



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

**“CALIBRACIÓN DEL SISTEMA CÁMARA-OJO PARA LA
RECONSTRUCCIÓN 3D DE ESTRUCTURAS EN IMÁGENES
DE RETINA”**

T E S I S

QUE PARA OBTENER EL GRADO DE:

**MAESTRO EN CIENCIAS
(COMPUTACIÓN)**

P R E S E N T A:

JAVIER ALEJANDRO ALDANA IUIT

**DIRECTORA DE TESIS:
CO-DIRECTOR DE TESIS:**

**DRA. MARÍA ELENA MARTÍNEZ PÉREZ
DR. ARTURO ESPINOSA ROMERO**

México, D.F.

2010



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Resumen

En este trabajo de tesis se presenta un sistema de reconstrucción tridimensional de vasos sanguíneos, presentes en la retina del ojo humano, con fines de visualización para asistencia a médicos. El sistema emplea imágenes de fondo de ojo capturadas en configuración de geometría epipolar para llevar a cabo el proceso de reconstrucción. La parte medular para la correcta reconstrucción es la calibración de la cámara, dicho proceso se realiza en cámaras convencionales mediante proyecciones de patrones conocidos, sin embargo, en el caso del sistema *cámara-ojo* esto no es posible debido a su naturaleza. Es por esto que la propuesta presentada consiste en dos diferentes formas de considerar los efectos de la óptica del ojo en la matriz de calibración de la cámara *fundus*.

Por una parte, se procede a calibrar la cámara *fundus* por métodos clásicos en tanto que la óptica ocular es involucrada tomando ciertas consideraciones en la captura de las imágenes, para minimizar los efectos indeseables de sus aberraciones. Entre estas consideraciones tenemos el uso de una lente de magnificación para corregir el astigmatismo, el ajuste del foco de la cámara para eliminar los errores refractivos (entiéndase miopía, hipermetropía y presbicia) y el cambio en el tamaño de la pupila de la cámara mediante el ajuste del diafragma, entre otras.

En contraparte, procedemos a calibrar el sistema *cámara-ojo* como un único sistema óptico mediante técnicas de auto-calibración, las cuales suponen que el sistema con el cual se adquirieron las imágenes no se encuentra disponible para su calibración. Haciendo uso de características en la escena capturada, estos métodos recuperan los parámetros intrínsecos de la cámara. En específico, los métodos de auto-calibración empleados en este trabajo son: las ecuaciones de Kruppa y el método de auto-calibración a partir de la matriz esencial. Se presenta un análisis comparativo de los resultados en reconstrucción con ambos métodos de auto-calibración y con las consideraciones en la captura.

Las reconstrucciones obtenidas mediante la auto-calibración por las ecuaciones de Kruppa resultan con menores índices de error, aunque calibrando únicamente la cámara *fundus* y minimizando las aberraciones del ojo se obtienen reconstrucciones de muy buena calidad. Los esqueletos en 3D de los vasos segmentados se presentan en su modelo de superficies.

Agradecimientos

Este trabajo de tesis no hubiera sido posible sin el apoyo de mis padres y de toda mi familia, que a pesar de la distancia siempre estuvieron conmigo.

A mi novia Karla por aguantar las largas jornadas de trabajo junto con los desvelos.

A mis tutores Elena y Arturo que sin su paciencia, consejos y apoyo no hubiera logrado terminar con exito este trabajo de investigación.

A CONACYT por otorgarme la beca de maestría para mi sustento y por parte del posgrado así como por parte del proyecto 83088 de la convocatoria de Ciencias Básicas 2007.

Al Dr. Rufino Díaz por brindarme amablemente su completa ayuda y asesoría académica en los momentos requeridos.

A Amalia, Lulú y Diana que se preocuparon siempre por mi, tanto en lo personal como en lo académico.

A Fernando, Jose Luis, David, Licho y Miguel que durante el tiempo que compartieron departamento conmigo hicieron más gratos los momentos difíciles.

Índice general

Índice general	VII
1. Introducción	1
1.1. Motivación	1
1.2. Conceptos de la reconstrucción 3D	2
1.3. El ojo humano	3
1.4. El sistema cámara-ojo y sus implicaciones	6
1.5. Estado del arte	8
1.6. Contribución y relevancia	10
1.7. Objetivos	11
1.8. Síntesis y organización de la tesis	11
2. Preprocesamiento	13
2.1. Notación	13
2.2. ADICOS	14
2.2.1. Polinomio de aproximación	15
2.2.2. Modos de mapeo	17
2.2.3. Desarrollo del sistema	19
2.3. GTM	24
2.3.1. Algoritmo de fuerza bruta	26
2.3.2. Algoritmo optimizado	27
2.3.3. Ejemplo de funcionamiento	28
3. Calibración de cámaras	31
3.1. Conceptos básicos	31
3.2. Algoritmo de Tsai	34
3.2.1. Conversión de unidades	34
3.2.2. Relación en coordenadas de la cámara	35

3.2.3. Obtención de parámetros	36
3.3. Algoritmo de Kanatani	36
3.3.1. Determinación del punto principal	37
3.3.2. Determinación de la distancia focal	39
3.3.3. Modelo de ruido	40
3.4. Algoritmo de Zhang	41
3.4.1. Estimación de homografías	42
3.4.2. Estimación de los parámetros de la cámara	44
3.5. Calib-tool de MATLAB	45
4. Óptica del ojo	49
4.1. Polinomio de aberración a tercer orden	49
4.2. Efecto de las aberraciones	54
4.2.1. Esfericidad	55
4.2.2. Coma	56
4.2.3. Astigmatismo y curvatura de campo	56
4.2.4. Distorsión	57
4.3. Ametropías	58
4.3.1. Miopía	60
4.3.2. Hipermetropía	61
4.3.3. Astigmatismo	61
4.4. Profundidad de campo	62
4.5. Minimización de aberraciones	63
5. Métodos de Auto-calibración	67
5.1. El uso de auto-calibración en el proceso global de reconstrucción	67
5.2. Estructura algebraica de la auto-calibración	68
5.3. Las ecuaciones de Kruppa	70
5.4. Auto-calibración a partir de la matriz esencial	72
5.5. Comparación entre los métodos	75
6. Procesos posteriores a la calibración	85
6.1. Geometría epipolar	85
6.2. Triangulación lineal	87
6.3. Modelo de superficies	88

7. Experimentación en la reconstrucción	91
7.1. Preprocesamiento a las imágenes fundus	92
7.2. Reconstrucción con calibración clásica	98
7.3. Reconstrucción con auto-calibración	108
7.4. Comparación de resultados	112
7.5. Estructuras representadas con volúmenes	116
8. Conclusiones y trabajo futuro	119
8.1. Conclusiones	119
8.2. Trabajo futuro	122
Bibliografía	125

Capítulo 1

Introducción

En este capítulo nos situaremos en el problema a resolver, describiendo los motivos por los cuales se desea reconstruir en un ambiente virtual de tres dimensiones, los vasos sanguíneos que se encuentran en la retina del ojo humano. Entiéndase por reconstrucción, para este trabajo de tesis, a la obtención de una aproximación de las superficies capturadas en las imágenes de la escena, es decir, a partir de los puntos observados se pretende recuperar la superficie que rodea a un volumen no completamente visible o medible.

Presentaremos breves descripciones de sus componentes, el funcionamiento de ojo humano y del sistema *cámara-ojo*. De igual forma mencionaremos el método utilizado para llevar a cabo la reconstrucción y cuáles son los inconvenientes al aplicarlo usando imágenes capturadas a través de este sistema. Revisaremos trabajos previos con diferentes enfoques de este problema para basar nuestra contribución al mismo. Por último presentaremos la organización del contenido de esta tesis.

1.1. Motivación

La vascularidad de la retina es la única en todo el cuerpo humano que puede ser observada a detalle directamente desde afuera, sin la necesidad de realizar daño a tejido sano. Las estructuras que se anteponen a la retina, en dirección a la parte anterior del globo ocular, son transparentes e incoloras, es por eso que podemos estudiarla *in vivo* sin procedimientos invasivos.

La reconstrucción 3D de la vascularidad retinal promete ser de gran utilidad para la valoración del estado de salud del ojo humano, debido a que numerosos padecimientos oculares

se manifiestan en sus etapas tempranas a nivel de retina. La presencia de *neo-vasos* es un ejemplo de un síntoma indicador de la presencia de una enfermedad en el ojo, como pueden ser la retinopatía diabética proliferante y la trombosis isquémica de las venas retinianas. La tortuosidad de los vasos sanguíneos y su grosor son parámetros que permiten detectar la presencia de la leucemia, y clasificar el grado de avance de la retinopatía hipertensiva [Bonafonte y Bonafonte, 2006].

Éstos, en conjunto, son ejemplos de indicadores morfológicos detectables en vasos sanguíneos a partir de imágenes conocidas como de *fondo de ojo*. La información implícita en la presencia de estos factores, es de suma importancia en el diseño de herramientas de detección de estas desviaciones a la morfología de la red sanguínea ocular para asistir a los médicos en la identificación y el diagnóstico temprano de enfermedades que amenazan al buen funcionamiento del sentido de la vista.

Otro factor importante que incentiva al análisis de la retina es que está íntimamente relacionada con el cerebro. Existe evidencia de que la microvascularidad retiniana y cerebral tienen propiedades morfológicas y fisiológicas en común; y por lo tanto, en cierta medida puede considerarse como un monitor del estado del cerebro [Patton *et al.*, 2005]. Podemos suponer que al realizar una valoración de la retina, se puede determinar la necesidad de realizar estudios más específicos para el cerebro. Lo anterior es sumamente conveniente pues los estudios de imagenología médica del cerebro, tales como la *tomografía computarizada* y la *resonancia magnética*, son excesivamente costosos en comparativa con el estudio de *fondo de ojo* que es barato si ya se cuenta con el equipo de adquisición, conocido como *cámara fundus*.

La herramienta desarrollada aquí brinda mayor información visual al médico tomando en cuenta que las imágenes de fondo de ojo no permiten percibir profundidad, ya que son proyecciones del mundo a un plano 2D. Al presentar una visualización de los vasos sanguíneos retinales en 3D se provee al médico de información de perspectiva permitiendo así detectar características no apreciables en las imágenes comunes.

1.2. Conceptos de la reconstrucción 3D

El proceso de reconstruir una escena 3D con base a dos o más imágenes 2D capturadas desde diferentes punto de vista, es decir, con desplazamientos de la cámara en relación al sistema de referencia del mundo, mejor conocido como *análisis de vistas múltiples*, involucra una serie de procedimientos aplicados a las imágenes de entrada con el objetivo de recuperar la posición de los puntos proyectados en las coordenadas del mundo. En esta aplicación se

usan únicamente dos imágenes de la escena a reconstruir y éstas son capturadas en una configuración llamada *geometría epipolar* (véase sección 6.1) y los requisitos para implementar este método son los siguientes:

1. Modelado de la distorsión radial introducida por el sistema de lentes de la cámara, para su corrección en las imágenes de entrada al proceso de reconstrucción.
2. Estimación de los parámetros intrínsecos y extrínsecos de la cámara.
3. Capturar dos imágenes de la escena en diferentes perspectivas y corregirles la distorsión con los mapas previamente calculados.
4. Encontrar las correspondencias entre los puntos característicos de la escena proyectados en ambas imágenes.
5. Eliminación de valores atípicos (*outliers*) de las correspondencias encontradas.
6. Calcular la matriz de proyección en perspectiva.
7. Calcular la posición del punto en el espacio para cada par de puntos en correspondencia.
8. Representar los esqueletos con superficies.

Los pasos anteriores serán vistos con mayor detalle en capítulos posteriores de esta tesis. Sin embargo, es importante mencionar que la parte medular de este proceso de reconstrucción es el mencionado en el punto número dos. A la estimación de los parámetros intrínsecos y extrínsecos de la cámara se le conoce como *calibración*. Su importancia radica en que los datos que provee permiten calcular la matriz de proyección que es la transformación proyectiva que mapea los puntos del plano de las imágenes a los puntos en el sistema coordenado del mundo (espacio tridimensional).

1.3. El ojo humano

El ojo humano promedio tiene un radio de 12 *mm* y la distancia entre la córnea y la región posterior del ojo (conocida como *longitud del eje pupilar*) mide entre 23 y 25 *mm*. Las regiones principales del ojo (basandonos en [Rajendra Acharya *et al.*, 2008]) se describen brevemente a continuación y se ilustran la Figura 1.1.

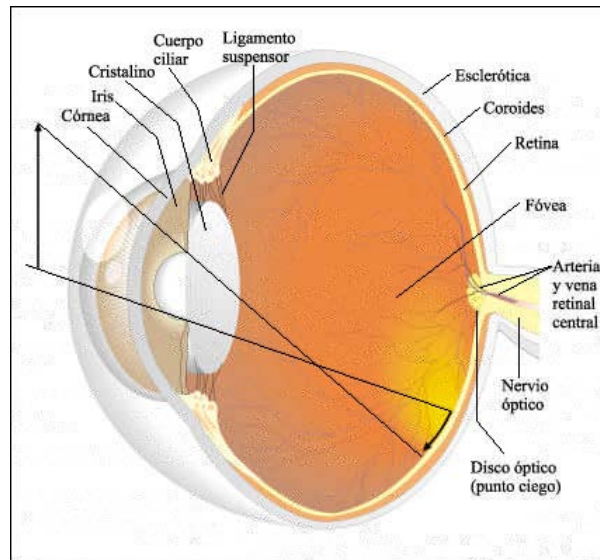


Figura 1.1: Anatomía del ojo humano. Figura extraída de <http://iescarin.educa.aragon.es>.

La córnea

Es un medio transparente con una composición del 78 % de agua. La córnea es elipsoidal por naturaleza tiene un diámetro vertical y horizontal de 11 y 12 *mm*, respectivamente. Su grosor no es uniforme, por el contrario es más delgada en la región central (530 μm en promedio) que en la periferia (710 μm en promedio). No contiene vasos sanguíneos y su función es refractar y transmitir luz al interior del ojo.

El humor acuoso

Consiste en una cámara que contiene fluido líquido, el cual provee de nutrientes a la vascularidad de la cornea y al cristalino. Su principal composición es agua. El flujo acuoso genera una presión intraocular de alrededor de 15 *mmHg* para un ojo sano.

El iris

Se conforma de tejido fibrovascular pigmentado conocido como *estroma*, el cual conecta a un conjunto de músculos contractores y dilatadores que modifican el tamaño de la pupila. Su función es controlar dicho tamaño de la pupila, ajustándolo a las condiciones de iluminación. En él se manifiesta el color de ojos de cada individuo. Con alta luminosidad la pupila mide como mínimo 2 *mm* de diámetro y en condiciones bajas de luz su diámetro mide 9 *mm*.

El cristalino

Es biconvexo con un grosor central entre 3.5 y 5 *mm*, sin acomodación, dependiendo de la edad del individuo. Su función es enfocar los rayos de luz que entran al ojo sobre la retina. Su composición es 65% agua y 35% proteína y es una región avascular. Su índice de refracción no es homogéneo, lo que permite corregir aberraciones. Su potencia dióptrica ($1/f$), es modificada por los músculos ciliares para enfocar a diferentes distancias (proceso de acomodación).

El humor vítreo

Es la región más grande del ojo. Consiste en una solución acuosa transparente (de 98.5% a 99.7% de agua) que llena el espacio comprendido entre el cristalino y la retina. Su longitud es de aproximadamente 16.5 *mm* y da la forma esférica al ojo.

La esclera

Es la capa exterior blanca del ojo y cubre cinco sextos de la parte posterior del globo ocular. Su grosor no es uniforme y su máximo es de 1 *mm* en el polo posterior. Su suministro de sangre es mínimo.

La corioide

Es una capa que se encuentra debajo de la esclera y es una estructura altamente vascularizada pues contiene un lecho capilar conocido como *coriocapilar*. Cuenta con el 85% del flujo sanguíneo ocular total.

La retina

Es la capa sensible a la luz y, por lo tanto, es la región en donde la energía luminosa es transformada en señales eléctricas neuronales. Está compuesta por células neuronales que recubren la parte posterior interna del globo ocular. Es la única parte del sistema nervioso central que puede ser vista directamente. Los principales tipos de células presentes en ella son: las fotorreceptoras (bastones y conos), las bipolares y las células ganglionares. Otras células retinales (horizontales, amacrinas y las neuronas interplexiformes) modifican e integran la señal antes de salir del ojo.

Cuando comparamos la retina central con la retina periférica encontramos que la zona de la fovea es considerablemente más gruesa que la retina periférica. Esto se debe a que en la fovea se encuentran densamente empaquetados gran cantidad de fotorreceptores (sobre todo conos), lo cual hace que incluso se desplacen lateralmente hacia fuera sus células bipolares y ganglionares asociadas. En la porción central de la retina existen sobre todo conos, mientras que en la retina periférica predominan los bastones. La nuclear externa, donde se encuentran los cuerpos celulares de los conos y bastones presenta un espesor semejante en toda la retina. Sin embargo en la retina periférica el número de cuerpos de bastones es muy superior al de los conos, y lo opuesto ocurre a nivel de la retina central. A nivel de la retina central los conos presentan unos axones oblicuos que se encuentran desplazados lateralmente respecto a sus pedículos sinápticos a nivel de la plexiforme externa.

La región periférica de la retina detecta movimientos bruscos sin percibir detalles pues los bastones (contienen rodopsina) son responsables de la visión en condiciones de baja luminosidad, presentando un pico de mayor sensibilidad hacia la longitud de onda de los 500 nm (luz verde azulada). El área central esta especializada en la actividad visual pues los conos son la base de la percepción del color. Estos contienen tres tipos diferentes de opsinas, una con mayor sensibilidad para las longitudes de onda largas (luz roja), otra que es sensible a las longitudes de onda medias (luz verde) y otra con mayor sensibilidad a las longitudes de onda cortas (luz azul).

Los vasos sanguíneos de la retina

Son los vasos que nutren la capa de la retina, entran al globo ocular a través del nervio óptico, y son los vasos de interés en esta tesis.

1.4. El sistema cámara-ojo y sus implicaciones

Las imágenes procesadas para llevar a cabo la reconstrucción son capturadas con cámaras especiales conocidas como cámaras *fundus* o de *fondo de ojo*, las cuales cuentan con un sistema óptico similar al de un microscopio. Cuando se realiza el estudio de fondo de ojo, el oftalmólogo utiliza un *lente de magnificación* que está en contacto directo con el globo ocular, específicamente con la cornea, y que cuenta con un conjunto de espejos que aumentan el campo visual del médico permitiéndole observar la periferia más alejada en la retina. Adicionalmente, los haces de luz que se proyectan en las imágenes pasan a través de las partes que conforman la óptica de la esfera ocular, es decir, los rayos también son refractados por el sistema óptico el ojo. A estos tres sistemas ópticos de naturaleza distinta, en conjunto, es

a lo que llamamos el sistema *cámara-ojo*. En la Figura 1.2 podemos observar un diagrama de la descripción anterior.

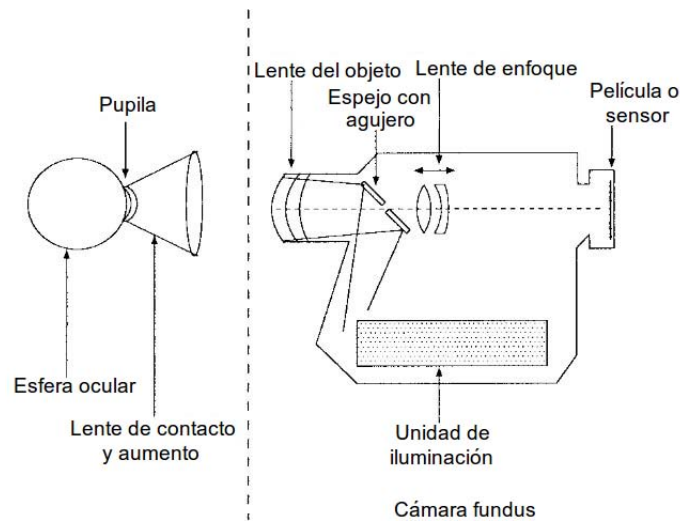


Figura 1.2: Componentes simplificados del sistema *cámara-ojo*. A la izquierda de la línea punteada se observa el globo ocular y el lente de magnificación en contacto directo. A la derecha de la línea punteada se presenta un diagrama de la cámara *fundus*, con sus principales componentes [Deguchi *et al.*, 1998].

Como se mencionó en la sección 1.2, sabemos que para reconstruir una escena por geometría epipolar necesitamos la calibración de la cámara, entonces al tratarse del sistema *cámara-ojo* tenemos que encontrar sus parámetros considerando los tres sistemas ópticos antes mencionados en conjunto. Sin embargo, solamente la cámara fundus y el lente de magnificación pueden ser calibrados por métodos clásicos pues es posible fotografiar a través de ellos patrones de referencia, lo cual no ocurre con la óptica del ojo pues en una prueba *in vivo* no podemos introducir en el un plano con un patrón geoméricamente conocido.

Con lo anterior podemos concluir que la calibración del sistema *cámara-ojo* no puede realizarse de la misma forma que con una cámara convencional debido a las implicaciones que conlleva involucrar al globo ocular. Con el fin de considerar al ojo en el cálculo de los parámetros intrínsecos del sistema completo debemos tener presente que la óptica del ojo, al no ser ideal y al variar de sujeto a sujeto, presenta aberraciones y éstas tienen que ser compensadas para minimizar sus efectos en las matrices de calibración de la cámara *fundus* y el lente de magnificación.

1.5. Estado del arte

Se han realizado diferentes trabajos (sin llegar a ser numerosos ya que se trata de una aplicación específica poco estudiada en el área de visión computacional y procesamiento digital de imágenes) con el objeto de reconstruir en 3D estas estructuras vasculares como es el caso de [Deguchi *et al.*, 1998], en los cuales se introduce la problemática de la imposibilidad de utilizar técnicas simples de visión estéreo, debido a que las imágenes de fondo son capturadas a través del sistema óptico del ojo y un lente de contacto usado por los médicos para aumentar el ángulo del campo visual (FOV por las siglas en inglés de *Field Of View*).

