvermatrices	H	M	Esta función plantea las ecuaciones lineales que se van a resolver a partir de las homografías obtenidas anteriormente. La entrada es la homografía para cada par de imágenes y la salida es el conjunto de ecuaciones para cada homografía.
Matrizcalibracion	x,H	A	Función con la cual se encuentra la matriz de calibración. La entrada es parte del resultado del <i>singular value decomposition</i> , este resultado es la solución del sistema de ecuaciones planteado en la función Resolvermatrices. Esta solución es la variable x, H es la respectiva homografía entre la <i>n</i> ésima imagen y la de referencia. La salida A es la matriz de calibración que contiene los parámetros intrínsecos de la cámara.

Nombre Función	Entradas	Salidas	Descripción
harris	im, sigma, thresh, radius	Cim,r,c	<p>Algoritmo detector de esquinas desarrollado por (Kovesi.) Las entradas se explican a continuación:</p> <ul style="list-style-type: none"> • im: imagen que será procesada. • Sigma: Desviación estándar de <i>smoothing Gaussian</i>. Los valores típicos están entre 1-3. Este filtro elimina el ruido de la imagen. • thresh: Es el threshold utilizado para hacer la discriminación. • Radius: radio de la región considerada como non maximal suppression (opcional). Los valores típicos están entre 1-3 píxeles. Esto se realiza para suprimir toda la información en la imagen que no pertenece al máximo local. <p>Salidas :</p> <ul style="list-style-type: none"> • cim: Imagen binaria mostrando las esquinas. • r: coordenadas en x de los puntos. • c: coordenadas en y de los puntos.
matchbycorrelation	(im1, [r1';c1'], im2, [r2';c2'], w,dmax)	M1,m2	<p>Esta función genera las posibles correspondencias de los puntos previamente encontrados. Esto se realiza por correlación.</p> <p>Entradas:</p> <ul style="list-style-type: none"> • im1, im2: Imágenes. • p1, p2: Coordenadas de los puntos encontrados por el algoritmo de Harris. • w: Tamaño de la ventana con la cual se hace el barrido. • dmax: Máximo radio de búsqueda. Este valor puede ser omitido. <p>Salidas:</p> <ul style="list-style-type: none"> • m1, m2: Coordenadas de los puntos que tienen correspondencia.

Nombre Función	Entradas	Salidas	Descripción
ransacfithomography	(x1, x2, t);	H, inliers	<p>Ajusta una homografía utilizando RANSAC.</p> <p>Entradas</p> <ul style="list-style-type: none"> • x1: Matriz de 2xN o 3xN que contiene los puntos en coordenadas homogéneas. Si los datos son 2xN se asume que el factor de escala es uno. • x2: Matriz de 2xN o 3xN que contiene los puntos en coordenadas homogéneas para que $x1 \leftrightarrow x2$. • t: Distancia de threshold entre los puntos y el modelo para decidir si es un inlier o no. <p>Salidas:</p> <ul style="list-style-type: none"> • H: homografía. • Inliers: Una matriz con los elementos x1, x2 que se consideraron como inliers en el mejor modelo de Homografía.

Nombre Función	Entradas	Salidas	Descripción
RANSAC	x, fittingfn, distfn, degenfn s, t, feedback, maxDataTrials, maxTrials	M, inliers	<p>Ajusta un modelo utilizando el algoritmo de RANSAC.</p> <p>Entradas:</p> <ul style="list-style-type: none"> • x: Los datos con los cuales se está buscando ajustar un modelo. Es una matriz del tamaño [dx Npts] donde d es la dimensión de los puntos y Npts es el número de puntos. • fittingfn: Función al cual se le ajustará un modelo utilizando x, en este caso la función es la homografía. • distfn: Función que evalúa la distancia del modelo a los datos en x. Esta función evalúa las distancias entre los puntos y el modelo. El criterio para decidir si son <i>inliers</i> es t. • degenfn: Función que determina si un grupo de puntos producen un modelo degenerativo. • s: El mínimo numero de muestras de x necesitadas para ajustar el modelo. • t: Distancia de threshold entre el punto y el modelo para decidir si es un <i>inlier</i> o no. • feedback: Opcional, 0 o 1. Con 1 se muestra el número de intentos necesitados para ajustar un modelo al grupo de puntos x. • maxDataTrials: Máximo número de intentos con distintos grupos de puntos. Este parámetro es opcional y está ajustado a 100 como predeterminado. • maxTrials: Máximo número de iteraciones. Este parámetro es opcional y está predeterminado a 1000. <p>Salidas:</p> <ul style="list-style-type: none"> • M: el modelo que contiene el mayor numero de <i>inliers</i>. • <i>Inliers</i>: Matriz que contiene el número de elementos que fueron considerados como <i>inliers</i> en el mejor modelo obtenido.
svd	r,0	U,S,V	<p>Función de Singular Value Decomposition. Donde r es el sistema de ecuaciones escrito en forma matricial y al cual se le realiza el SVD. Del resultado se utiliza la matriz S para encontrar la solución del sistema de ecuaciones.</p>
chol	W	K	<p>Factorización de Cholesky con la cual se obtiene la matriz de calibración. La entrada es la cónica absoluta.</p>

Nombre Función	Entradas	Salidas	Descripción
Pantiltzoom	R	T	Función que descompone la matriz de rotación R en sus componentes pan tilt zoom. Se utiliza para poder graficar la rotación de la cámara.
Matrizdefinitva	DistanciaFocal, Px, Py, Skew	K	Obtiene el promedio de todas las matrices de calibración. La salida es la matriz promedio de calibración para todo el conjunto de imágenes utilizadas.
graficarresultados	A,R,N	-	Función que grafica los resultados respectivos de los parámetros intrínsecos y los parámetros extrínsecos. También se necesita el numero de imágenes que esta es la variable N.