El trabajo anterior está basado en la hipótesis de que la superficie del fondo del ojo (la retina) es parte de una esfera. Se identifican los parámetros ópticos del sistema y se re proyectan los puntos de las imágenes para interceptar una superficie cuadrática. Se corrige la asimetría de las intersecciones de las líneas de vista (rayos definidos por los centros de las cámaras y los puntos en las imágenes) con la superficie cuadrática, ajustando esta última. Una vez caracterizada la superficie en la cual se encuentra posicionada la imagen del fondo, se re proyectan las imágenes estéreo para recuperar la forma esférica de la retina en las coordenadas del mundo. Como resultado tenemos un mapeo de textura en 3D de la vascularidad retinal sobre la superficie esférica antes mencionada, pero no así los vasos como estructuras en 3D.

Adicionalmente, la distorsión introducida por la cámara (y únicamente la cámara) es corregida en el proceso de calibración por métodos clásicos, y el sistema del lente de contacto (para aumento del FOV) en conjunto con el sistema óptico del ojo es considerado difícil de calibrar porque las propiedades ópticas del ojo humano son diferentes de individuo a individuo y son variantes con el tiempo. Este es el motivo por el cual se opta por modelarlos en conjunto como un único lente simple. De este modelo se calculan sus parámetros ópticos con base a la consideración de la forma esférica del fondo de ojo.

En la siguiente propuesta de [Deguchi *et al.*, 2000] para este mismo problema, se utiliza una serie de imágenes tomadas desde diferentes puntos de vista. De manera concreta, la determinación de la forma de la superficie cuadrática sobre la cual se re proyectará la imagen de la retina se realiza mediante las correspondencias de puntos característicos en cada una de las imágenes en la secuencia.

De igual forma se determina la pose relativa de cada cámara, respecto al globo ocular. Luego se identifican los parámetros ópticos del lente ocular de tal forma que la superficie cuadrática debería ser originada a partir de una superficie del fondo esférica. Como resulta-

do final obtenemos una superficie tridimensional amplia del fondo reconstruida a partir del conjunto de imágenes.

En trabajos más recientes como [Martinez-Perez y Espinosa-Romero, 2004], se propone realizar la reconstrucción 3D de la vascularidad ocular haciendo uso de métodos altamente robustos de segmentación. El método empleado en este trabajo, para la segmentación de los árboles de vasos retinales, se basa en el análisis de la representación multi-escala de las imágenes, propuesto en [Martinez-Perez *et al.*, 2007]. La información geométrica y morfológica de los árboles es extraída por la técnica propuesta en [Martinez-Perez *et al.*, 2002]. Con esta información, las estructuras en 3D de la red de vasos son generadas tomando en consideración las posibles superposiciones entre las ramas de los arboles sanguíneos, es decir, se recupera y respeta la relación espacial y de profundidad en el proceso de segmentación.

En relación con las implicaciones de realizar reconstrucción en el sistema *cámara-ojo*, los autores proponen calibrar de manera independiente la cámara fundus por métodos clásicos. Adicionalmente, la calibración del sistema completo, incluyendo la óptica del globo ocular, se obtiene mediante el método de auto-calibración llamado las ecuaciones de Kruppa [Lourakis *et al.*, 1999]. El método de auto-calibración es iterativo y converge dando como resultado una aproximación de la matriz de calibración del sistema global.

Otro enfoque de este mismo problema se presenta en [Chanwimaluang y Fan, 2007], trabajo en donde se propone el uso del modelo de cámara afin para el cálculo de la curvatura de la retina, debido a que se toman imágenes de ella con un campo visual de 30° de acuerdo a los protocolos emitidos por los Institutos Nacionales de la Salud (NIH por las siglas en inglés de *National Institutes of Health*) de los Estados Unidos, permitiendo asumir que la curvatura sufre de cambios muy pequeños en el área cubierta con esa apertura. El problema de considerar la óptica ocular es enfrentado primeramente incluyendo una medida de error de ajuste en su algoritmo, y posteriormente, compensando la distorsión con una función de costo.

Aunado a este modelo de cámara, se utiliza información de la geometría fundamental de la retina provista por imágenes múltiples de la misma en diferentes regiones. En concreto, se desarrolla un proceso de optimización limitada que considera tanto la forma geométrica de la retina humana como todos los elementos que causen una distorsión no-lineal significativa en las imágenes de fondo de ojo. Este método hace uso de una idea parecida a la utilizada en [Deguchi *et al.*, 1998, 2000] para el procedimiento de aproximar la imagen a una superficie cuadrática, con la añadidura de la corrección de distorsión.

Con esta misma herramienta de optimización (funciones de costo), en [Liu *et al.*, 2008] se estima la distancia focal de todo el sistema *cámara-ojo* por un método de auto-calibración. Para auto-calibrar se definen restricciones sobre los parámetros intrínsecos de la cámara, por ejemplo, la razón de aspecto se considera igual a 1, la asimetría del sensor (*skew*) se considera nula y los píxeles se suponen cuadrados. Además, por las condiciones de captura, la calibración se asume constante en las diferentes imágenes tomadas, a pesar del cambio del punto de vista.

1.6. Contribución y relevancia

En la actualidad, el estudio de los vasos sanguíneos retinales se ha hecho con el uso de métodos de procesamiento digital de imágenes, con el fin de segmentar los árboles vasculares en las imágenes de retinales, como en [Martinez-Perez *et al.*, 2007], trabajo en el cual se realiza un análisis multi-escala de las imágenes de fondo de ojo en tonos de grises. La información devuelta por el sistema implementado en [Martinez-Perez *et al.*, 2002] es utilizada para realizar una reconstrucción 3D de estos arboles vasculares usando geometría epipolar y auto-calibración, como se propone en [Martinez-Perez y Espinosa-Romero, 2004].

El sistema de reconstrucción propuesto en [Martinez-Perez y Espinosa-Romero, 2004] es la base de este proyecto de tesis para completar o refinar ciertas partes del proceso. El método adicionado en este proyecto es la corrección de distorsión radial en imágenes digitales previamente a la calibración de la cámara puesto que este proceso de corrección no se llevaba a cabo en la propuesta inicial, para lo cual se desarrolló el *Sistema Automático de Corrección de Distorsión* (ADICOS por sus siglas en inglés, *Automatic DIstortion COrrrection System*).

La incorporación de un algoritmo automático para la eliminación de valores atípicos en las correspondencias es otra aportación hecha en este proyecto de tesis. En específico utilizaremos el algoritmo llamado *Correspondencias por Transformación de Gráficas* (GTM por las siglas en inglés de *Graph Transformation Matching*) [Aguilar, 2006]. Por último presentamos la propuesta de minimizar los efectos de las aberraciones de la óptica ocular con un método indirecto para no afectar los parámetros internos de la cámara fundus y presentándolo junto con una comparativa de resultados en reconstrucción con dos métodos de auto-calibración.

1.7. Objetivos

- Conocer los tipos de aberraciones presentadas por el sistema óptico del ojo humano sano.
- Conocer las repercusiones de las patologías refractivas más comúnmente presentadas por las personas (miopía, hipermetropía y astigmatismo) en las imágenes de fondo de ojo.
- Comprobar si la óptica del ojo afecta significativamente al proceso de calibración del sistema *cámara-ojo* para la reconstrucción 3D de la vascularidad retiniana.
- Proponer un método para obtener la matriz de calibración del sistema *cámara-ojo* que incluya de manera intrínseca la calibración de la cámara fundus y del sistema de lentes del ojo.
- Comparar los resultados de reconstrucción usando métodos clásicos de calibración contra métodos de auto-calibración.

1.8. Síntesis y organización de la tesis

El problema a enfrentar consiste en reconstruir los vasos retinales a partir de un conjunto de imágenes de fondo de ojo. Debido a que el método de reconstrucción elegido para esta tarea necesita de un modelo de la cámara con la cual se adquirieron las imágenes. Los métodos más usados para calcular este modelo, requieren de imágenes de un patrón de referencia (métodos clásicos). Entonces, por la naturaleza del sistema *cámara-ojo* tenemos que descartar la idea de usar métodos clásicos si deseamos caracterizar el sistema completo, pues claramente no podemos introducir un patrón geométrico físico dentro del ojo. La propuesta alterna para modelar el sistema consiste en usar un método de auto-calibración, en los cuales se supone que no se cuenta con la cámara con la que se adquirieron las imágenes. Es por eso que dos métodos de auto-calibración fueron sujetos a pruebas de reconstrucción con el fin de saber cual de los dos era más robusto y apropiado para nuestra aplicación.

Colateralmente al modelado de la cámara, se desarrolló un conjunto de consideraciones en la toma de las imágenes para minimizar las aberraciones que introduce la óptica del ojo cuando se añade al sistema global. Las consideraciones fueron diseñadas con base al análisis del modelo de aberraciones a tercer orden en sistemas centrados de lentes con simetría de revolución. El motivo de minimizar las aberraciones es la suposición de que sin ellas, la

presencia de la óptica del ojo no causa cambios en los parámetros de la cámara *fundus*, permitiéndonos obtener reconstrucciones euclidianas, lo cual es nuestra meta final. Al conocerse el método de auto-calibración más conveniente para el problema, éste se emplea para modelar el sistema *cámara-ojo* usando imágenes capturadas bajo las condiciones establecidas para garantizar el correcto desempeño del proceso de auto-calibración y contar con la presencia mínima de las aberraciones de todo el sistema.

Para presentar a detalle las partes de este proyecto, se decidió organizar la tesis de la siguiente manera: En el Capítulo 2 se presentan las descripciones de los métodos usados para procesar las imágenes de fondo de ojo para extraer de ellas la información necesaria para reconstruir la vascularidad en un ambiente 3D. En el Capítulo 3 se detallan los fundamentos de calibración de cámaras, tres métodos para calibrar cámaras, y finalmente se presenta el sistema de calibración considerado el estado del arte en este tópico. En el Capítulo 4 se introducen las aberraciones en los sistemas ópticos como el ojo humano y las consideraciones que se toman para su minimización.

En contra parte a la calibración clásica, en el Capítulo 5 se muestran dos métodos de auto-calibración para identificar el más apropiado a esta aplicación y comprobar sus ventajas y desventajas respecto a los métodos clásicos de calibración. En el Capítulo 6 se describen los métodos usados para obtener la visualización final de las estructuras reconstruidas. Como parte medular, el Capítulo 7 presenta las pruebas realizadas para reconstruir en 3D usando la minimización de las aberraciones de la óptica ocular junto con la calibración de la cámara *fundus* comparándola con la reconstrucción basada en métodos de auto-calibración, es decir, incluyendo a todo el sistema *cámara-ojo*. Finalmente, en el Capítulo 8 presentamos los resultados obtenidos de la experimentación y las conclusiones generadas después del análisis de resultados. De igual forma se manifiesta el trabajo que a nuestra consideración queda por realizar en este tópico de investigación.

Capítulo 2

Métodos de preproceso para las imágenes

En este capítulo se describen los métodos utilizados para adecuar las imágenes de entrada para una correcta extracción de datos, con el fin de reconstruir los objetos de interés conservando su geometría original. En primer lugar, se presenta la notación usada en las ecuaciones que se incluyen en esta tesis. Posteriormente, se describe el funcionamiento del sistema usado para corregir la distorsión de las imágenes, y por último, el algoritmo para la eliminación de valores atípicos en las correspondencias entre imágenes.

2.1. Notación

Una imagen puede definirse como una función bidimensional I definida en una región compacta Ω de una superficie bidimensional, que adopta valores reales positivos. En el caso de las imágenes digitales, tanto el dominio Ω como el contradominio \mathbb{R}_+ son discretos. Por lo tanto, I es una función

$$I : \Omega \subset \mathbb{Z}^2 \rightarrow \mathbb{Z}_+; (x, y) \mapsto I(x, y), \quad (2.1)$$

donde, I se le conoce como intensidad o brillo (formalmente se le conoce como irradiancia) y (x, y) son coordenadas espaciales en un plano (imagen), es decir, la posición de un píxel. Las imágenes digitales se representa con matrices por su naturaleza discreta, es por eso que nos referiremos a una imagen como una matriz $\mathbf{I} \in \mathbb{Z}^{M \times N}$, donde M es el número de filas y N es el número de columnas.

La posición de un punto en un plano en el espacio se representa mediante el vector columna

$$\mathbf{u} = (u_1, u_2)^T \in \mathbb{R}^2, \quad (2.2)$$

y la posición de un punto en el espacio euclidiano tridimensional se representa con el vector columna

$$\mathbf{x} = (x_1, x_2, x_3)^T \in \mathbb{R}^3, \quad (2.3)$$

a los cuales se les añadirá eventualmente un subíndice como referencia, según convenga al tema tratado.

En el caso de gráficas, definiremos \mathbf{ab} como la arista que une a los vértices $\mathbf{a} = (a_1, a_2)^T$ y $\mathbf{b} = (b_1, b_2)^T$, dichos vértices son puntos que se encuentran en un mismo plano pues las gráficas con las que trabajaremos son bidimensionales.

Ahora definiremos la notación de las operaciones vectoriales. Sean \mathbf{a} , \mathbf{b} y \mathbf{c} dos vectores pertenecientes a \mathbb{R}^N , el producto punto o escalar de dichos vectores se indica con la expresión:

$$(\mathbf{a}, \mathbf{b}), \quad (2.4)$$

y el producto vectorial se indica con la expresión:

$$\mathbf{a} \times \mathbf{b}, \quad (2.5)$$

y el triple producto escalar se expresa como:

$$|\mathbf{a}, \mathbf{b}, \mathbf{c}| = (\mathbf{a}, \mathbf{b} \times \mathbf{c}). \quad (2.6)$$

2.2. Sistema automático de corrección de distorsión ADICOS

La distorsión es un tipo de aberración monocromática presente en los sistemas de lentes. Este fenómeno ocasiona que las partes de la escena proyectadas en la imagen, sufran de magnificaciones diferentes. Si la distorsión tiene patrón radial, a menor distancia de los puntos de la escena al eje óptico de la cámara, mayor será la magnificación en la proyección, o viceversa. A diferencia de los otros tipos de aberraciones ópticas, la distorsión no causa emborronamiento en la imagen, sino únicamente deforma el objeto capturado. Las imágenes de fondo de ojo, con las cuales trabajaremos, presentan alto grado de distorsión que es necesario corregir, si se desea reconstruir objetos con su geometría verdadera. El tema de aberraciones se presenta con mayor detalle en el capítulo 4.

El sistema automático de corrección de distorsión *ADICOS* es una herramienta para la corrección de distorsión radial introducida por los sistemas de lentes en las cámaras convencionales. La supresión de esta aberración es muy importante debido a que los objetos proyectados en las imágenes se encuentran deformados por consiguiente si son reconstruidos no contarán con su geometría verdadera. Se debe dar importancia al modelado de las distorsiones debidas a las lentes para aplicaciones de visión, especialmente cuando la cámara se desea utilizar como dispositivo de medición. Para el desarrollo de este sistema nos basamos fundamentalmente en [Asari *et al.*, 1999; Helferty *et al.*, 2001], porque el método propuesto modela la distorsión de cámaras de endoscopía. Esta distorsión presenta un patrón radial muy similar al introducido por las cámaras de fondo de ojo, dispositivo con el cual se adquieren las imágenes usadas en esta aplicación.

El objetivo principal de *ADICOS* es modelar automáticamente la distorsión radial de una cámara, usando como datos de entrada una sola imagen de un patrón de referencias plano capturada bajo estrictas condiciones. Se pretende obtener un polinomio que caracterice la deformación ocasionada por la distorsión. El producto final son mapas de redireccionamiento de píxeles que eliminan la aberración de las imágenes. Detalles del modelado y corrección hechas por el sistema se presentan en secciones siguientes.

2.2.1. Polinomio de aproximación

En el modelo de cámara simple llamado de *pin-hole*, un punto en la escena (\mathbf{x}) es proyectado como un punto en el plano de la imagen (\mathbf{u}). El rayo que va del punto en el mundo al plano de la imagen nunca es desviado en su trayecto, es decir, el punto en la escena, el centro de la cámara y la imagen del punto son colineales. Este modelo (mostrado en la Figura 2.1) se describe en las ecuaciones, conocidas como proyección en perspectiva, presentadas a continuación:

$$u_1 = \frac{f}{x_3}x_1, \text{ y } u_2 = \frac{f}{x_3}x_2, \quad (2.7)$$

donde, el plano de la imagen se halla en f , la distancia focal de la cámara si el objeto se encuentra en el infinito. Las ecuaciones (2.7) son buenas aproximaciones si f es pequeña. Por otra parte, es importante notar que en la cámara *pinhole* la magnificación es negativa, por lo que la escena capturada aparece invertida en la imagen. Entonces, las ecuaciones (2.7) hacen referencia a la imagen virtual (no invertida) y en adelante seguiremos haciendo referencia a esta imagen.

En la realidad este modelo no se cumple, ya que en las cámaras convencionales el factor de magnificación es una función no lineal cuya variable independiente es la distancia del

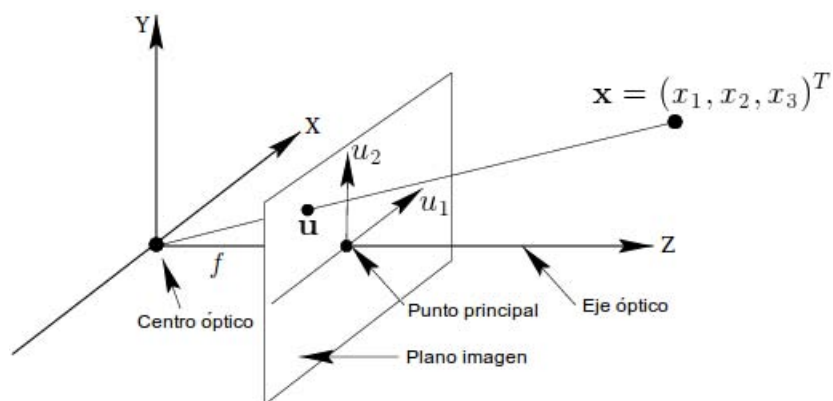


Figura 2.1: Modelo de cámara ideal conocido como *pin-hole* [Hartley y Zisserman, 2004].

punto en la escena al eje óptico y la variable dependiente es el factor de escala a multiplicar por la posición de dicho punto. En otras palabras, el punto en la escena, el centro de la cámara y el punto imagen no son colineales debido a la distorsión introducida por el sistema de lentes de la cámara (la cual no es ideal).

La meta común de los diferentes métodos para modelar la distorsión es encontrar los coeficientes del polinomio de aproximación que corrige esta no-linealidad en el factor de escala. En *ADICOS*, el polinomio está en coordenadas cartesianas ya que es lo más común en tareas de visión computacional como son: registro de imágenes, transformaciones afines, homografías, entre otras. Adicionalmente, se implementa el modelo polinomial a tercer orden. El contradominio de este polinomio contiene las nuevas posiciones de los píxeles en la imagen resultante libre de distorsión.

La motivación para elegir a tercer grado el polinomio de aproximación radica en que la literatura en óptica geométrica indica que el análisis de las aberraciones en los sistemas de lentes con simetría de revolución se tiene que hacer fuera de la aproximación paraxial, es decir, es necesario usar una aproximación a tercer orden. Debido a la simetría de revolución, el polinomio tiene que ser de orden impar y específicamente a tercer orden (polinomio de tercer grado). De acuerdo a lo anterior, *ADICOS* emplea un modelo polinomial de la forma mostrada en las siguientes ecuaciones:

$$u'_1 = \sum_{i=0}^m \sum_{j=0}^m a_{ij} u_1^i u_2^j, \quad (2.8)$$

$$u'_2 = \sum_{i=0}^m \sum_{j=0}^m b_{ij} u_1^i u_2^j, \quad m = i + j \leq 3, \quad (2.9)$$

donde, a_{ij} y b_{ij} son los coeficientes del modelo polinomial para calcular las nuevas coordenadas u'_1 e u'_2 , respectivamente.

Evaluando las posiciones de píxeles de una imagen de igual tamaño al de las capturadas por la cámara analizada, se generan tablas de memoria donde se almacenan las nuevas posiciones de los píxeles para compensar la distorsión del sistema de lentes. A estas tablas se les conoce como mapas. Los mapas obtenidos para una cámara, pueden ser reutilizados para suprimir la distorsión de todas las imágenes provenientes de la misma cámara, siempre y cuando el foco no haya sido modificado de la posición donde se encontraba cuando se realizó el modelado.

2.2.2. Modos de mapeo

Entendamos por modos de mapeo, al criterio seguido para aplicar los mapas almacenados en las matrices \mathbf{M}_x y \mathbf{M}_y denominadas funciones de transferencia de mapeo (TMF por las siglas en inglés de *Transfer Mapping Function*). Las TMFs son generadas de manera automática por ADICOS tras haber modelado la distorsión presente en las imágenes. De igual forma, se involucra el sentido de la transformación con el cual fueron calculados los mapas, es decir, al realizar el ajuste del polinomio se debe decidir qué imagen será nuestro origen y cuál nuestro destino de transformación. En este caso en particular, puede calcularse un mapa de transformación que nos lleve de la imagen distorsionada a la imagen correcta o viceversa.

Al modo de mapeo implementado en este sistema lo denominamos *mapeo inverso*, pues en esencia consiste en aplicar el mapa de transformación de la imagen corregida (puntos del patrón sin distorsión) a la imagen distorsionada (puntos del patrón distorsionados), es decir, el mapa inverso. Aplicamos el mapa inverso porque queremos asignar a los píxeles de la imagen corregida los valores de intensidad de la imagen distorsionada. Las direcciones son dictadas por el mapa, para cada uno de los píxeles de la imagen resultante, evitando así los huecos o píxeles con valor de intensidad indeterminado (véase Figura 7.1(b) y (e)). Estos huecos se presentan como resultado del cambio en el tamaño de la imagen original

debido a la transformación correctiva por mapeo directo y a que las direcciones almacenadas en los mapas son calculadas usando interpolación del vecino más cercano. Sin embargo, la introducción de artefactos en las imágenes corregidas al “estirar ó encoger” la imagen original no se resuelve usando únicamente el mapeo inverso.

Para lo anterior se adiciona la implementación de una interpolación bilineal como se presenta en [Press *et al.*, 1995]. Se toman los cuatro vecinos más cercanos a la posición del píxel bajo análisis indicada por los mapas. En la Figura 2.2 tenemos una ilustración de los píxeles involucrados en la interpolación. El símbolo “+” indica la posición del sub-píxel bajo análisis.

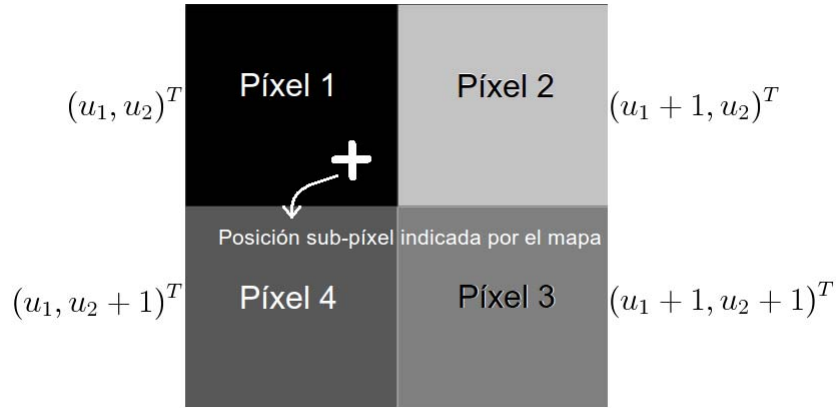


Figura 2.2: Vecindad de píxeles usada en la interpolación bilineal.

Siguiendo la definición de una interpolación bilineal, procedemos a realizar una suma ponderada de los niveles de gris de los píxeles en la vecindad para obtener la intensidad del píxel reposicionado. La imagen interpolada resultante \mathbf{I}' se obtiene aplicando la interpolación a cada uno de sus píxeles, tomando los valores de gris de la imagen \mathbf{I} , como se muestra en la siguiente ecuación:

$$\mathbf{I}'(u_1, u_2) = (1 - r)(1 - s)\mathbf{I}(u_{m1}, u_{m2}) + r(1 - s)\mathbf{I}(u_{m1} + 1, u_{m2}) + rs\mathbf{I}(u_{m1} + 1, u_{m2} + 1) + (1 - r)s\mathbf{I}(u_{m1}, u_{m2} + 1), \quad (2.10)$$

donde,

$$u_{m1} = \lfloor \mathbf{M}_x(u_1, u_2) \rfloor, \quad u_{m2} = \lfloor \mathbf{M}_y(u_1, u_2) \rfloor, \\ r = \mathbf{M}_x(u_1, u_2) - \lfloor \mathbf{M}_x(u_1, u_2) \rfloor \quad \text{y} \quad s = \mathbf{M}_y(u_1, u_2) - \lfloor \mathbf{M}_y(u_1, u_2) \rfloor.$$

En la ecuación (2.10), $\mathbf{u}_m = (u_{m1}, u_{m2})^T$ corresponde a la posición almacenada en los mapas de corrección correspondiente al píxel 1 de la vecindad. Por último, r y s son los factores de ponderación usados en la interpolación y representan el complemento de la distancia normalizada al píxel de referencia respecto a las abscisas y a las ordenadas, respectivamente.

Como es de esperarse, las imágenes resultantes presentan un nivel de difuminación (*blur*) debido a la interpolación bilineal. El efecto de emborronamiento es muy ligero y casi imperceptible, sin embargo, este proceso soluciona el problema de los artefactos después de la corrección de distorsión. La Figura 2.3(a) muestra la imagen corregida con mapeo inverso y la Figura 2.3(b) la imagen corregida con interpolación bilineal.

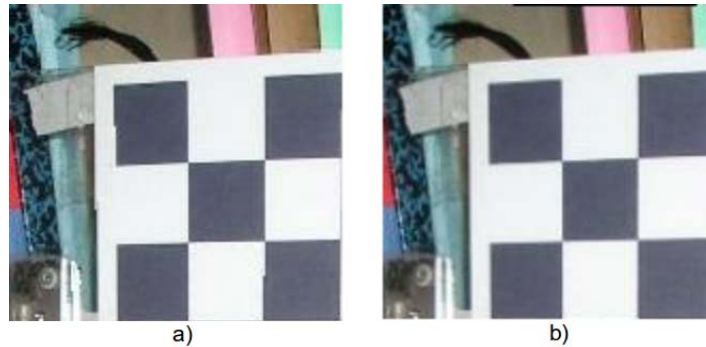


Figura 2.3: Resultado de mapeo inverso con interpolación bilineal para eliminar los artefactos del procesamiento. (a) Imagen corregida con mapeo inverso sin interpolación y (b) la imagen corregida con mapeo inverso usando interpolación bilineal.

2.2.3. Desarrollo del sistema

El funcionamiento de *ADICOS* se basa en un conjunto de algoritmos de procesamiento de imágenes y posteriormente en métodos numéricos de ajuste y minimización de errores, los cuales son descritos a continuación.

Imágenes de entrada

Consisten en imágenes de un *patrón de referencia* que consta de una rejilla rectangular de puntos posicionados en un plano. Los puntos equidistan unos con otros vertical y horizontalmente en 8 *mm*. El patrón se fotografía garantizando que el eje óptico sea perpendicular al plano del patrón y que las filas (y columnas) estén orientadas respecto a la horizontal (y

a la vertical) de la imagen resultante, lo cual se consigue mediante un montaje experimental como el mostrado en la Figura 2.4.

El patrón se define como una función discreta $g(\mathbf{u})$ donde, $\mathbf{u} = (u_1, u_2) \in \mathbb{Z}_+^2$ y $g \in \mathbb{Z}_+$, que representa el muestreo cada $\Delta x, \Delta y = 8mm$ de una función escalon unitario bidimensional, que es continua en el espacio.

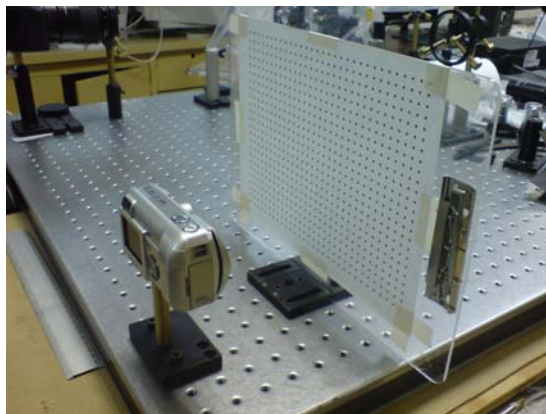


Figura 2.4: Montaje experimental para tomar imágenes del patrón de referencia.

Procesamiento de imágenes

Una vez capturadas las imágenes de la rejilla de puntos se procede a segmentarla con el fin de trabajar únicamente con las coordenadas de los centros de gravedad de los puntos. Para la segmentación se usan los siguientes métodos.

- **Remoción de fondo.** Restamos una imagen del fondo a la imagen del patrón de puntos. Para imágenes que no requieren enmascaramiento, la imagen de fondo consiste en una aproximación obtenida mediante el filtrado lineal RMS (*Root Media Square*) de la imagen del patrón. Y para las imágenes que requieren enmascaramiento, es necesario tomar otra imagen sin la presencia de los puntos del patrón.
- **Binarización por umbralización global.** El valor de este umbral es calculado automáticamente con el algoritmo propuesto en [Otsu, 1979], pues la imagen consiste en puntos negros con fondo blanco, es decir, el histograma solo tiene dos modas y como las condiciones de iluminación del patrón de referencia mejoran el contraste de las imágenes, hay poca presencia de ruido.

El umbralaje se define de la siguiente forma:

$$\mathbf{I}'(u_1, u_2) = \begin{cases} 255 & \mathbf{I}(u_1, u_2) \leq thr \\ 0 & \mathbf{I}(u_1, u_2) > thr \end{cases}, \quad (2.11)$$

donde, thr es el umbral calculado automáticamente.

- **Supresión de ruido.** La binarización introduce ruido sal y pimienta el cual es eliminado con un filtro no lineal como el filtro mediana. Dicho filtro fue elegido pues elimina los valores extremos presentes en la ventana muestreada, que en nuestro caso son los cambios fuertes en la forma circular de los puntos segmentados. Lo anterior modifica la posición de los centros de gravedad, lo que es indeseable.

El efecto de las operaciones y filtros aplicados a las imágenes del patrón de puntos se ilustra en la Figura 2.5, para las dos clases de imágenes de fondo, es decir, para la obtenida por el filtro pasa-bajas o por el capturado con la cámara directamente. De izquierda a derecha se muestran los resultados en cascada obtenidos por las operaciones con las imágenes. Primero se presenta la sustracción de la imagen de fondo a la imagen a segmentar, ilustrada con el símbolo “−”, y a la derecha del símbolo “=” tenemos el resultado de esta operación. Posteriormente, se presenta el resultado de la binarización por umbralaje global y finalmente, en el extremo derecho, se muestra el resultado de aplicar el filtro mediana a la imagen binaria.

Estimación del centro de distorsión

Las coordenadas de los centros de gravedad de los puntos del patrón segmentado, son calculados promediando las posiciones de todos los píxeles pertenecientes a un mismo punto de la rejilla. Para agrupar los píxeles por cada uno de los puntos, se recorre la imagen binaria en búsqueda de píxeles con valor 255. Cuando se encuentra un píxel de interés, se almacena su posición y se visita a los píxeles dentro de su vecindad 8. Si se encuentra otro píxel con valor de intensidad de 255, se introduce su posición en una cola para visitar su respectiva vecindad 8 más adelante. Cuando la cola queda vacía se promedian las posiciones almacenadas previamente. El algoritmo termina cuando se recorre completamente la imagen y no se encuentra más píxeles de interés. Una vez obtenidos los centros de gravedad en la imagen distorsionada, se genera la rejilla de puntos libre de distorsión y se superpone a la rejilla distorsionada, mediante los siguientes procesos:

- **Detección del cambio de signo en la curvatura.** Primeramente, generamos una base de datos con las coordenadas de los puntos agrupados por filas y columnas, en orden lexicográfico (de izquierda a derecha y de arriba hacia abajo) y descendente (de

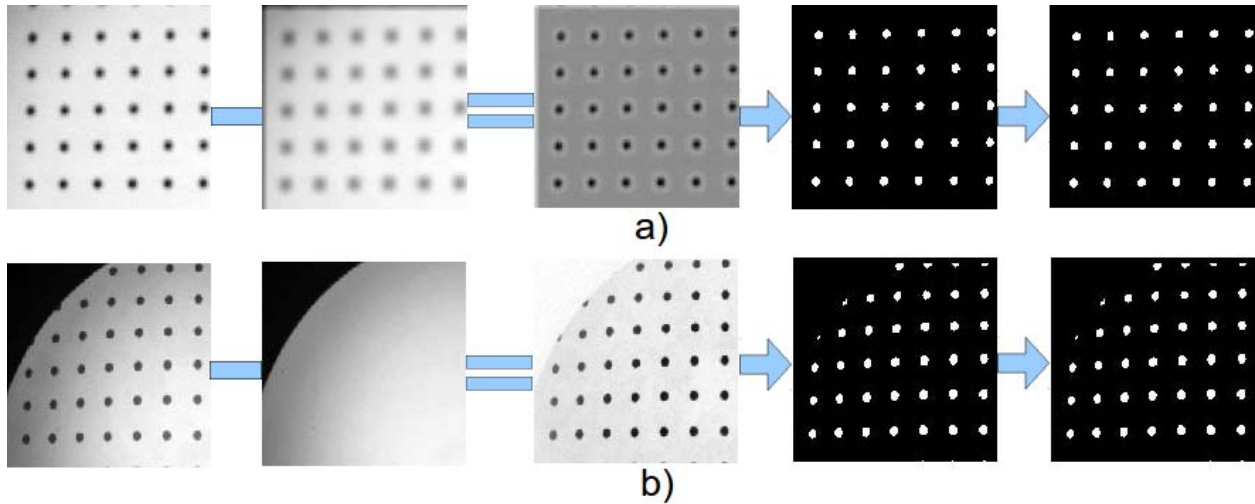


Figura 2.5: Secciones de imágenes de las diferentes etapas de la segmentación de los puntos del patrón. Empezando por la izquierda con la imagen original, imagen de fondo, resultado de la sustracción del fondo, binarización y finalmente, el resultado de aplicar el filtro mediana. La secuencia (a) corresponde a la imagen que no requiere enmascaramiento y la secuencia (b) a la que sí lo requiere, como es el caso de las imágenes de fondo de ojo.

arriba hacia abajo y de izquierda a derecha), respectivamente.

Posteriormente, para estimar el centro de distorsión primero se realiza una regresión lineal para encontrar un polinomio de segundo grado ($h(x)$) por cada fila y columna, que mejor aproxime la función que pasa por todos los puntos del conjunto. Obtenida esta función, aplicamos la siguiente ecuación, propuesta en [Asari *et al.*, 1999], empleada para el cálculo de la curvatura (Q) de una línea contando con su polinomio de aproximación (ver Figura 2.6).

$$Q_{rn} = \frac{\left. \frac{\delta^2 h_{rn}(x)}{\delta x^2} \right|_{(x_{rn}, y_{rn})^T}}{\left\{ 1 + \left(\left. \frac{\delta h_{rn}(x)}{\delta x} \right|_{(x_{rn}, y_{rn})^T} \right)^2 \right\}^{3/2}}, \quad (2.12)$$

donde, el subíndice r significa que estamos analizando la curva de ajuste que pasa por los puntos pertenecientes a una fila, si se tratase de columnas el subíndice sería c . n es el índice de las dos filas (o columnas) más cercanas al centro de distorsión, por lo

tanto, $n = 1, 2$. El punto $(x_{rn}, y_{rn})^T$ nos indica que a pesar de que las operaciones para calcular la curvatura están basadas en la función unidimensional $h_{rn}(x)$, el dominio en el cual estamos trabajando es bidimensional, y corresponde a la imagen $(\mathbf{I}(x_{rn}, y_{rn}))$.

Cuando se encuentra un cambio en el signo de la curvatura, al pasar de una fila (columna) a la siguiente se asigna a estas filas (columnas) los índices $n = 1$ y $n = 2$, para denotar que entre ellas se encuentra la fila (columna) que pasa por el centro de distorsión. Las filas y columnas marcadas se pueden observar en la Figura 2.6.

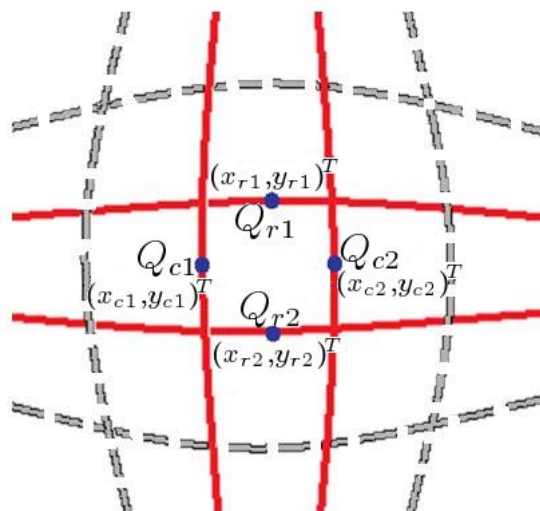


Figura 2.6: Filas y columnas más cercanas al centro de distorsión

- Determinación del centro de distorsión.** El centro de distorsión $(x_d, y_d)^T$ se estima mediante una interpolación lineal de las curvaturas Q_{r1} , Q_{r2} , Q_{c1} y Q_{c2} , según muestran las ecuaciones, propuestas en [Asari *et al.*, 1999], mostradas a continuación:

$$x_d = \frac{Q_{c1}x_{c1} + Q_{c2}x_{c2}}{Q_{c1} + Q_{c2}}, \quad (2.13)$$

$$y_d = \frac{Q_{r1}y_{r1} + Q_{r2}y_{r2}}{Q_{r1} + Q_{r2}}, \quad (2.14)$$

donde, los puntos $(x_{rn}, y_{rn})^T$ y $(x_{cn}, y_{cn})^T$ corresponden a los puntos sobre las líneas de cambio de signo de la curvatura, situados justo en medio de las intersecciones de las

curvas, véase la Figura 2.6, pues asumimos que son los puntos más cercanos al centro de distorsión real.

Cálculo de los mapas para corrección de distorsión

Generamos una rejilla regular (sin distorsión) de puntos, con una distancia de separación igual a la presentada entre los cuatro puntos (en la rejilla original) más cercanos al centro de distorsión. Los puntos de ambas rejillas se ponen en correspondencia para generar el sistema de ecuaciones de los coeficientes del polinomio de aproximación. Los coeficientes del polinomio de tercer grado que modela la distorsión, expresados en las ecuaciones (2.8) y (2.9), se calculan resolviendo el sistema por el método de mínimos cuadrados.

Posteriormente, se generan las TMFs en tablas que almacenan los resultados de la evaluación del polinomio antes obtenido. En primera instancia, se genera el *mapa directo* que transforma los puntos distorsionados a los puntos correctos, para conocer las dimensiones de la imagen corregida. Después, calculamos el *mapa inverso* (mapea los puntos correctos a sus posiciones distorsionadas) el cual, como se comentó en la sección 2.2.2, nos sirve para corregir la imagen original evitando la presencia de “huecos” y permitiendo adicionar un proceso de interpolación.

Para finalizar esta sección presentamos un ejemplo del resultado de corrección de distorsión con *ADICOS* al patrón de puntos, en la Figura 2.7. Es importante mencionar que una vez que se modela la distorsión de una cámara, es decir, se calculan los TMFs, es posible aplicarlos a todas las imágenes capturadas con esa misma cámara y la misma configuración de lentes. Por lo tanto, en el caso particular de la cámara de fondo de ojo, la cual cuenta con diferentes ángulos de campo visual, es necesario modelar la distorsión para cada uno de ellos. Esto se debe a que cada campo visual utiliza un conjunto diferente de lentes. Ejemplos de correcciones hechas a imágenes de fondo de ojo se mostrarán en la sección 7.1.

2.3. Correspondencias por transformación de gráficas GTM

Como se verá a mayor detalle en la sección 6.1, la reconstrucción por geometría epipolar requiere puntos característicos en correspondencia entre ambas imágenes. Los puntos característicos usualmente son extraídos por métodos como filtros espaciales detectores de esquinas, como el detector de Harris presentado en [Harris y Stephens, 1988], o por algoritmos más sofisticados y robustos como SIFT [Lowe, 1999]. Contando con los puntos característicos en

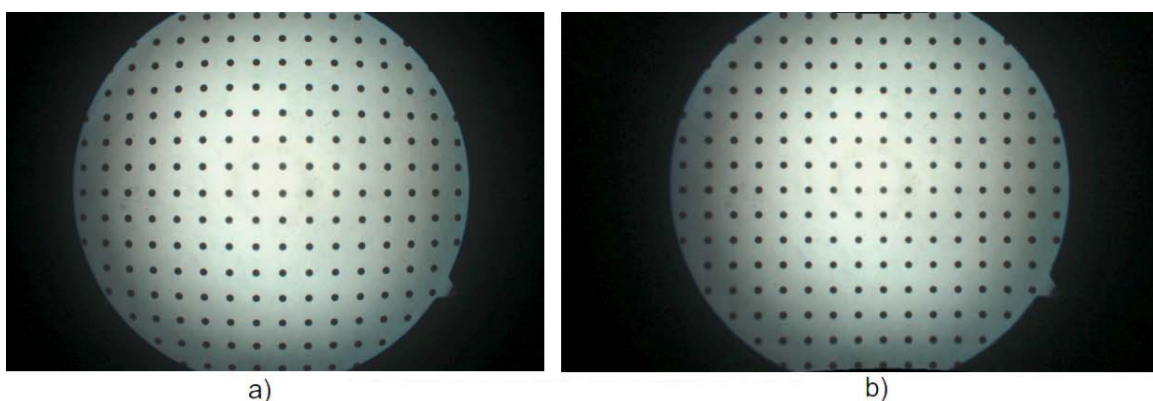


Figura 2.7: Ejemplo de los resultados de *ADICOS*. En (a) tenemos la imagen original del patrón y en (b) tenemos la imagen libre de distorsión.

ambas imágenes se procede a realizar un proceso de emparejamiento, es decir, a cada punto en la primera imagen se le relaciona con otro punto en la segunda imagen ambos correspondiendo al mismo punto de la escena. Lo anterior nos conduce a un problema de búsqueda del vecino más cercano basada en comparaciones de descriptores, como ejemplo tenemos el uso de la distancia euclidiana entre los vectores (descriptores). La búsqueda puede acelerarse usando algoritmos como el de los árboles-*kd* presentado en [Bentley, 1975]. Por otra parte, si recurrimos a búsquedas aproximadas para disminuir el tiempo de procesamiento, existe la alternativa presentada por el algoritmo *Best Bin First* (BBF) en [Beis y Lowe, 1997], el cual es 100 veces más rápido que la búsqueda exhaustiva, sin embargo, presenta un 5% de soluciones que no resultan ser las óptimas.

En la práctica, varias de las correspondencias encontradas por los algoritmos antes mencionados son erróneas, es decir, representan valores atípicos (del inglés *outliers*). Para la reconstrucción por geometría epipolar estas correspondencias erróneas son sumamente perjudiciales, al igual que en otros procesos como el cálculo de transformaciones entre imágenes. Es por eso que es crucial su eliminación y para ello existen algunos métodos como son *Random Sample Consensus* (RANSAC) propuesto en [Fischler y Bolles, 1981] y *Least Median of Squares* (LMedS), el cual puede estudiarse a detalle en [Rousseeuw y Leroy, 1987].

En este trabajo de tesis se eligió trabajar con el algoritmo llamado *Correspondencias por Transformación de Gráficas* [Aguilar *et al.*, 2007] (GTM por las siglas del inglés, *Graph Transformation Matching*) por su fácil implementación y porque se ha demostrado que el algoritmo es robusto al ruido de correspondencias erróneas. Para un 54% de correspondencias

iniciales incorrectas, encuentra un 40% de las correspondencias correctas con respuestas menores a 2 errores. Detalles del algoritmo se presentan a continuación.

2.3.1. Algoritmo de fuerza bruta

El algoritmo inicia usando como datos de entrada un emparejamiento inicial de puntos característicos en ambas vistas del objeto, calculados por el método de correlación. Se cuenta con las posiciones (u_{i1}, u_{i2}) y (u'_{i1}, u'_{i2}) de cada uno de los puntos en los respectivos sistemas de referencia de las imágenes. La etapa inicial se conoce como *verificación* y realiza un proceso de refinación de las correspondencias iniciales, es decir, elimina los falsos positivos comparando una representación estructural, que describe a la escena, iterativamente.

Se construyen en total dos gráficas no dirigidos y sin pesos, uno para cada imagen. Estos son construidos usando a los píxeles de los puntos característicos como *vértices*. La conexión (*arista*) entre dos vértices, sean $\mathbf{u}_i = (u_{i1}, u_{i2})^T$ y $\mathbf{u}_j = (u_{j1}, u_{j2})^T$, se definirá si se cumplen dos condiciones:

1. Si $d(\mathbf{u}_i, \mathbf{u}_j)$ es de las k menores dejando fijo el vértice \mathbf{u}_i y cambiando \mathbf{u}_j .
2. Si $d(\mathbf{u}_i, \mathbf{u}_j)$ es menor o igual que la *mediana* de las distancias entre todos los puntos característicos.

En las condiciones anteriores $d(\mathbf{u}_i, \mathbf{u}_j)$ es la norma ℓ^2 del vector resultante de la sustracción ($\|\mathbf{u}_i - \mathbf{u}_j\|_2$). Continuando con el algoritmo, si uno de los puntos no tiene k vecinos que cumplan las anteriores condiciones, es eliminado del conjunto de vértices junto con las aristas relacionadas con él.

Ahora bien, para cada gráfica se construye una *matriz de adyacencia* (\mathbf{A}), definida como sigue:

$$\mathbf{A}(i, j) = \begin{cases} 1 & \text{si } \mathbf{u}_i \mathbf{u}_j \text{ es una arista de la gráfica,} \\ 0 & \text{de otra forma,} \end{cases}$$

donde, \mathbf{u}_i y \mathbf{u}_j pertenecen a la misma gráfica.

Una vez construidas las gráficas \mathbf{A} y \mathbf{A}' de las dos imágenes, se eliminan de correspondencias erróneas detectando cuales son los vértices (uno en cada imagen) que más rompen con la estructura de la gráfica ingresando aristas no compartidas entre ambas gráficas. Después de eliminar dichos vértices, se reconstruyen las gráficas nuevamente.

El procedimiento para detectar los vértices (de la correspondencia errónea) a eliminar es la construcción de una matriz de diferencia (\mathbf{D}), definida como sigue:

$$\mathbf{D} = |\mathbf{A} - \mathbf{A}'|, \quad (2.15)$$

donde, las barras verticales indican que se aplica el valor absoluto elemento a elemento de la matriz $\mathbf{A} - \mathbf{A}'$.

Las casillas de \mathbf{D} con 1's representan las aristas no comunes en ambas gráficas, por lo tanto, el vértice cuya fila contiene el mayor número de 1's, es él que debe ser eliminado. Posteriormente, se reconstruyen las gráficas y se repite el proceso hasta que las gráficas sean idénticas, es decir, que la matriz \mathbf{D} sólo contenga ceros. La complejidad del algoritmo es $O(n^3 \log(n))$, de ahí la necesidad de optimizarlo.

2.3.2. Algoritmo optimizado

La complejidad cúbica del algoritmo recae en la regeneración de las gráficas en cada iteración. La optimización se presenta en la etapa de búsqueda del vértice con mayor número de aristas no comunes y en la reconstrucción de los gráficas. A continuación presentamos una descripción muy simple de las modificaciones hechas al algoritmo de fuerza bruta, para mayor detalle consultar [Aguilar, 2006].

Sustituiremos las matrices de adyacencia por tres estructuras de datos nuevas:

- Una matriz \mathbf{S} de $N \times N$ en la cual cada renglón corresponde a un vértice y cada casilla contiene el índice del vértice con el cual se define una arista de salida (ordenadas de menor a mayor distancia).
- Un arreglo \mathbf{e} de listas ligadas de longitud N . Cada lista contiene los índices de los vértices con los cuales se definen aristas de entrada con el vértice correspondiente.
- Un arreglo \mathbf{n} (vector) de longitud N que contiene la posición (columna) en \mathbf{S} que almacena al siguiente vértice a ser conectado. Esta estructura de datos es análoga a un arreglo de apuntadores.

Para encontrar el vértice a eliminar, uniremos las k aristas de salida (almacenadas en \mathbf{S}) y las aristas de entrada (almacenadas en \mathbf{e}) de cada vértice, procedemos a comparar este conjunto de aristas con el respectivo de la otra imagen. El vértice que resulte con el mayor número de aristas distintas será eliminado.

La eliminación del vértice y la actualización de las gráficas se realiza con los siguientes pasos:

- Se suprime el vértice de todas las listas contenidas en \mathbf{e} en las cuales esté presente.
- Se elimina la fila respectiva del vértice de \mathbf{S} , impidiendo el acceso a los índices correspondientes a esas localidades de memoria.
- Eliminamos las aristas que involucran al vértice en \mathbf{S} , mediante la actualización de los índices en \mathbf{n} .

2.3.3. Ejemplo de funcionamiento

En esta sección presentamos un ejemplo (presentado en [Aguilar, 2006]) del funcionamiento del algoritmo de correspondencias por transformación de gráficas en su versión de fuerza bruta, pues por simplicidad y brevedad sugerimos referirse a [Aguilar, 2006] para ver a detalle un ejemplo del algoritmo optimizado.

En la Figura 2.8 se muestra un par de imágenes con las correspondencias iniciales señaladas, donde los vértices \mathbf{u}_i y \mathbf{u}'_i están emparejados. Para este ejemplo haremos $k = 2$. Tras calcular todas las distancias euclidianas entre los puntos, ordenarlas y encontrar a los 2 vecinos más cercanos de cada punto tenemos como resultado las gráficas presentadas en la Figura 2.9.



Figura 2.8: Conjunto de correspondencias iniciales para ser refinado con GTM [Aguilar, 2006].



Figura 2.9: Gráficas iniciales generadas con una $k = 2$. La correspondencia errónea esta marcada con un círculo azul [Aguilar, 2006].

Teniendo construidas ambas gráficas, procedemos a elaborar sus respectivas matrices de adyacencia, presentadas en las ecuaciones 2.16 y 2.17.

$$\mathbf{A} = \begin{matrix} & \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 & \mathbf{u}_5 & \mathbf{u}_6 \\ \mathbf{u}_1 & \left(\begin{array}{cccccc} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{array} \right) & & & & & \\ \mathbf{u}_2 & & & & & & & & & & & \\ \mathbf{u}_3 & & & & & & & & & & & \\ \mathbf{u}_4 & & & & & & & & & & & \\ \mathbf{u}_5 & & & & & & & & & & & \\ \mathbf{u}_6 & & & & & & & & & & & \end{matrix} \quad (2.16)$$

$$\mathbf{A}' = \begin{matrix} & \mathbf{u}'_1 & \mathbf{u}'_2 & \mathbf{u}'_3 & \mathbf{u}'_4 & \mathbf{u}'_5 & \mathbf{u}'_6 \\ \mathbf{u}'_1 & \left(\begin{array}{cccccc} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{array} \right) & & & & & \\ \mathbf{u}'_2 & & & & & & & & & & & \\ \mathbf{u}'_3 & & & & & & & & & & & \\ \mathbf{u}'_4 & & & & & & & & & & & \\ \mathbf{u}'_5 & & & & & & & & & & & \\ \mathbf{u}'_6 & & & & & & & & & & & \end{matrix} \quad (2.17)$$

Después de restar ambas matrices y aplicar el valor absoluto, obtenemos la siguiente

matriz de diferencia \mathbf{D} presentada en la ecuación 2.18.

$$\mathbf{D} = |\mathbf{A} - \mathbf{A}'| = \begin{matrix} & \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 & \mathbf{u}_5 & \mathbf{u}_6 \\ \mathbf{u}_1 & \left(\begin{array}{cccccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \end{array} \right) & & & & & \\ \mathbf{u}_2 & & & & & & \\ \mathbf{u}_3 & & & & & & \\ \mathbf{u}_4 & & & & & & \\ \mathbf{u}_5 & & & & & & \\ \mathbf{u}_6 & & & & & & \end{matrix} \quad (2.18)$$

En este ejemplo la correspondencia que desvía más la estructura de la gráfica es la (u_6, u'_6) , por lo tanto, se le considera errónea y se elimina. Se reconstruyen las gráficas, resultando idénticos (véase la Figura 2.10). Finalmente, las matrices de adyacencia son idénticas, condición con la cual finaliza el algoritmo.



Figura 2.10: Gráficas finales sin la presencia de correspondencias erróneas, resultado final de GTM [Aguilar, 2006].

Ahora que sabemos que las imágenes usadas para reconstruir están libres de distorsión y que las correspondencias entre las diferentes vistas de la escena no tienen errores, nos encontramos en condiciones de ejecutar el proceso de reconstrucción. Sin embargo, como se mencionó con anterioridad, la parte medular de la geometría epipolar es la calibración de la cámara, es por eso que abundaremos en este proceso en el siguiente capítulo.

Capítulo 3

Calibración de cámaras

Los métodos de *calibración* tiene como objetivo estimar el valor de los parámetros internos (intrínsecos) y externos (extrínsecos) de la cámara que se pretende modelar, haciendo uso de patrones geoméricamente conocidos capturados en imágenes. Nos referiremos a estos métodos de calibración como métodos clásicos. Introduciremos los conceptos básicos del modelado de una cámara proyectiva. Se presentarán tres métodos, pues consideramos que son los más representativos de este tema de visión computacional, y se describirán en orden cronológico de aparición. Dichos métodos son: el algoritmo de Tsai [Tsai, 1987], de Kanatani [Kanatani, 1993] y por último el de Zhang [Zhang, 1999]. Finalizaremos presentando el sistema desarrollado por *calib-tool* Bouguet [2008] implementado en MATLAB, el cual empleamos para calibrar las cámaras usadas en este trabajo.

3.1. Conceptos básicos

Los parámetros extrínsecos de la cámara indican su posición y orientación con respecto al marco de referencia del mundo. Estos parámetros son la matriz de rotación \mathbf{R} y el vector de traslación \mathbf{t} que constituyen la transformación de cuerpo rígido que superpone el sistema de referencia de la cámara con el sistema de referencia del mundo, aplicada como se muestra a continuación:

$$\mathbf{x}_c = \mathbf{R}\mathbf{x}_w + \mathbf{t}, \quad (3.1)$$

donde, $\mathbf{x}_c = (x_{c1}, x_{c2}, x_{c3})^T$ es un punto bajo el sistema de referencia de la cámara y $\mathbf{x}_w = (x_{w1}, x_{w2}, x_{w3})^T$ es un punto con referencia al sistema del mundo.

Los parámetros intrínsecos no dependen de la posición de la cámara, sino únicamente de su construcción y de la posición de los lentes que conforman su sistema óptico. Estos

parámetros se introducen en una matriz de 3×3 que representa una proyección en perspectiva, tal como se enseñó en la ecuación (2.7). Esta transformación calcula la posición de las proyecciones en el plano de la imagen de los puntos tridimensionales. Además convierte las unidades de longitud de metros (si es el caso) a píxeles, para modelar la captura de las imágenes con el sensor CCD, que por la naturaleza de la digitalización convierte las unidades de los puntos. Dicha transformación se le conoce como modelo de *cámara ideal* y se presenta a continuación:

$$u_1 = \frac{f}{s_x} \frac{x_{c1}}{x_{c3}} + u_{o1}, \quad u_2 = \frac{f}{s_y} \frac{x_{c2}}{x_{c3}} + u_{o2}, \quad (3.2)$$

donde,

- $(u_1, u_2)^T$ es un punto en la imagen,
- f es la distancia del centro de la cámara al plano de la imagen y es igual a la distancia focal del sistema óptico si el objeto se encuentra en el infinito,
- $(u_{o1}, u_{o2})^T$ es el punto de intersección del eje óptico de la cámara con el plano de la imagen y se conoce como *punto principal* y
- s_x e s_y son las distancias entre los centros de las celdas CCD del sensor en metros o bien, el tamaño de los píxeles en cada una de sus dimensiones.

El modelo de la cámara *pin-hole* o *ideal* con los parámetros internos y externos es un conjunto de transformaciones lineales en cascada que cambian el marco de referencia de los puntos en la escena, primero del sistema del mundo al sistema de la cámara, y de éste último al sistema de la imagen. Si se añaden dos factores (ϵ, γ , descritos más adelante) que representan a fenomenos relacionados con la construcción de la cámara real, que nos desvian del caso ideal, la expresión matricial del modelo se puede escribir como se muestra a continuación:

$$\lambda \begin{bmatrix} u_1 \\ u_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \epsilon f/s_x & \gamma & u_{o1} \\ 0 & f/s_y & u_{o2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} x_{w1} \\ x_{w2} \\ x_{w3} \\ 1 \end{bmatrix}, \quad (3.3)$$

dicha ecuación se puede reescribir de la siguiente forma:

$$\lambda \begin{bmatrix} u_1 \\ u_2 \\ 1 \end{bmatrix} = \mathbf{K} \mathbf{D}_u \mathbf{G} \begin{bmatrix} x_{w1} \\ x_{w2} \\ x_{w3} \\ 1 \end{bmatrix}, \quad (3.4)$$

donde,

- \mathbf{K} se conoce como matriz de calibración o de parámetros,
- \mathbf{D}_u es una matriz que permite realizar la multiplicación de matrices y en inglés se le conoce como matriz *dummy*,
- \mathbf{G} es la matriz correspondiente a la transformación euclídeana de los parámetros extrínsecos,
- ϵ es un factor de escala conocido como razón de aspecto (*aspect ratio*) y se refiere a la proporción entre el ancho y el alto de los píxeles, y
- γ representa el efecto de asimetría (*skew*) en la cuadrícula de píxeles en el sensor de la cámara ocasionado por los primeros captador de imágenes (*frame grabber*), en los cuales el escaneo horizontal era línea por línea del sensor. En los modernos capturadores la lectura se hace en paralelo, por lo tanto, la asimetría se considera nula.
- λ es el factor de escala al cual son invariantes los puntos en coordenadas homogéneas. Este factor de escala queda indeterminado porque a una misma restricción geométrica corresponden infinitas parejas de imágenes.

A la matriz resultante de la concatenación de las matrices de parámetros intrínsecos y extrínsecos se le conoce como matriz de *proyección en perspectiva* ($\mathbf{\Pi}$) y relaciona a un punto en el espacio proyectivo (\mathbb{P}^3) con otro punto en el plano proyectivo (\mathbb{P}^2). La representación a bloques de $\mathbf{\Pi}$ se presenta a continuación:

$$\mathbf{\Pi} = \mathbf{K}\mathbf{D}_u\mathbf{G}. \quad (3.5)$$

La ecuación (3.2) se obtiene a partir de la ecuación (3.3), si suponemos que los píxeles son cuadrados ($\epsilon = 1$), la asimetría es nula en la distribución de las celdas del sensor ($\gamma = 0$) y $\lambda = 1$ (sin pérdida de generalidad) puesto que estamos trabajando con coordenadas homogéneas.

Una vez presentados los conceptos básicos del modelado de cámaras, continuamos con la presentación de 3 métodos de calibración de cámaras, los cuales están ordenados cronológicamente según sus fechas de publicación. Los algoritmos que se describirán son la base de los métodos más recientes en este tema de visión computacional.

3.2. Algoritmo de Tsai (1987)

Este método de calibración necesita al menos 7 puntos no coplanares de referencia en la escena 3D, para lo cual se toman fotografías de un patrón de referencia como el mostrado en la Figura 3.1. El funcionamiento de este algoritmo se basa en el uso de pares de coordenadas del mismo punto, el primero (\mathbf{x}_w) bajo el sistema de referencia del mundo y el segundo (\mathbf{u}) bajo el sistema de referencia de la imagen, pues sabemos que ambos vectores están relacionados por la matriz $\mathbf{\Pi}$. El sistema coordenado del mundo se posiciona de acuerdo al patrón de referencia para facilitar la medición de la posición de los puntos sobresalientes. La relación dada por la proyección se muestra a continuación:

$$\mathbf{u} = \mathbf{\Pi}\mathbf{x}_w. \quad (3.6)$$

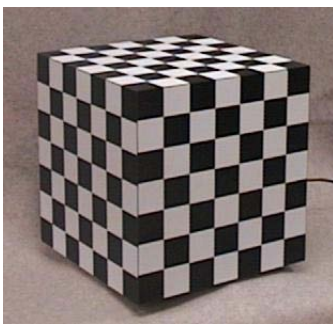


Figura 3.1: Ejemplo de patrón de calibración usado para el método de Tsai. Tomado de www.cs.rochester.edu/~sanders/cali_box/.

3.2.1. Conversión de unidades

Aplicamos las operaciones inversas a la ecuación (3.2), puesto que se desea realizar una reproyección de los puntos que se encuentran en las imágenes. La conversión se realiza con las siguientes ecuaciones (conservando las definiciones hechas en la sección 3.1 en el modelo de la cámara *pin-hole*):

$$x_{c1} = \frac{s_x(u_1 - u_{o1})}{\epsilon}, \quad x_{c2} = s_y(u_2 - u_{o2}), \quad (3.7)$$

donde, $(x_{c1}, x_{c2})^T$ es un punto en la imagen bajo el sistema de referencia de la cámara en metros, y $(u_1, u_2)^T$ es el mismo punto pero en píxeles y bajo el sistema de referencia de la

imagen. En este método se considera que las imágenes no han sido corregidas de la distorsión radial, pero se modela esta aberración en etapas posteriores del algoritmo. Para el cálculo de estas coordenadas supondremos que $\gamma = 0$, $f = 1$ y $\epsilon = 1$. Si las imágenes estuvieran libres de distorsión, estas coordenadas al ser reproyectadas representarían la posición original de los puntos en el espacio salvo un factor de escala desconocido.

3.2.2. Relación en coordenadas de la cámara

Procedemos a relacionar estos puntos en las imágenes con los puntos en el sistema de referencia del mundo, lo anterior en términos de la transformación de cuerpo rígido aplicada a las coordenadas en el sistema del mundo, véase las siguientes ecuaciones y observese que el parámetro ϵ es nuevamente incluido para su estimación:

$$\frac{x_{c1}}{\epsilon} = \frac{r_{11}x_{w1} + r_{12}x_{w2} + r_{13}x_{w3} + t_1}{r_{31}x_{w1} + r_{32}x_{w2} + r_{33}x_{w3} + t_3}, \quad (3.8)$$

$$x_{c2} = \frac{r_{21}x_{w1} + r_{22}x_{w2} + r_{23}x_{w3} + t_2}{r_{31}x_{w1} + r_{32}x_{w2} + r_{33}x_{w3} + t_3}. \quad (3.9)$$

Obtenemos una sola ecuación despejando el denominador de la ecuación (3.9) y sustituyéndolo en la ecuación (3.8). Ahora, si ordenamos los términos de la ecuación resultante como sigue:

$$x_{c1} = \frac{1}{t_2} (\epsilon x_{c2} (r_{11}x_{w1} + r_{12}x_{w2} + r_{13}x_{w3} + t_1) - x_{c1} (r_{21}x_{w1} + r_{22}x_{w2} + r_{23}x_{w3})), \quad (3.10)$$

podemos definir a:

$$\mathbf{E} = \begin{bmatrix} x_{c2}x_{w1} & x_{c2}x_{w2} & x_{c2}x_{w3} & x_{c2} & -x_{c1}x_{w1} & -x_{c1}x_{w2} & -x_{c1}x_{w3} \end{bmatrix} \text{ y} \quad (3.11)$$

$$\mathbf{l} = \frac{1}{t_2} [\epsilon r_{11}, \epsilon r_{12}, \epsilon r_{13}, \epsilon t_1, r_{21}, r_{22}, r_{23}]^T, \quad (3.12)$$

donde, \mathbf{E} es una matriz de 1×7 que contiene las coordenadas de los puntos en la imagen y en el mundo, y \mathbf{l} es un vector de \mathbb{R}^7 que contienen parámetros desconocidos de la cámara. Reescribiendo la ecuación (3.10) en términos de las ecuaciones (3.11) y (3.12), obtenemos:

$$x_{c1} = \mathbf{E}\mathbf{l}. \quad (3.13)$$

3.2.3. Obtención de parámetros

Debido a que la ecuación (3.13) tiene 7 incógnitas necesitamos mínimo 7 puntos de referencia tomados del patrón de referencia; sin embargo, para minimizar los errores por ruido es necesario usar un mayor número de puntos para formar un sistema sobre-determinado de ecuaciones, el cual se presenta abajo:

$$\mathbf{r} = \mathbf{E}'\mathbf{l}, \quad (3.14)$$

donde, $\mathbf{r} = (x_{c1,1}, x_{c1,2}, \dots, x_{c1,N})^T$ es un vector perteneciente a \mathbb{R}^N y $\mathbf{E}' = [\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_N]^T$ es una matriz de $N \times 7$, donde N es el número de pares de puntos conocidos. Este sistema de ecuaciones se puede resolver por el método propuesto en [Moore, 1920], cuya solución sigue el criterio de mínimos cuadrados y se presenta abajo:

$$\mathbf{l} = \left(\mathbf{E}'^T \mathbf{E}' \right)^{-1} \mathbf{E}'^T \mathbf{r}. \quad (3.15)$$

Una vez obtenido el vector \mathbf{l} se procede a calcular los parámetros t_1 , t_2 y ϵ utilizando propiedades de ortonormalidad de la matriz de rotación. La tercera fila de la matriz \mathbf{R} se obtiene con el producto cruz de sus dos primeras filas, las cuales son extraídas de \mathbf{l} con anterioridad. Finalmente, f y t_3 se calculan construyendo un sistema sobre-determinado de ecuaciones como la presentada en la ecuación (3.9) multiplicada en su lado derecho por f . El sistema nuevamente puede ser resuelto con el método de la inversa de *Moore-Penrose* [Moore, 1920].

Ahora contamos con todos los parámetros intrínsecos y extrínsecos de la cámara usando como datos de entrada una sola imagen del patrón de referencia y los puntos sobresalientes localizados en la imagen. Para más detalles del algoritmo consultar [Tsai, 1987].

3.3. Algoritmo de Kanatani (1993)

En este algoritmo, un punto en el espacio o en la imagen se representa con un vector \mathbf{v} asociado a él. Dicho vector es unitario, se posiciona en el origen del sistema coordenado del mundo y apunta en dirección al punto que representa. Una línea es representada mediante un vector \mathbf{w} , el cual es unitario, se posiciona en el origen del sistema coordenado y esta en dirección normal al plano que contiene al origen y a la línea representada. A estas representaciones de puntos y líneas en \mathbb{P}^3 se les conoce como N-vectores.

Este método está basado en conceptos bastante simples de geometría proyectiva, y fue presentado [Kanatani, 1993] como una aplicación en visión computacional. Las propiedades geométricas entre puntos y líneas usadas principalmente son: la ortogonalidad entre

N-vectores de puntos de fuga, la *conjugación* y la *polaridad*.

Un punto de fuga, correspondiente a una línea en el espacio, se define como el límite de la proyección de un punto que recorre dicha línea en la escena en una sola dirección. Por otro lado, se dice que dos puntos son conjugados, si al transformar en una línea al primer punto, aplicando una correlación, ésta pasa por el segundo punto y viceversa. Una correlación se define como una transformación que mapea el conjunto de todos los puntos imagen al conjunto de todas las líneas imagen y viceversa. También se dice que una correlación es una polaridad cuando transforma siempre un mismo punto (*polo*) en una misma línea (*polar*) y viceversa. Teniendo en cuenta lo anterior, procedamos a describir el algoritmo de calibración.

3.3.1. Determinación del punto principal

El punto principal (no confundir con uno de los puntos cardinales en los sistemas ópticos) se define como la intersección del eje óptico de la cámara con el plano imagen. Su determinación se basa en las propiedades que tienen los N-vectores de los puntos de fuga de líneas que en la escena son ortogonales y cómo éstos pueden formar un triángulo *auto-polar* en la imagen, el cual se define como el triángulo cuyas polares de sus vértices pasan por los otros dos vértices, es decir, son mutuamente conjugados. El siguiente teorema y sus derivaciones fundamentan el algoritmo para el cálculo del punto principal.

Teorema 3.1 *El ortocentro de un triángulo auto-polar se encuentra en el punto principal de la imagen [Kanatani, 1993].*

Del anterior teorema podemos deducir lo siguiente (ver Figura 3.2(b)):

- El punto principal (\mathbf{v}_o) de la imagen es el ortocentro del triángulo definido por tres puntos de fuga de líneas mutuamente ortogonales en la escena.
- El punto principal (\mathbf{v}_o) de la imagen es el ortocentro del triángulo definido por tres líneas de fuga de superficies planas mutuamente ortogonales en la escena.

El algoritmo de estimación del punto principal inicia con una imagen de un cubo, en la cual sean visibles tres de sus caras, como se puede ver en la Figura 3.2(a), después ajustamos rectas a cada una de sus aristas y calculamos con sus intersecciones los tres puntos de fuga. Al ser mutuamente ortogonales los tres conjuntos de rectas paralelas en la escena, los N-vectores de sus puntos de fuga también lo serán formando así un triángulo *auto-polar* en el plano imagen.

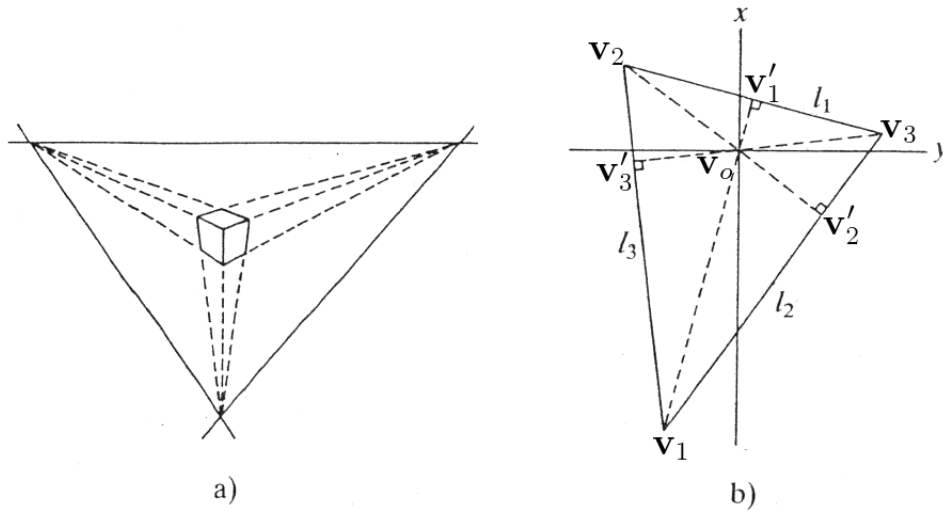


Figura 3.2: (a) Vista del patrón de referencia para estimación del punto principal de la cámara y (b) triángulo auto-polar formado por tres puntos de fuga de líneas ortogonales en la escena [Kanatani, 1993].

Cada vértice del triángulo con su respectiva arista opuesta forman una pareja polo-polar respecto a la polaridad estándar (matriz identidad). Por lo tanto, cada vértice del triángulo es conjugado de los otros dos, confirmando así el Teorema 3.1. El triángulo auto-polar en la imagen se presenta en la Figura 3.2(b), donde \mathbf{v}_i son los puntos de fuga de las líneas en la escena, \mathbf{v}'_i son las intersecciones de las alturas del triángulo con las aristas y \mathbf{w}_i son los N-vectores correspondientes a las líneas l_i que pasan por las aristas, para $i = 1, 2, 3$.

Al contar con los N-vectores de los puntos de fuga \mathbf{v}_i , sólo nos resta encontrar los N-vectores de los puntos \mathbf{v}'_i sobre las líneas polares. El punto sobre la línea polar debe formar con el respectivo polo una recta que sea perpendicular a dicha polar, es decir, la distancia $d(\mathbf{v}_i, \mathbf{v}'_i)$ representa la distancia del polo a su línea polar. Con estos dos N-vectores podemos calcular el N-vector de la recta que pasa por el ortocentro con la ecuación (3.18) para cada uno de los tres pares. El punto de concurrencia de esas tres rectas será el ortocentro de tal triángulo y puede estimarse bajo el siguiente teorema:

Teorema 3.2 *Las líneas imagen l_i , $i = 1, \dots, N$, son concurrentes si y sólo si el menor de los eigenvalores de su matriz de momento*

$$M = \sum_{i=1}^N W_i \mathbf{w}_i \mathbf{w}_i^T, \quad (3.16)$$

es cero, donde $\{W_i\}$ es un conjunto de constantes positivas. El eigenvector unitario asociado \mathbf{v} es el N-vector del punto común de intersección [Kanatani, 1993].

Siendo \mathbf{v}_i el N-vector de uno de los polos en la imagen, y siendo \mathbf{v}'_i el punto sobre la polar l_i que forma con \mathbf{v}_i una línea perpendicular a dicha polar l_i que tiene el N-vector \mathbf{w}_i asociado. \mathbf{v}'_i se calcula con la ecuación que sigue:

$$\mathbf{v}'_i = \pm N \left[\mathbf{k} - (\mathbf{w}_i, \mathbf{k})\mathbf{w}_i - \frac{|\mathbf{w}_i, \mathbf{v}_i, \mathbf{k}|}{(\mathbf{v}_i, \mathbf{k})}\mathbf{w}_i \times \mathbf{k} \right], \quad (3.17)$$

donde, $\pm N[\cdot]$ es un operador que normaliza la magnitud de los vectores a 1 y \mathbf{k} es el vector unitario en dirección del eje Z.

3.3.2. Determinación de la distancia focal

Como primer paso, tenemos que capturar una imagen de un rectángulo localizado en un superficie plana en la escena, como se muestra en la Figura 3.3.

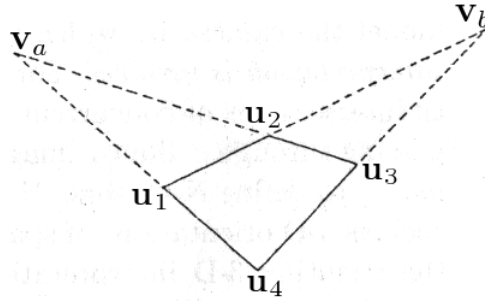


Figura 3.3: Diagrama de un rectángulo posicionado en la escena usado como patrón para estimar la distancia focal [Kanatani, 1993].

Siendo \mathbf{u}_i para $i = 1, \dots, 4$, los vértices del rectángulo en la imagen ordenados en el sentido de las manecillas del reloj. Calculamos sus respectivos N-vectores \mathbf{v}_i usando un valor tentativo de distancia focal f . Ahora calculamos los N-vectores de las líneas que pasan por los lados del rectángulo $\mathbf{u}_1\mathbf{u}_2$, $\mathbf{u}_2\mathbf{u}_3$, $\mathbf{u}_3\mathbf{u}_4$ y $\mathbf{u}_4\mathbf{u}_1$, por medio de la siguiente ecuación:

$$\mathbf{w} = \pm N[\mathbf{v} \times \mathbf{v}']. \quad (3.18)$$

Teniendo los N-vectores de las aristas procedemos a calcular los N-vectores de los puntos de fuga de los dos pares de líneas paralelas, es decir, la intersección de $\mathbf{u}_1\mathbf{u}_2$ y $\mathbf{u}_3\mathbf{u}_4$, y la

intersección de $\mathbf{u}_2\mathbf{u}_3$ y $\mathbf{u}_4\mathbf{u}_1$, como sigue:

$$\mathbf{v} = \pm N[\mathbf{w} \times \mathbf{w}']. \quad (3.19)$$

Finalmente, calculamos la distancia focal verdadera \hat{f} usando dos propiedades de los N-vectores. La primera consiste en la modificación de los N-vectores cuando cambia la distancia focal, propiedad mostrada a continuación:

$$\mathbf{v}' = \pm N \left[\begin{pmatrix} v_1 \\ v_2 \\ \left(\frac{f'}{f}\right) v_3 \end{pmatrix} \right], \quad \mathbf{w}' = \pm N \left[\begin{pmatrix} w_1 \\ w_2 \\ \left(\frac{f'}{f}\right) w_3 \end{pmatrix} \right], \quad (3.20)$$

donde, los N-vectores $\mathbf{v} = (v_1, v_2, v_3)^T$ y $\mathbf{w} = (w_1, w_2, w_3)^T$ son calculados inicialmente con f , y los N-vectores $\mathbf{v}' = (v'_1, v'_2, v'_3)^T$ y $\mathbf{w}' = (w'_1, w'_2, w'_3)^T$ están calculados con f' que es la nueva distancia focal. La segunda propiedad consiste en que los puntos de fuga (\mathbf{v}_a y \mathbf{v}_b) de líneas ortogonales en la escena, son conjugados entre ellos bajo la cónica absoluta (matriz identidad). Usando las dos propiedades descritas, obtenemos la siguiente ecuación para calcular la distancia focal verdadera:

$$\hat{f} = f \sqrt{-\frac{v_{a1}v_{b1} + v_{a2}v_{b2}}{v_{a3}v_{b3}}}. \quad (3.21)$$

3.3.3. Modelo de ruido

Como los puntos de fuga son calculados de una imagen real, una cierta cantidad de ruido debe ser considerada. El ruido en los puntos de fuga está descrito por sus respectivas matrices de covarianza $\mathbf{V}[\mathbf{v}_a]$ y $\mathbf{V}[\mathbf{v}_b]$ presentadas a continuación:

$$\mathbf{V}[\mathbf{v}_a] = \frac{\vartheta/f^2}{2(1+r^2/f^2)} \left(\mathbf{p}\mathbf{p}^T + \frac{\mathbf{q}\mathbf{q}^T}{1+r^2/f^2} \right), \quad (3.22)$$

donde,

$$\mathbf{k} = (0, 0, 1)^T, \quad \mathbf{p} = \pm \sqrt{(1+f^2/r^2)} \mathbf{v}_a \times \mathbf{k}, \quad \mathbf{q} = \pm \mathbf{p} \times \mathbf{v}_a,$$

f es la distancia del centro de la cámara al plano imagen, r es la distancia del punto imagen al punto principal, y ϑ se le conoce como la precisión de la imagen y es el valor *RMS* del ruido presente en la imagen, el cual se considera presente en todos los píxeles de la imagen y con la misma probabilidad en todas las orientaciones.

La varianza en la estimación de la distancia focal esta definida como sigue:

$$\mathbf{V}[f] = \frac{f^2 (\mathbf{v}_b, \mathbf{V}[\mathbf{v}_a] \mathbf{v}_b) + (\mathbf{v}_a, \mathbf{V}[\mathbf{v}_b] \mathbf{v}_a)}{4 (v_{a3} v_{b3})}. \quad (3.23)$$

Debido a que la confiabilidad en la estimación de la distancia focal disminuye al aumentar la distancia entre los puntos de fuga y el centro de la imagen, se opta por hacer un promedio ponderado de distancias focales estimadas en imágenes con diferentes puntos de vista del plano de referencia. A este promedio se le conoce como longitud focal óptima estimada \bar{f} , y se define en la siguiente ecuación:

$$\bar{f} = \sum_{i=1}^N W_i f_i, \quad \sum_{i=1}^N W_i = 1, \quad (3.24)$$

donde, f_i para $i = 1, \dots, N$ son las distancias focales estimadas de diferentes imágenes. Los pesos son determinados en función de la varianza asociada a la distancia focal estimada $\mathbf{V}[f_i]$ con la siguiente ecuación:

$$W_i = \frac{1}{\mathbf{V}[f_i] \sum_{j=1}^N \frac{1}{\mathbf{V}[f_j]}}, \quad (3.25)$$

y la varianza de la longitud focal óptima estimada está definida como sigue:

$$\mathbf{V}[\bar{f}] = \frac{1}{\sum_{i=1}^N \frac{1}{\mathbf{V}[f_i]}}. \quad (3.26)$$

Para mayores detalles sobre este modelo estadístico de ruido referirse a [Kanatani, 1993].

3.4. Algoritmo de Zhang (1999)

Este método de calibración de cámaras presenta la ventaja de utilizar como datos de entrada imágenes de un patrón de referencia plano, como un tablero de ajedrez, el cual es más sencillo de construir que el empleado en el método presentado en [Tsai, 1987] (véase sección 3.2). El método está basado en el cálculo de homografías, de ahí el uso de un patrón de calibración plano, que transforman los puntos característicos en el plano del patrón a su posición en el plano de la imagen. A continuación describimos los pasos que conforman el algoritmo.

3.4.1. Estimación de homografías

Para el cálculo de las homografías asumiremos, al igual que en los otros métodos, que el sistema de referencia del mundo se posiciona en el patrón de referencia, sin embargo, en este caso al ser plano se le considera localizado en $Z_w = 0$, sin pérdida de generalidad. Lo anterior se puede observar en la Figura 3.4.

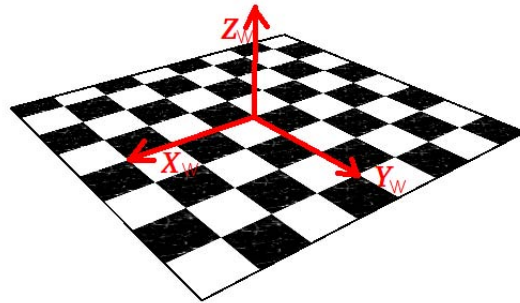


Figura 3.4: Posicionamiento del marco de referencia del mundo sobre el patrón de ajedrez.

Los puntos característicos extraídos del patrón tienen coordenadas espaciales de la forma $\mathbf{x} = (x_1, x_2, 0)^T$ y después de proyectarse en el plano de la imagen resultan de la forma $\mathbf{u} = (u_1, u_2)^T$. Bajo el modelo de *cámara ideal*, presentado en la ecuación (3.3), esta proyección se puede escribir como la ecuación (3.27).

$$\begin{aligned} \lambda \begin{bmatrix} u_1 \\ u_2 \\ 1 \end{bmatrix} &= \mathbf{K} \begin{bmatrix} \mathbf{r}_{c1} & \mathbf{r}_{c2} & \mathbf{r}_{c3} & \mathbf{t} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 0 \\ 1 \end{bmatrix} \\ &= \mathbf{K} \begin{bmatrix} \mathbf{r}_{c1} & \mathbf{r}_{c2} & \mathbf{t} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}, \end{aligned} \quad (3.27)$$

donde, $\mathbf{r}_{ci} = (r_{1,i}, r_{2,i}, r_{3,i})^T$ para $i = 1, 2, 3$, son los vectores columna de la matriz de rotación de los parámetros extrínsecos y $r_{m,n} = \mathbf{R}(m, n)$. Podemos observar que la matriz $\mathbf{H} = \mathbf{K} \begin{bmatrix} \mathbf{r}_{c1} & \mathbf{r}_{c2} & \mathbf{t} \end{bmatrix}$ es de 3×3 , representa una homografía, ya que proyecta un plano (patrón de referencia) sobre otro (plano de la imagen), es independiente a un factor de escala y ahora relaciona a dos puntos ambos en \mathbb{P}^2 .

La relación entre un punto en el patrón y un punto en el plano de la imagen establecida por la homografía antes definida se puede escribir como una relación de colinealidad, presentada en la ecuación (3.28), debido a que los vectores solo difieran por un factor de escala.

$$\mathbf{u}_j \times (\mathbf{H}\mathbf{x}_j) = \mathbf{0}, \quad (3.28)$$

donde, j es el índice del punto característico tomado del patrón de referencia y su correspondiente posición en la imagen.

Aplicando la homografía y desarrollando el producto cruz entre los vectores de la ecuación (3.28), construimos un sistema homogéneo de ecuaciones en donde el vector de incógnitas está conformado por los vectores fila de \mathbf{H} , los cuales se definen como,

$$\mathbf{h}_{ri} = (h_{i,1}, h_{i,2}, h_{i,3})^T, \quad \text{para } i = 1, 2, 3, \quad (3.29)$$

donde, $h_{m,n} = \mathbf{H}(m, n)$ para $m, n = 1, 2, 3$, son los elementos de la homografía. El sistema de ecuaciones se puede escribir como sigue,

$$\begin{bmatrix} \mathbf{0}^T & -\mathbf{x}_j^T & u_{2j}\mathbf{x}_j^T \\ \mathbf{x}_j^T & \mathbf{0}^T & -u_{1j}\mathbf{x}_j^T \\ -u_{2j}\mathbf{x}_j^T & u_{1j}\mathbf{x}_j^T & \mathbf{0}^T \end{bmatrix} \begin{bmatrix} \mathbf{h}_{r1} \\ \mathbf{h}_{r2} \\ \mathbf{h}_{r3} \end{bmatrix} = \mathbf{0}, \quad (3.30)$$

donde, j es el índice del punto analizado. Sin embargo, podemos observar que el tercer renglón de la matriz izquierda es una combinación lineal de las dos primeras, por lo tanto, la eliminamos para quedarnos con una matriz de 2×9 ,

$$\begin{bmatrix} \mathbf{0}^T & -\mathbf{x}_j^T & u_{2j}\mathbf{x}_j^T \\ \mathbf{x}_j^T & \mathbf{0}^T & -u_{1j}\mathbf{x}_j^T \end{bmatrix} \begin{bmatrix} \mathbf{h}_{r1} \\ \mathbf{h}_{r2} \\ \mathbf{h}_{r3} \end{bmatrix} = \mathbf{0}. \quad (3.31)$$

Por cada punto en el patrón, y con su correspondiente imagen, tenemos dos ecuaciones y como una de las 9 entradas de \mathbf{H} es fija por el factor de escala, mínimo necesitamos 4 puntos característicos para resolver el sistema. Con un mayor número de puntos se mejora la aproximación al tener un sistema sobre-determinado cuya solución la podemos encontrar con descomposición por valores singulares (SVD por las siglas del inglés, *Singular Value Decomposition*) [Press *et al.*, 1995]. Sin embargo, existe un límite en el número de puntos en correspondencia considerados para mejorar la solución del sistema de ecuaciones y se debe a la cantidad de ruido presente en las mediciones.

3.4.2. Estimación de los parámetros de la cámara

Ya que conocemos los valores de las entradas de \mathbf{H} , procedemos a extraer los parámetros internos y externos de la cámara. Definamos los vectores columna de la matriz \mathbf{H} como sigue:

$$\mathbf{h}_{ci} = (h_{1,i}, h_{2,i}, h_{3,i})^T, \quad \text{para } i = 1, 2, 3, \quad (3.32)$$

por lo tanto,

$$\mathbf{H} = \mathbf{K} \begin{bmatrix} \mathbf{r}_{c1} & \mathbf{r}_{c2} & \mathbf{t} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_{c1} & \mathbf{h}_{c2} & \mathbf{h}_{c3} \end{bmatrix}, \quad (3.33)$$

De la ecuación 3.33 se obtienen las siguientes ecuaciones:

$$\mathbf{h}_{c1} = \mathbf{K}\mathbf{r}_{c1}, \quad \mathbf{h}_{c2} = \mathbf{K}\mathbf{r}_{c2}, \quad \mathbf{h}_{c3} = \mathbf{K}\mathbf{t}. \quad (3.34)$$

Despejando los vectores \mathbf{r}_{c1} y \mathbf{r}_{c2} de las ecuaciones 3.34, ya que la matriz de parámetros intrínsecos es invertible, y sustituyendolos en las siguientes ecuaciones:

$$\mathbf{r}_{c1}^T \mathbf{r}_{c2} = 0, \quad \mathbf{r}_{c1}^T \mathbf{r}_{c1} = \mathbf{r}_{c2}^T \mathbf{r}_{c2}, \quad (3.35)$$

que expresan las propiedades de ortogonalidad de las matrices de rotación, obtenemos:

$$\mathbf{h}_{c1}^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{h}_{c2}^T = 0, \quad \mathbf{h}_{c1}^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{h}_{c1}^T = \mathbf{h}_{c2}^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{h}_{c2}^T. \quad (3.36)$$

Al igual que en otras ocasiones, nos vemos en la tarea de despejar las incógnitas de estas ecuaciones agrupandolas en un vector y resolviendo un sistema de ecuaciones sobre-determinado por SVD. Brevemente describiremos a continuación este procedimiento.

Definimos la matriz $\mathbf{J} = \mathbf{K}^{-T} \mathbf{K}^{-1}$ como la matriz simétrica que contiene a las incógnitas. Ya que la matriz inversa de parámetros intrínsecos se define como en la ecuación (3.37), podemos encontrar las expresiones de las entradas de \mathbf{J} en función de dichos parámetros.

$$\mathbf{K}^{-1} = \begin{bmatrix} \frac{1}{f_x} & -\frac{\gamma}{f_x f_y} & \frac{\gamma u_{o2} - f_y u_{o1}}{f_x f_y} \\ 0 & \frac{1}{f_y} & -\frac{u_{o2}}{f_y} \\ 0 & 0 & 1 \end{bmatrix}, \quad (3.37)$$

donde $f_x = \frac{f}{s_x}$ y $f_y = \frac{f}{s_y}$ que representan la distancia focal expresada en píxeles, tomando en cuenta las dimensiones de ancho y alto del píxel, respectivamente.

Los coeficientes $\mathbf{J}(m, n) = j_{m,n}$ se presentan a continuación:

$$j_{1,1} = \frac{1}{f_x^2}, \quad j_{1,2} = \frac{-\gamma}{f_x^2 f_y}, \quad j_{1,3} = \frac{\gamma u_{o2} - f_x u_{o1}}{f_x^2 f_y},$$

$$j_{2,2} = \frac{\gamma^2}{f_x^2 f_y^2} + \frac{1}{f_y^2} \quad j_{2,3} = \frac{-\gamma(\gamma u_{o2} - f_y u_{o1})}{f_x^2 f_y^2} - \frac{u_{o2}}{f_y^2}. \quad (3.38)$$

Ahora en base a las ecuaciones (3.36), construimos una ecuación que relaciona cualquiera dos vectores columna de \mathbf{H} con la matriz \mathbf{J} , y que agrupe los coeficientes de la homografía ($h_{m,n}$) en un vector y los coeficientes de la matriz de incógnitas ($j_{m,n}$) en otro. Esta relación se muestra a continuación:

$$\mathbf{h}_{ci}^T \mathbf{J} \mathbf{h}_{cj} = \mathbf{g}_{ij}^T \mathbf{j}, \quad (3.39)$$

donde,

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

y

$$\mathbf{j} = [j_{1,1}, j_{1,2}, j_{1,3}, j_{2,2}, j_{2,3}, j_{3,3}]^T.$$

Ahora podemos reescribir las condiciones de ortogonalidad de las ecuaciones (3.36) usando la ecuación (3.39), obteniendo como resultado las siguientes ecuaciones:

$$\mathbf{g}_{12}^T \mathbf{j} = 0, \quad (\mathbf{g}_{11}^T - \mathbf{g}_{22}^T) \mathbf{j} = 0. \quad (3.40)$$

Entonces, tenemos dos ecuaciones por homografía estimada, una por cada imagen del patrón de referencia, y buscamos los 6 parámetros de \mathbf{j} (en realidad 5 pues uno está fijo pues es invariante a la escala). Necesitamos al menos 3 imágenes para estimar los parámetros internos y externos con este algoritmo.

Finalmente, después de resolver el sistema de ecuaciones por SVD, despejamos los parámetros intrínsecos de las ecuaciones (3.38) y despejando las ecuaciones (3.34) obtenemos los extrínsecos, completando así el modelado de la cámara. Para mayores detalles de este algoritmo consultar [Zhang, 1999].

3.5. Calib-tool de MATLAB

Calib-tool [Bouguet, 2008] es una herramienta para calibración de cámaras, desarrollada en MATLAB [Little y Moler, 2009], que cuenta con una interfase gráfica de usuario muy simple y amigable para uso no restringido a personas expertas en este tema de visión computacional. Esta implementación es empleada en este trabajo de tesis para modelar la cámara por parte de métodos clásicos.

Esta herramienta (*toolbox*) es considerado el estado del arte en cuestión de calibración de cámaras, pues cuenta con la implementación de un conjunto de métodos, todos publicados independientemente, que combinados alcanzan un alto grado de robustez y confiabilidad en su estimación de parámetros. Los métodos más importantes son los presentados en este capítulo, si se desea conocer más detalles de este sistema refierase a http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/ref.html.

Calib-tool requiere un mínimo de 2 imágenes (sin un número máximo) de un patrón de referencia plano como él de un tablero de ajedrez (véase la Figura 2.3), sin embargo, mientras mayor sea el número de imágenes menor será el error en las estimaciones. Cuenta con una etapa de inicialización en la cual estima las homografías usando el algoritmo de [Zhang, 1999] descrito en la sección 3.4.

El cálculo de los parámetros intrínsecos a partir de las homografías es ligeramente distinto, ya que utiliza las propiedades de ortogonalidad de los puntos de fuga de líneas perpendiculares en la escena por el algoritmo de Kanatani (véase sección 3.3) para estimarlos. Posteriormente con esos parámetros internos y las homografías se extraen los parámetros externos. La etapa de refinamiento del algoritmo de Zhang basado en la *estimación de máxima verosimilitud* se realiza de la misma forma como se encuentra documentado en [Zhang, 1999], es decir, se considera que cada punto característico en las imágenes del patrón esta corrompido por una distribución de ruido independiente y forma común, con el cálculo de la diferencia en posición de estos puntos con los proyectados en las imágenes obtiene una función a minimizar por el método Levenberg-Marquardt [Moré, 1977].

Al igual que en los tres métodos anteriores, *Calib-tool* no considera la distorsión en su estimación inicial, pero en la etapa de optimización no-lineal utiliza un modelo de parámetros intrínsecos propuesto en [Heikkila y Silven, 1997], el cual considera un modelo polinomial par de distorsión radial de sexto orden, más dos coeficientes extras que corresponden a la distorsión tangencial.

De esta forma el *toolbox* calcula los parámetros intrínsecos y extrínsecos, y adicionalmente nos provee de un modelo de distorsión radial y tangencial que puede ser usado para corregir todas las imágenes capturadas con la misma cámara. Sin embargo, en este trabajo no empleamos al *Calib-tool* para corregir distorsión, pues para eso las imágenes fueron tratadas con *ADICOS* y después fueron procesadas en *Calib-tool*.

Gran parte de los principios teóricos usados en el tema de calibración de cámaras han sido cubiertos por los tres métodos clásicos descritos en este capítulo. Es por eso que el

mótipo de presentar estos algoritmos, es entender completamente el problema del modelado de cámaras y presentar las técnicas con las cuales se ha resuelto este problema. Teniendo así, las herramientas para desarrollar por nuestra cuenta un sistema de calibración de cámaras, aunque en este trabajo no se realizó por estar fuera de los objetivos del mismo.

Ahora que ya sabemos que los métodos clásicos de calibración requieren de patrones conocidos, es importante estudiar los fenómenos ópticos en el ojo que podrían modificar los valores estimados por la calibración previa de la cámara *fundus*. Lo anterior, con el fin de minimizar esos efectos de la óptica del ojo en los parámetros intrínsecos. El estudio de algunos conceptos de óptica geométrica aplicados en oftalmología, se presentan en el siguiente capítulo.

Capítulo 4

Óptica del ojo en el proceso de calibración

En este capítulo describiremos de manera sencilla el tema de mayor interés, en relación a este trabajo de tesis, sobre la óptica del ojo humano, denominado *aberraciones*. El objetivo de entender qué son las aberraciones y cómo se modelan, para saber si afectan o no en los parámetros internos de la cámara *fundus*. En el caso de cambiar la calibración, se encontrará un modo de minimizar las aberraciones al momento de capturar las imágenes del fondo de ojo.

Ya que el ojo, visto como una cámara muy sofisticada, cuenta con un sistema de lentes bastante complejo que es necesario modelar para analizar las aberraciones introducidas por él, optamos por considerarlo como un tipo de sistema óptico más simple. Para el análisis de aberraciones asumiremos que es un sistema óptico centrado y con simetría de revolución, es decir, los diferentes lentes con los que está conformado tienen superficies esféricas y sus centros de curvatura se encuentran sobre la misma línea, la cual se conoce como eje óptico. En la Figura 4.1, podemos observar el esquema de un sistema óptico de este tipo.

4.1. Polinomio de aberración a tercer orden

Para entender qué son las aberraciones, necesitamos conocer cómo se refractan los rayos al atravesar cada una de las interfaces (superficie límite entre dos medios con índices de refracción diferentes) y así poderlos seguir hasta su llegada al plano de la imagen, a lo que se le conoce como *trazo exacto de rayos*. Para lo anterior, se puede recurrir al método L-U [Kingslake, 2010], del cual solo se presentaran sus ecuaciones más importantes.

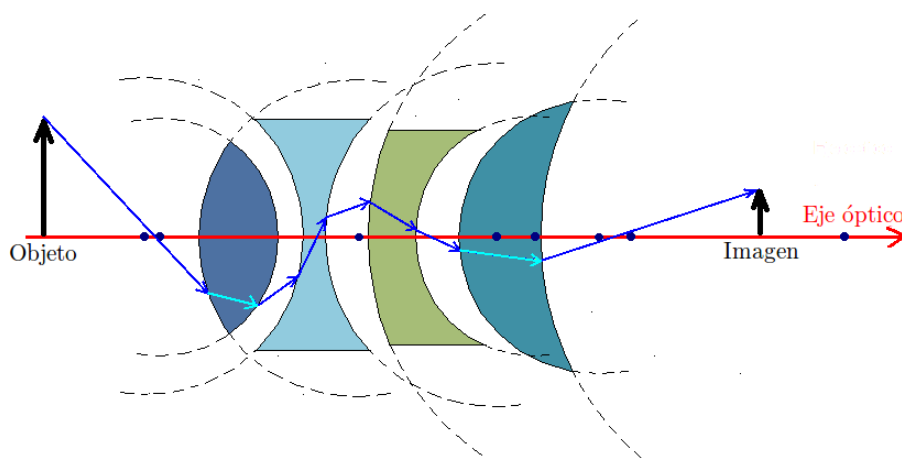


Figura 4.1: Sistema de lentes centrado con superficies esféricas (simetría de revolución).

En la Figura 4.2 se presentan los ángulos y distancias que utiliza este método como datos de entrada para calcular el trazo de rayos. A es el punto objeto, B es el punto imagen, V es la intersección de la interfase con el eje óptico y se le conoce como vértice, C es el centro de curvatura y P es el punto de incidencia del rayo con la interfase. Con respecto a las distancias tenemos que: L es la distancia del objeto al vértice, L' es la distancia del vértice al objeto y R es el radio de la lente esférica.

Por último los ángulos de interés son: U (formado entre el rayo incidente, que proviene del punto objeto, y el eje óptico), U' (el ángulo formado entre el rayo refractado y el eje óptico), α (ángulo auxiliar comprendido entre el eje óptico y el segmento radial CP), β (ángulo auxiliar comprendido entre los rayos incidente y refractado), I e I' (el ángulo de incidencia y el ángulo de transmisión o refracción, respectivamente, medidos a partir de la dirección normal N a la superficie en el punto de incidencia).

El ángulo U' y la distancia L' son los datos que se desean obtener por el método L-U, pues con ellos se puede conocer la dirección del rayo incidente después de ser desviado por la lente (rayo refractado). Usando propiedades de triángulos semejantes y leyes de los senos, se obtienen las *ecuaciones de refracción* presentadas a continuación:

$$\begin{aligned} \text{sen } I &= \frac{R - L}{R} \text{sen } U, & \text{sen } I' &= \frac{n}{n'} \text{sen } I, \\ U' &= U - I + I', & L' &= R \left(1 - \frac{\text{sen } I'}{\text{sen } U'} \right), \end{aligned} \quad (4.1)$$

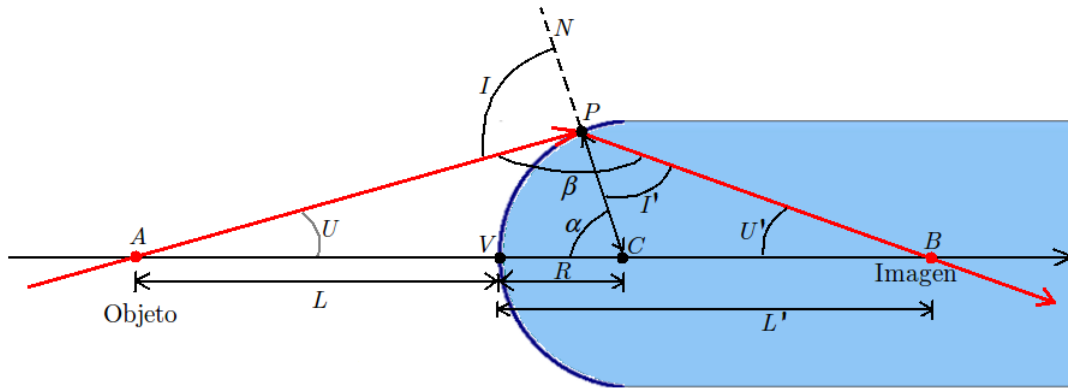


Figura 4.2: Parámetros usados por el método L-U.

donde, n y n' son los índices de refracción del primer y segundo medio. Para el diagrama de la Figura 4.2, el primer medio (donde se encuentra el punto objeto) es el aire y el segundo medio es material del lente, es decir, el primer índice de refracción es mayor que el segundo lo que implica que los rayos tenderán a doblarse hacia el segundo medios, en el cual viajan con mayor facilidad.

En las ecuaciones (4.1) podemos notar que la posición del punto imagen a una distancia L' sobre el eje óptico y respecto a la interfase esta en función de L , U , n , n' , y R ; por lo tanto, los rayos que parten de un mismo punto objeto (L fija) no se proyectarán en un mismo punto imagen ya que L' cambiará en tanto cambie I' . Esto nos indica que no tendremos una imagen estigmática, es decir, un punto en el objeto no se proyectará en un punto en la imagen, si no que presentará un emborronamiento o desenfoque. En vez de tener una imagen formada por puntos bien definidos, se formará con puntos extendidos o manchas pues los rayos provenientes del objeto se dispersan, no quedando concentrados en un único punto.

Ahora bien, teniendo en cuenta que la función seno puede descomponerse en la serie de potencias:

$$\text{sen } \theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \dots, \quad (4.2)$$

podemos simplificar las ecuaciones de refracción sustituyendo dicha operación por el primer elemento de la serie de potencia, teniendo presente que la aproximación es válida únicamente para ángulos pequeños. Sustituyendo la aproximación $\text{sen } \theta \approx \theta$ en las ecuaciones (4.1), obtenemos:

$$i = \frac{R-L}{R}u, \quad i' = \frac{n}{n'}i,$$

$$u' = u - i + i', \quad l' = R \left(1 - \frac{i'}{u'} \right). \quad (4.3)$$

A las ecuaciones (4.3) se les conoce como *ecuaciones de trazo de rayos en la aproximación paraxial* (o de primer orden), porque restringen la validez de esta aproximación con ángulos pequeños, es decir, aplica únicamente para rayos muy cercanos al eje óptico, de ahí el nombre de *paraxial*. En la práctica, este ángulo debe estar preferentemente entre 5° y 10° , sin embargo ángulos de 15° pueden considerarse dentro de la aproximación paraxial. En nuestra aplicación se controla la entrada al sistema óptico de rayos paraxiales mediando el ajuste de la pupila de la cámara digital. Se estimaron ángulos de 10.2° , aproximadamente, por lo que consideramos válida la aproximación ideal para nuestro caso. Por convención, las variables a primer orden o paraxiales se escriben en minúsculas. Mediante un proceso algebraico podemos reducir las ecuaciones de refracción en la aproximación paraxial:

$$\frac{n'}{l'} - \frac{n}{l} = \frac{n' - n}{R}, \quad (4.4)$$

donde, l' depende de l , n , n' y R , y es independiente de u , por lo tanto, todos los rayos que provienen de un mismo punto objeto se proyectan en un mismo punto imagen. Esto quiere decir que los rayos paraxiales sí forman imágenes estigmáticas o perfectas.

A toda desviación de la imagen estigmática causada por los sistemas de lentes se conocen como *aberraciones*. En la Figura 4.3 podemos observar la aberración transversal (Δx) y la longitudinal (ΔL), las cuales se miden tomando como referencia el punto donde se interceptan el eje óptico con los rayos paraxiales, que son los que no presentan aberración por lo que se les considera ideales.

En el modelado de las aberraciones en los sistemas de lentes centrados utilizaremos tres marcos de referencia. El primer marco se encuentra en el plano imagen, el segundo en el plano de refracción para una sola superficie y el tercero en el plano de la imagen. El tercer eje es común a los tres sistemas coordenados y lo llamaremos eje Z . Las coordenadas en plano objeto se identifican con las letras X e Y , las del plano de refracción (para un sistema complejo que involucra varias superficies, se utilizan las coordenadas del plano de la pupila) se identifican con las letras griegas ξ e η ; y por último, para las coordenadas del plano imagen usaremos las letras x e y . Un diagrama de estos sistemas coordenados se presenta en la Figura 4.4.

Igualmente podemos observar que, cuando el ángulo formado por el rayo proveniente del punto objeto con el eje óptico (U) cambia de signo, Δx cambia de signo y ΔL siempre

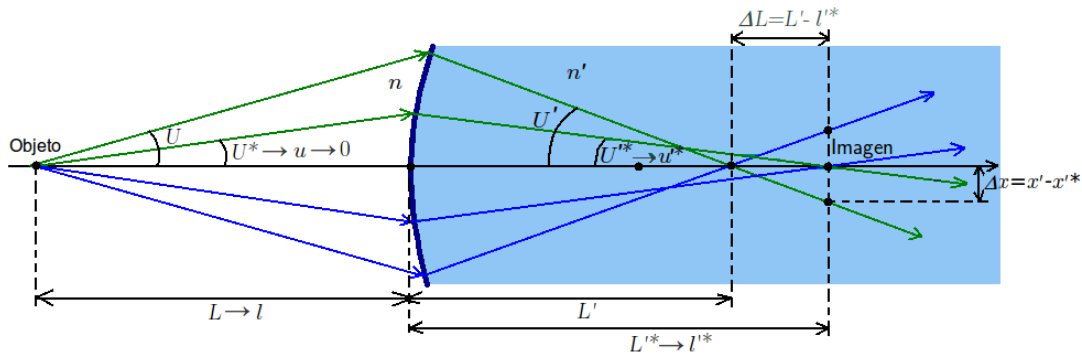


Figura 4.3: Diagrama de trazado de rayos en el cual se distingue la aberración transversal ΔL y longitudinal Δx .

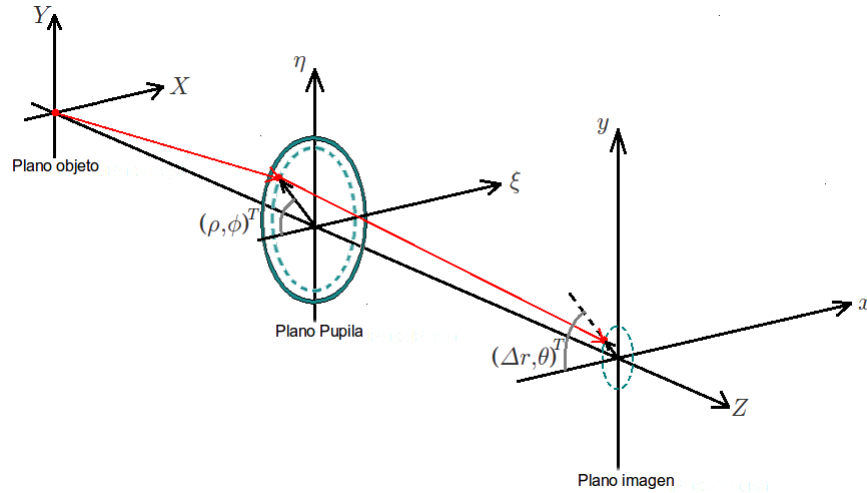


Figura 4.4: Marcos de referencia de los polinomios de aberración.

conserva el mismo signo. Debido a este comportamiento en los signos y a que L , n , n' y R son fijos, tenemos que Δx debe ser función impar de U o bien, de la altura ξ en que intercepta el rayo a la superficie del lente. Tal función es de la forma:

$$\Delta x = A_0\xi + A_1\xi^3 + A_2\xi^5 + A_3\xi^7 + A_4\xi^9 + \dots, \quad (4.5)$$

donde, A_i son los coeficientes del polinomio de grado impar, del cual $A_0 = 0$ pues pertenece al término a primer orden o paraxial y en él no hay aberraciones. Los valores de los coeficientes dependen de las características del sistema óptico. Tomando hasta el segundo término, que tiene grado tres, obtenemos una muy buena aproximación sin complicar demasiado los

cálculos. Con esta aproximación y el hecho de que los sistemas de lentes con los que trabajaremos tienen la simetría de revolución, se obtienen *los polinomios de aberración a tercer orden* presentados a continuación:

$$\begin{aligned} \Delta x = & A_E (\xi^2 + \eta^2) \xi + A_C [X (\xi^2 + \eta^2) + 2\xi (X\xi + Y\eta)] + \\ & 2A_A (X^2\xi + XY\eta) + A_{CC} (X^2 + Y^2) \xi + A_D (X^2 + Y^2) X, \end{aligned} \quad (4.6)$$

$$\begin{aligned} \Delta y = & A_E (\xi^2 + \eta^2) \eta + A_C [Y (\xi^2 + \eta^2) + 2\eta (X\xi + Y\eta)] + \\ & 2A_A (XY\xi + Y^2\eta) + A_{CC} (X^2 + Y^2) \eta + A_D (X^2 + Y^2) Y, \end{aligned} \quad (4.7)$$

donde, los coeficientes representan a cada una de las clases de aberraciones primarias o de tercer orden y se les denomina como sigue:

- A_E es *Esfericidad*
- A_C es *Coma*
- A_A y A_{CC} son *Astigmatismo* y *curvatura de campo*, respectivamente y
- A_D es *Distorsión*.

La caracterización de estas aberraciones primarias se mostrarán en secciones posteriores.

4.2. Efecto de las aberraciones

Para poder analizar el efecto que cada clase de aberración ocasiona en la imagen, se evalúan los polinomios haciendo cero todos los coeficientes excepto el que se encuentra bajo análisis, y suponiendo que el punto objeto se encuentra descansando sobre el eje X (sin pérdida de generalidad). Sin entrar en mayores detalles, describiremos los efectos y características de cada clase de aberración.

A modo de demostración se presentan gráficas de la simulación del polinomio de aberración completo tomando como plano objeto a una rejilla regular de puntos artificiales presentada en la Figura 4.5(a) y las coordenadas en el plano de la pupila consideradas para su evaluación se muestran en la Figura 4.5(b), donde cada punto representa a un rayo proveniente de un punto objeto que ingresa al sistema de lentes.

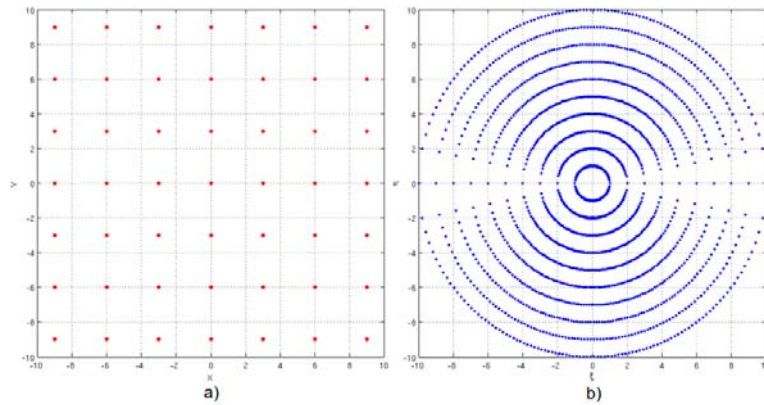


Figura 4.5: (a) Rejilla artificial de puntos, que representa al plano objeto, usado para simular el efecto de las aberraciones y (b) plano de la pupila marcado con las posiciones evaluadas en el polinomio de aberración.

4.2.1. Esfericidad

En este caso consideramos $A_E \neq 0$; $A_C = A_A = A_{CC} = A_D = 0$, por lo tanto, las ecuaciones (4.6) y (4.7) se pueden reescribir como sigue:

$$\begin{aligned}\Delta x &= A_E(\xi^2 + \eta^2)\xi = A_E\rho^3 \cos \phi, \\ \Delta y &= A_E(\xi^2 + \eta^2)\eta = A_E\rho^3 \sin \phi,\end{aligned}\tag{4.8}$$

cuya versión en coordenadas polares es:

$$\Delta r = A_E\rho^3,\tag{4.9}$$

la cual, si ρ permanece constante es la ecuación de un círculo.

De la ecuación (4.9) podemos ver que este tipo de aberración es la única que no depende de la posición del objeto, tiene el mismo valor en todo el campo de observación y que se presenta en menor grado si la pupila está más cerrada. El resultado de aplicar esfericidad a la imagen de un objeto como la rejilla de prueba, se presenta en la Figura 4.6.

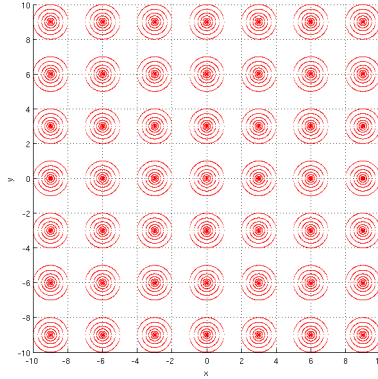


Figura 4.6: Imagen de la rejilla con aberración de esfericidad.

4.2.2. Coma

En este caso consideramos $A_C \neq 0$; $A_E = A_A = A_{CC} = A_D = 0$, por lo tanto, después de sustituir los coeficientes en las ecuaciones (4.6) y (4.7) y simplificarlas, obtenemos:

$$\begin{aligned}\Delta x &= A_C X \rho^2 (2 + \cos 2\phi), \\ \Delta y &= A_C X \rho^2 \sin 2\phi.\end{aligned}\quad (4.10)$$

Elevando ambas ecuaciones 4.10 al cuadrado y sumándolas obtenemos:

$$(\Delta x + 2A_C X \rho^2)^2 + (\Delta y)^2 = A_C^2 X^2 \rho^4, \quad (4.11)$$

que consiste en la ecuación de una circunferencia, si dejamos a ρ constante, con radio $R = A_C X \rho^2$ y cuyo centro se encuentra en $(2R, 0)$. De igual forma, la ecuación (4.11) describe círculos con centro y radio variables con la distancia del rayo al eje óptico, en la pupila. El resultado de aplicar la aberración de coma a la imagen de la rejilla de prueba, se presenta en la Figura 4.7.

4.2.3. Astigmatismo y curvatura de campo

En este caso consideramos $A_A, A_{CC} \neq 0$; $A_E = A_C = A_D = 0$, por lo tanto, las ecuaciones (4.6) y (4.7) se reescriben como sigue:

$$\begin{aligned}\Delta x &= (2A_A + A_{CC})X^2\xi, \\ \Delta y &= A_{CC}X^2\eta.\end{aligned}\quad (4.12)$$

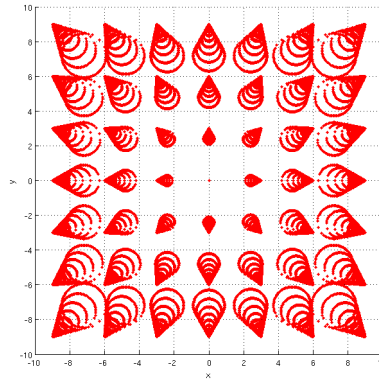


Figura 4.7: Imagen de la rejilla con aberración de coma.

Nuevamente, si elevamos al cuadrado ambas ecuaciones y después las sumamos, obtenemos la ecuación siguiente:

$$\frac{(\Delta x)^2}{(2A_A + A_{CC})^2 X^4 \rho^2} + \frac{(\Delta y)^2}{A_{CC}^2 X^4 \rho^2} = 1, \quad (4.13)$$

la cual, consiste en la ecuación de una elipse paralela a los ejes XY , en el plano de la imagen, cuyos semiejes son $a = (2A_A + A_{CC})X^2\rho$ y $b = A_{CC}X^2\rho$. Combinando valores para A_A y A_{CC} , se pueden presentar los siguientes casos:

1. Si $A_A = 0$, $a, b = A_{CC}X^2\rho$, la elipse se convierte en un círculo ($D \neq 0$). Se conoce como el mejor foco.
2. Si $A_{CC} = 0$, $b = 0$; la elipse degenera en una línea horizontal. Se conoce como el foco sagital.
3. Si $A_{CC} = -2A_A$, $a = 0$; la elipse degenera en una línea vertical. Se conoce como el foco tangencial.

El comportamiento del astigmatismo con respecto a cambios en la posición en Z del plano de la imagen se presenta en la Figura 4.8, y el resultado de aplicar astigmatismo, en los tres casos especiales, a la imagen de la rejilla de prueba, se presenta en la Figura 4.9.

4.2.4. Distorsión

En este caso consideramos $A_D \neq 0$; $A_E = A_C = A_A = A_{CC} = 0$, sin embargo, en esta ocasión no supondremos que la posición del punto objeto está sobre el eje X . Las ecuaciones

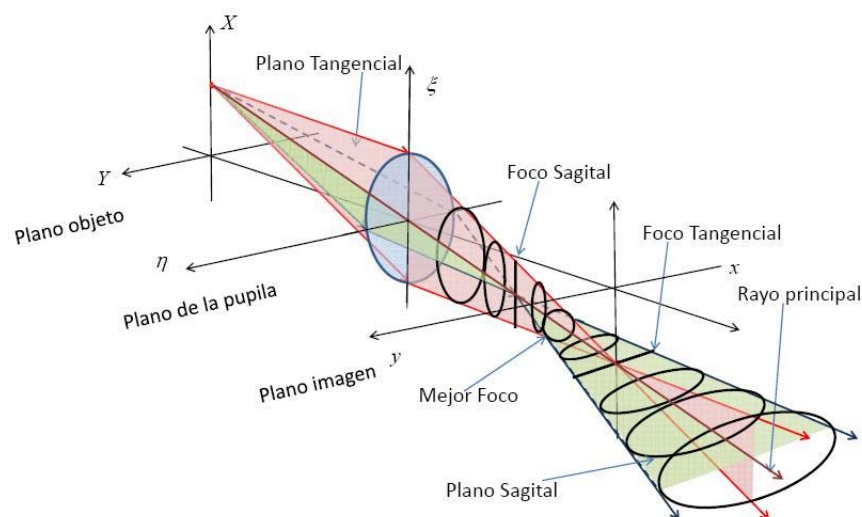


Figura 4.8: Efecto del astigmatismo cuando el plano de la imagen se mueve a lo largo del eje Z [Díaz Uribe, 2009].

(4.6) y (4.7) se reescriben, después de haberlas simplificado, como sigue:

$$\begin{aligned}\Delta x &= A_D(X^2 + Y^2)X, \\ \Delta y &= A_D(X^2 + Y^2)Y.\end{aligned}\tag{4.14}$$

De las ecuaciones 4.14 podemos observar que este tipo de aberración no depende de ξ ni de η , entonces, la imagen de un punto objeto es también un punto, en otras palabras, la distorsión no genera emborronamiento en la imagen simplemente desplaza de su posición correcta a los puntos proyectados en el plano imagen. El resultado de aplicar distorsión a la rejilla de prueba, en sus dos casos característicos, se presenta en la Figura 4.10.

4.3. Ametropías

Otro factor que impide obtener imágenes estigmáticas de la retina con la cámara *fundus*, son los problemas de enfoque del ojo humano. El ojo *emétrope* es aquel que tiene la capacidad de enfocar todos los rayos provenientes de un objeto distante (rayos paralelos) en la superficie de la retina en condiciones de relajación, es decir, que no se encuentra en acción el fenómeno de *acomodación* (función de enfoque fino llevado a cabo por el cristalino que se encuentra sujetado por unos ligamentos que están conectados con los *músculos ciliares*, los cuales al contraerse modifican el grosor del cristalino y la curvatura de sus superficies,

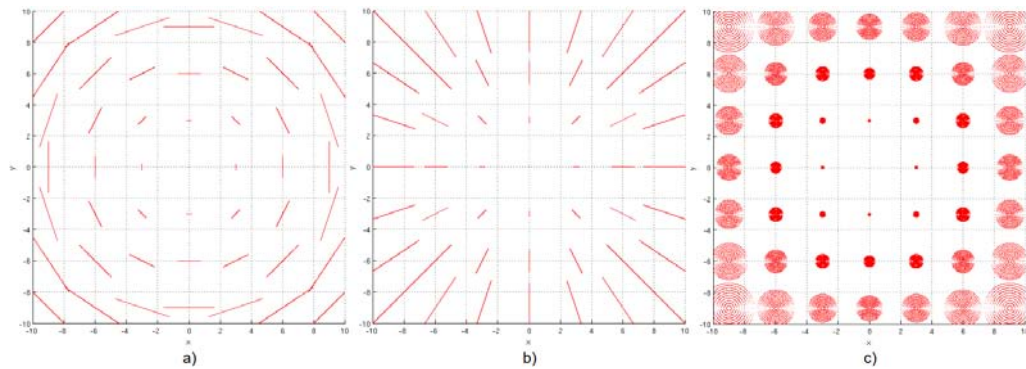


Figura 4.9: Imagen de la rejilla con astigmatismo en (a) foco tangencial, (b) foco sagital y (c) mejor foco.

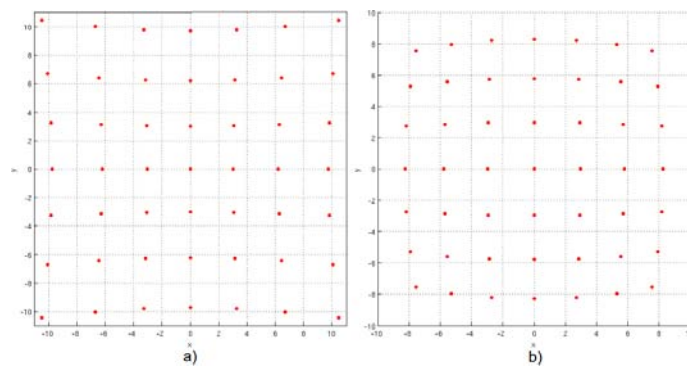


Figura 4.10: Imagen de la rejilla con distorsión de (a) corset o cojín y (b) de barril.

ocasionando que su distancia focal sea menor).

Lo anterior significa que en un ojo normal (*emétrope*) sin acomodación, la superficie de la retina se encuentra a una distancia del cristalino igual a la distancia focal de este último. Sin embargo, esta condición no siempre ocurre, teniendo que el punto focal del cristalino no se encuentra en la retina, ocasionando padecimientos como hipermetropía, miopía, presbicia y astigmatismo. El principal motivo para la presencia de estos padecimientos son modificaciones en el tamaño del ojo, y en específico cambios en su longitud, cambiando así la distancia entre el cristalino y la retina.

A continuación describiremos brevemente estos defectos refractivos (miopía, hipermetropía y astigmatismo) por separado, para dejar claro en que consisten y como se corrigen.

4.3.1. Miopía

La miopía se presenta cuando, sin acomodación, el foco del sistema de lentes queda enfrente de la retina (véase Figura 4.11(a)), en otras palabras, la distancia focal es muy pequeña para la separación entre el cristalino y la retina. También puede interpretarse como que la potencia de la óptica del ojo es demasiado grande para la longitud del ojo. Los objetos lejanos se proyectan en un plano imagen posicionado delante de la retina. El punto lejano (punto objeto más lejano que puede observarse nítidamente por un ojo *emétrope* sin acomodación) está más cerca que el infinito (ésto depende de cada persona pues cada uno tiene diferente cantidad de miopía) y los puntos que se encuentren más lejos que él, se proyectan borrosos en la superficie de retina. Los objetos apartados se podrán ver enfocados siempre y cuando estén a una distancia menor como para que el cristalino relajado proyecte en la retina todos los rayos provenientes del punto objeto (véase Figura 4.11(b))

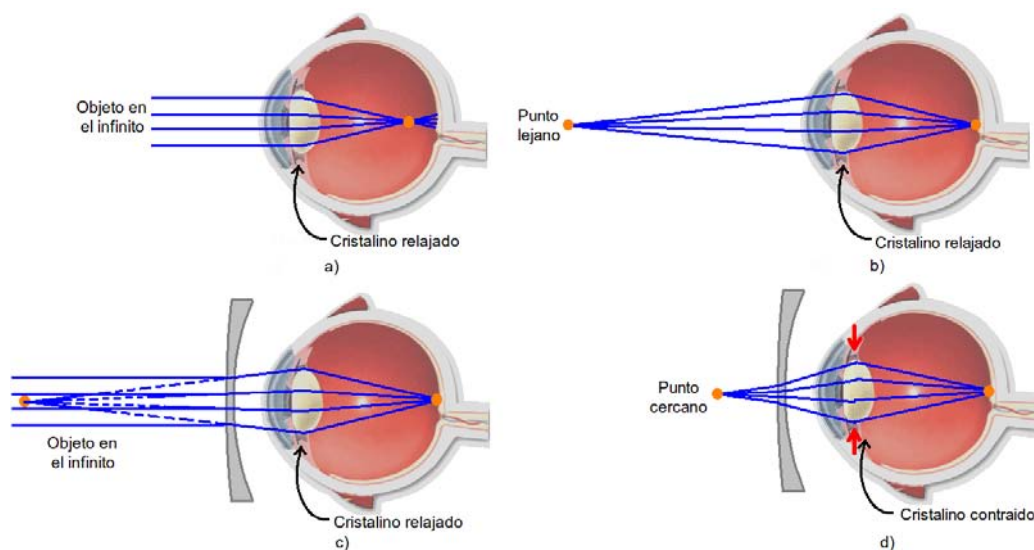


Figura 4.11: Ojo miope observando (a) un punto objeto en el infinito y (b) un punto lejano. En (c) vemos la corrección de miopía con una lente negativa y en (d) vemos como cuando se observa un punto objeto cercano, la acomodación permite ajustar la distancia focal del cristalino para que el punto imagen caiga en la retina.

Por el contrario, los objetos que se encuentran a corta distancia se pueden ver claramente con este padecimiento. La miopía se corrige empleando lentes negativas o lentes divergentes, pues lo que se pretende es doblar los rayos paralelos en forma de abanico como si provinieran

de un objeto más cercano. Así obtenemos un punto focal de la combinación de la óptica del ojo y del lente externo que se encuentra justo en la retina (véase Figura 4.11(c)). Cuando se usa la lente correctiva y se observa un punto objeto cercano, el fenómeno de acomodación evita que la distancia imagen sea mayor que la separación entre el cristalino y la retina, aumentando la potencia del primero haciendo que los puntos imagen caigan en la retina, como se ilustra en la Figura 4.11(d).

4.3.2. Hipermetropía

Este defecto refractivo, conocido en inglés como *hyperopia*, en el cual la distancia focal de la óptica del ojo cae detrás de la retina (véase la Figura 4.12(a)), es decir, la distancia focal es mayor que la separación de cristalino y la superficie de la retina. La causa de la hipermetropía es el acortamiento de la longitud del ojo, es decir, el cristalino se encuentra demasiado cerca de la retina. Si la hipermetropía no es muy severa, se pueden ver puntos objeto lejanos mediante el fenómeno de acomodación, como se ve en la Figura 4.12(b). Sin embargo cuando el objeto está muy próximo, la capacidad de la acomodación para aumentar la potencia refractiva del cristalino llega a su límite, presentando de nuevo vista borrosa (véase la Figura 4.12(c)).

Para corregir este defecto, se antepone una lente positiva o convergente al ojo, la cual dobla los rayos provenientes del punto objeto tendiéndolos a ser paralelos, con lo que se crea el efecto de alejamiento del punto próximo o cercano. Con lo anterior se le permite al ojo bajo acomodación visualizar nítidamente el punto objeto, pues su potencia vuelve a ser adecuada, como se muestra en la Figura 4.12(d).

4.3.3. Astigmatismo

Este defecto refractivo, forma parte de las diferentes clases de aberraciones descritas en la sección 4.2 y fué modelada por medio del *polinomio de aberración a tercer orden*. En el caso práctico del ojo humano, se debe a la curvatura desigual de la córnea, es decir no tiene simetría de revolución. Una forma de apreciarlo es pensar que se hacen cortes en la córnea con planos meridionales, entonces los perfiles obtenidos no tendrán la misma medida de curvatura. Podemos decir que la potencia del sistema de lentes cambia respecto a la orientación de los planos meridionales en los cuales se encuentren los rayos, por lo que se identifican dos planos meridionales principales. El primero donde se presente la curvatura o potencia refractiva máxima y el segundo donde se presente la mínima curvatura.

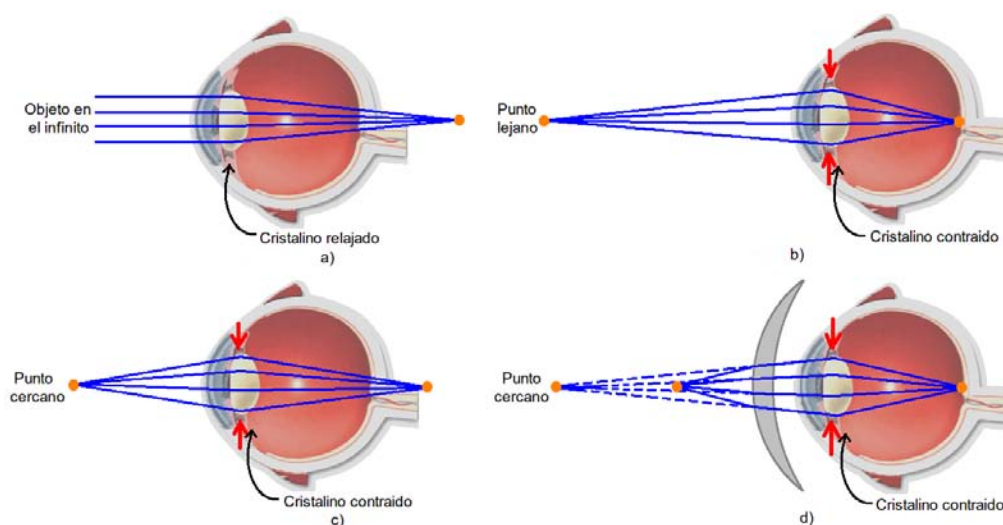


Figura 4.12: Ojo hipermetrope observando (a) un punto en el infinito, (b) un punto alejado pero a menor distancia que el punto en infinito con acomodación y (c) un punto muy cercano al ojo donde se sobre pasa el limite de la acomodación. Y en (d) vemos la corrección de este defecto por medio de lentes positivas.

Se dice que el astigmatismo es *regular* si los dos planos meridionales principales (curvatura mínima y máxima) son ortogonales, en este caso el astigmatismo es corregible. Por el contrario, si no son ortogonales, se dice que es irregular y es muy complicado de corregir. Los lentes usados para corregir el astigmatismo regular se les conoce como *anamórficas* y su característica distintiva es que cuentan con dos diferentes potencias refractivas o distancias focales en sus planos meridionales principales. La Figura 4.13 muestra un ejemplo, generado artificialmente por computadora, del efecto de las lentes anamórficas. Este ejemplo muestra el cambio en la magnificación según la orientación, pero no se muestra el efecto de desenfoque diferenciado entre las direcciones X e Y.

4.4. Profundidad de campo

Otro fenómeno ocasionado por los sistemas de lentes en las cámaras es la profundidad de campo. Su efecto ocasiona que no todos los objetos de la imagen aparezcan nítidos o enfocados, debido a que los objetos no se encuentran a la misma profundidad o distancia en Z respecto a la cámara. Se dice que una cámara tiene poca profundidad de campo, si a pequeñas variaciones en la coordenada Z de los puntos de la escena, algunos de estos puntos



Figura 4.13: Efecto de los lentes anamórficos, aparentemente contrae la imagen de la escena original en dirección horizontal. Imagen tomada de *Computer desktop encyclopedia 2004, The computer language Co. Inc.*

aparecen emborronados y otros nítidos. Por el contrario, una cámara tiene mucha profundidad de campo cuando todos los objetos de la escena aparecen nítidos en la imagen, a pesar de que la variación en sus coordenadas Z es grande.

En la Figura 4.14 se muestra un diagrama muy simplificado del fenómeno de la profundidad de campo, en la cual aparecen tres puntos sobre el eje óptico con diferentes profundidades. El plano de la imagen se encuentra a una distancia al lente adecuada para enfocar el punto verde, sin embargo, los puntos rojo y azul aparecen desenfocados pues no se encuentran a la misma distancia que el punto verde. Como resultado, los puntos rojo y azul se proyectan como círculos de diámetro considerable y no como puntos, lo que es deseable.

Para aumentar la profundidad de campo y disminuir el emborronamiento de partes de la escena, se necesita reducir el área efectiva del lente (o sistema de lentes) ajustando la apertura del diafragma. La Figura 4.15 muestra el efecto de disminuir la apertura. Los rayos provenientes de los puntos objeto que llegan al lente en las regiones más alejadas al eje óptico son rechazados. Solamente los rayos más cercanos al eje óptico llegan a la imagen, produciendo una reducción en el diámetro de los círculos que representan las proyecciones de los puntos fuera de foco.

4.5. Consideraciones en la captura para minimización de aberraciones

El impedimento de tomar imágenes de un patrón de referencia pasando a través de la óptica del ojo no nos permite caracterizar el efecto de las aberraciones en imágenes previa-

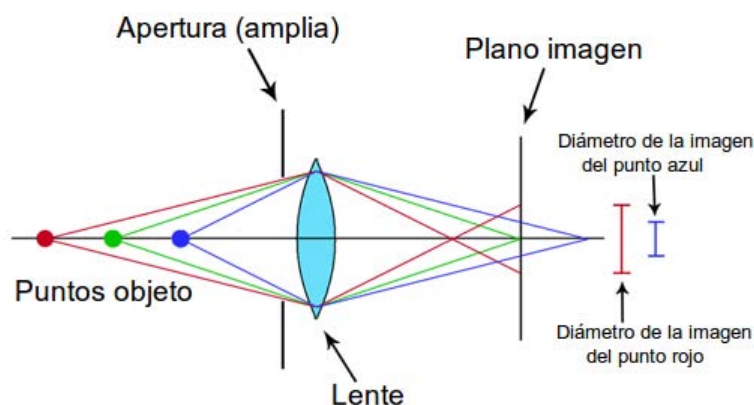


Figura 4.14: Fenomeno de la profundidad de campo. Se muestra que únicamente el punto verde aparece proyectado como un punto en la imagen. Los puntos rojo y azul que se encuentran a diferente profundidad respecto al punto verde, se proyectan como círculos en la imagen, es decir, sufren de desenfoco. Figura tomada de *www.trustedreviews.com*.

mente capturadas, como lo hacemos en el caso de distorsión con el sistema *ADICOS*. De ahí la propuesta de corregir o minimizar las aberraciones durante la captura, modificando ciertos parámetros de la cámara *fundus* según nos indique el comportamiento de las aberraciones por su naturaleza.

La presencia de aberraciones se traduce en emborronamiento de la imagen capturada (exceptuando distorsión), por lo tanto, si ajustamos la distancia focal de la cámara de fondo de ojo podemos corregir problemas refractivos como la miopía, hipermetropía y presbicia, ya que consisten en problemas de desenfoco. Podemos decir que si al capturar imágenes de retina éstas aparecen completamente nítidas, entonces las aberraciones son mínimas y pueden despreciarse. El mayor grado de nitidez en las imágenes es identificado por simple inspección, pues la aplicación de métricas cuantitativas de la calidad de las imágenes queda fuera de los alcances de esta tesis.

Siendo más rigurosos en el criterio de despreciar a las aberraciones y considerando que las aberraciones están presentes (aún siendo mínimas), podemos justificar la nulificación de los efectos de la esfericidad, astigmatismo y curvatura de campo por ser simétricos no cambian sus centros de gravedad. Aún con la presencia de las aberraciones anteriores, el proceso de segmentación realizado por el sistema RISA [Martínez-Pérez *et al.*, 2007] usa filtros gaussianos derivativos isotrópicos y después de un proceso de erosión para obtener los esqueletos, los cuales no cambiarán los resultados pues los puntos del objeto se proyectan en áreas más

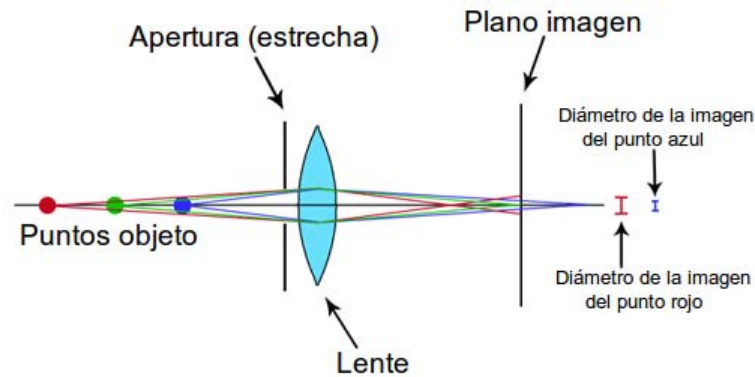


Figura 4.15: Aumento en la profundidad de campo. El desenfoque disminuye en la imagen de los puntos rojo y azul, como resultado de reducir el tamaño de la pupila. El diámetro de los círculos proyectados se reduce significativamente. Figura tomada de *www.trustedreviews.com*.

extensas a un punto (manchas) simétricas en la imagen.

La aberración de coma al no ser simétrica en ambos ejes modifica la posición del centro de gravedad de los objetos segmentados en la imagen, por lo que su minimización es de mayor importancia. Observando las ecuaciones (4.6) y (4.7), sabemos que la aberración de coma está en función del área efectiva del lente. Reduciendo el área mediante un diafragma de iris la aberración de coma disminuye, así como también el desenfoque por profundidad de campo. Para evitar el oscurecimiento de las imágenes por el cierre del diafragma se aumenta el tiempo de exposición o la intensidad de la fuente de luz de ser posible. Con lo mencionado anteriormente podemos considerar mínimas las aberraciones del sistema óptico del ojo. Es importante mencionar que el cierre del diafragma tiene un límite, puesto que, el diámetro de la pupila determina la contribución de la difracción y de las aberraciones a la calidad de las imágenes obtenidas. Cuando la abertura de la pupila es muy pequeña, la calidad óptica depende de la difracción ocasionando una pobre resolución espacial, sin embargo, el efecto de las aberraciones es casi nulo. Cuando la pupila tiene una gran abertura, la presencia de las aberraciones domina sobre el efecto de la difracción. En nuestros experimentos, se redujo la abertura de la pupila tanto como fue posible sin presentar artefactos por difracción.

El modelado de las aberraciones ópticas del ojo se puede llevar acabo por métodos experimentales como son la técnica de doble paso (*D-P*), el sensor de Shack-Hartmann (*SH*), las técnicas de trazado de rayos por láser (*Laser Ray Tracing*, o *LRT* por sus siglas en inglés) y el refractómetro de resolución espacial (*Spatially Resolved Refractometer*, o *SRR* por sus

siglas en inglés), como se describe en [Marcos, 2005]. Las técnicas anteriores son propias de temas de estudio de la óptica geométrica, requieren de equipo especializado y aunque se estimen los coeficientes de las aberraciones mediante alguno de ellos, corregirlos en imágenes de un área considerablemente grande de la retina, es un proceso numérico complejo y computacionalmente pesado. Es por ello que los consideramos fuera de los alcances de este trabajo de tesis, para mayores detalles del modelado de aberraciones en el ojo humano y la medida de la calidad óptica del mismo es importante consultar [Marcos, 2005].

Capítulo 5

Métodos de Auto-calibración

En este capítulo presentamos la justificación de emplear el proceso de auto-calibración cuando lidiamos con el problema de modelar el sistema *cámara-ojo*, así como también, se describen dos métodos de auto-calibración con la finalidad de ser comparados entre sí, y determinar cuál es más conveniente para ser empleado en la obtención de la matriz de calibración de dicho sistema. Estos métodos basan la estimación de los parámetros internos en restricciones, cuyo fundamento teórico es muy distinto, de ahí la importancia de la comparación entre ellos.

5.1. El uso de auto-calibración en el proceso global de reconstrucción

Primeramente, es importante saber que el proceso de auto-calibración consiste en estimar los parámetros intrínsecos a partir de un conjunto de imágenes capturadas con una o más cámaras no calibradas, es decir, no necesitamos usar patrones de calibración para calcular las propiedades métricas de las cámaras. Haciendo uso de imágenes de una escena cualquiera e imponiendo condiciones sobre los parámetros intrínsecos y/o extrínsecos, podemos realizar esta tarea a pesar de que la cámara haya sufrido cambios en enfoque, acercamiento o alejamiento (*zooming*), o bien, haya sido movida de su posición original. Entiéndase por “diferentes cámaras” no sólo a distintos equipos de captura, sino también, a cambios en los parámetros internos de la misma. Es por lo anterior, que cuando nos enfrentamos al problema de estimar los parámetros intrínsecos (y adicionalmente los extrínsecos) de un sistema óptico compuesto y tan complejo como lo es el del sistema *cámara-ojo* (véase sección 1.4), consideramos a la auto-calibración como una posible solución al problema.

Otro motivo para el uso de auto-calibración, en nuestra aplicación específica, es la ventaja que presenta cuando las diferentes vistas de la escena fueron capturadas con diferentes cámaras. Pues la estimación de las matrices de calibración de cada una de las cámaras se realiza por métodos de ajuste en paquete (*bundle adjustment*). Lo cual es aplicable a las imágenes de fondo de ojo usadas para la reconstrucción, ya que al fotografiar distintas regiones de la retina o mover el punto de vista de una misma región, los ejes ópticos de la cámara *fundus* y del ojo cambian su posición relativa entre ellos, modificándose así la posición del punto principal. De igual forma, al manipular la cámara *fundus* se ocasionan cambios en la distancia de separación entre la óptica del ojo y los lentes de la cámara, resultando en alteraciones a la distancia focal del sistema global.

5.2. Estructura algebraica de la auto-calibración

Cuando no se conoce la calibración de ninguna de las cámaras con las que se capturó una misma escena, ni la posición relativa de ellas con respecto a una, la reconstrucción 3D obtenida con los puntos en correspondencia entre las imágenes incurre en una ambigüedad que ocasiona deformaciones en el objeto reconstruido. A esto se le conoce como ambigüedad proyectiva, en otras palabras, los puntos en 3D reconstruidos (\mathbf{x}_{rj}) y los de la escena real (\mathbf{x}_{wj}) difieren (adicionalmente al factor de escala desconocido) por una transformación proyectiva en \mathbb{P}^3 que se conoce como homografía (\mathbf{H}). Bajo estas condiciones, en el mejor de los casos, los puntos 3D y las matrices de proyección en perspectiva ($\mathbf{\Pi}_i$) de las cámaras que serán calculadas, estarán bajo una transformación proyectiva. Es posible estimar \mathbf{x}_{rj} y $\mathbf{\Pi}_i$ bajo los efectos de \mathbf{H} utilizando los mismo puntos en las imágenes para la reconstrucción, esto se demuestra en la siguiente ecuación:

$$\mathbf{\Pi}_i \mathbf{x}_{rj} = (\mathbf{\Pi}_i \mathbf{H}^{-1})(\mathbf{H} \mathbf{x}_{rj}). \quad (5.1)$$

Ahora bien, el problema de la auto-calibración inicia suponiendo que se tiene la reconstrucción proyectiva de la escena, es decir, la matriz de proyección en perspectiva y la posición de los puntos reconstruidos y su notación es: $\{\mathbf{\Pi}_i, \mathbf{x}_{rj}\}$. Se desea calcular la \mathbf{H} que rectifique dicha reconstrucción ($\{\mathbf{\Pi}_i \mathbf{H}, \mathbf{H}^{-1} \mathbf{x}_{rj}\}$) de tal forma que recuperemos las propiedades métricas.

En base a la idea anterior definamos un conjunto de m cámaras cuyas matrices de proyección son $\mathbf{\Pi}_{E,i}$ para $i = 1, \dots, m$ y donde el subíndice E significa que las cámaras están calibradas y que el sistema de referencia del mundo es euclidiano (métrico) y su composición se muestra a continuación:

$$\mathbf{\Pi}_{Ei} = \mathbf{K}_i [\mathbf{R}_i | \mathbf{t}_i], \quad i = 1, \dots, m. \quad (5.2)$$

Como ya sabemos, de la reconstrucción proyectiva podemos recuperar las matrices de las cámaras no calibradas $\mathbf{\Pi}_{P_i}$, donde el subíndice P significa que fueron obtenidas de la reconstrucción proyectiva, que se relacionan con $\mathbf{\Pi}_{E_i}$ mediante \mathbf{H} (como se muestra en la ecuación (5.3)), la cual tenemos que calcular.

$$\mathbf{\Pi}_{E_i} = \mathbf{\Pi}_{P_i} \mathbf{H}, \quad i = 1, \dots, m. \quad (5.3)$$

Definiendo nuestras cámaras de referencia, tenemos que $\mathbf{\Pi}_{E_1} = \mathbf{K}_1 [\mathbf{I} | \mathbf{0}]$ y $\mathbf{\Pi}_{P,1} = [\mathbf{I} | \mathbf{0}]$ (forma canónica, donde $f = 1$). Únicamente para esta ecuación se redefine \mathbf{I} como la matriz identidad. Continuando, se despeja \mathbf{H} de la ecuación (5.3) para las cámaras de referencia, obtenemos la siguiente ecuación, donde se puede observar la homografía en su forma de bloque,

$$\mathbf{H} = \begin{bmatrix} \mathbf{K}_1 & \mathbf{0} \\ \mathbf{s}^T & 1 \end{bmatrix}. \quad (5.4)$$

Suponemos que $k = 1$ para fijar el factor de escala. El vector \mathbf{s} , que es causante del efecto no-lineal de las transformaciones proyectivas y es el que permite a estas transformaciones modelar los puntos de fuga [Hartley y Zisserman, 2004], nos determina junto con \mathbf{K}_1 la posición del plano al infinito (π_∞) en la reconstrucción proyectiva, como podemos ver a continuación:

$$\pi_\infty = \mathbf{H}^{-T} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\mathbf{K}_1^{-T} \mathbf{s} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{a} \\ 1 \end{bmatrix}, \quad (5.5)$$

donde, $\mathbf{a} = -\mathbf{K}_1^{-T} \mathbf{s}$. Después, se sustituye \mathbf{s} en la ecuación (5.4) para obtener la forma final de la homografía a calcular por auto-calibración, mostrada a continuación:

$$\mathbf{H} = \begin{bmatrix} \mathbf{K}_1 & \mathbf{0} \\ -\mathbf{a}^T \mathbf{K}_1 & 1 \end{bmatrix}. \quad (5.6)$$

Para finalizar esta sección, obtendremos las ecuaciones básicas de auto-calibración. Definamos a las matrices de las cámaras de la reconstrucción proyectiva como $\mathbf{\Pi}_{P_i} = [\mathbf{B}_i | \mathbf{b}_i]$. Sustituyendo lo anterior junto con las ecuaciones (5.2) y (5.6) en la ecuación (5.3), obtenemos que:

$$\mathbf{K}_i \mathbf{R}_i = (\mathbf{B}_i - \mathbf{b}_i \mathbf{a}^T) \mathbf{K}_1, \quad \text{para } i = 2, \dots, m, \quad (5.7)$$

lo cual, despejando \mathbf{R}_i y eliminando esta matriz de la expresión al multiplicarla por \mathbf{R}_i^T , nos queda:

$$\begin{aligned} \Omega_i^* &= (\mathbf{B}_i - \mathbf{b}_i \mathbf{a}^T) \Omega_1^* (\mathbf{B}_i - \mathbf{b}_i \mathbf{a}^T)^T, \\ \Omega_i &= (\mathbf{B}_i - \mathbf{b}_i \mathbf{a}^T)^{-T} \Omega_1 (\mathbf{B}_i - \mathbf{b}_i \mathbf{a}^T)^{-1}, \end{aligned} \quad (5.8)$$

donde $\mathbf{\Omega}_i^* = \mathbf{K}_i \mathbf{K}_i^T$ y $\mathbf{\Omega}_i = (\mathbf{\Omega}_i^*)^{-1}$, matrices a las cuales se les conoce como imagen dual de la cónica absoluta e imagen de la cónica absoluta, respectivamente. A las ecuaciones (5.8) se denominan las ecuaciones básicas de auto-calibración.

Los diferentes métodos de auto-calibración son variaciones en la resolución de estas ecuaciones, y en generalmente calculan $\mathbf{\Omega}_i$ y $\mathbf{\Omega}_i^*$ para extraer de éstas las matrices de calibración de todas las cámaras. Si se parametrizan las matrices \mathbf{K}_i , es posible usar métodos iterativos de ajuste en bloque, para estimarlas directamente.

En las secciones siguientes presentamos breves descripciones de dos métodos de auto-calibración, los cuales fueron comparados en sus resultados para determinar cuál utilizar en la aplicación específica de este trabajo de tesis.

5.3. Las ecuaciones de Kruppa

A las ecuaciones de Kruppa se les considera históricamente como el primer método de auto-calibración [Hartley y Zisserman, 2004] y fue presentado en [Faugeras *et al.*, 1992]. A *grosso* modo, consisten en restricciones entre dos vistas de la misma escena y se necesita la matriz fundamental \mathbf{F} , que relaciona las dos imágenes, para resolverlas.

La obtención de las ecuaciones se basa en la correspondencia de líneas epipolares tangentes a una cónica cualquiera, cuya curva se encuentra localizada en el plano al infinito, proyectada en ambas vistas. De manera más concreta, supongamos que existe una cónica en el mundo \mathbf{C}_w que se proyecta en dos cámaras formando sus imágenes \mathbf{C} y \mathbf{C}' , con sus respectivas cónicas duales \mathbf{C}^* y $\mathbf{C}^{*'}$. Las líneas epipolares tangentes las denominamos \mathbf{l}_1 y \mathbf{l}_2 para la primera cámara, y \mathbf{l}'_1 y \mathbf{l}'_2 para la segunda cámara.

Realizando el producto simétrico entre las líneas epipolares para cada imagen, obtenemos las cónicas degeneradas presentadas en la ecuación (5.9).

$$\begin{aligned} \mathbf{C}_d &= \hat{\mathbf{e}} \mathbf{C}^* \hat{\mathbf{e}}, \\ \mathbf{C}'_d &= \hat{\mathbf{e}}' \mathbf{C}^{*'} \hat{\mathbf{e}}', \end{aligned} \tag{5.9}$$

donde, $\hat{\mathbf{e}}$ y $\hat{\mathbf{e}}'$ son los epipolos en matrices anti-simétricas de la primera y segunda imagen, respectivamente. Ahora bien, tenemos que establecer la relación de correspondencia entre las líneas epipolares y esto se hace por medio la homografía (\mathbf{H}), inducida por un plano

cualquiera en el espacio, que mapea los puntos en la imagen 1 a la imagen 2. Dicha relación se escribe:

$$\widehat{\mathbf{e}}' \mathbf{C}^* \widehat{\mathbf{e}}' = \mathbf{H}^{-T} \widehat{\mathbf{e}} \mathbf{C}^* \widehat{\mathbf{e}} \mathbf{H}^{-1} = \mathbf{F} \mathbf{C}^* \mathbf{F}^T, \quad (5.10)$$

en la cual se substituyó $\mathbf{F} = \mathbf{H}^{-T} \widehat{\mathbf{e}}$, pues es claro que el mapeo de un punto en una imagen a su respectiva línea epipolar (intersección del plano de la imagen y el plano definido por los centros de las dos cámaras y el punto de la escena) en la otra imagen, es la tarea que define a la matriz fundamental \mathbf{F} , definida en la sección 6.1.

Como se ha comentado anteriormente, en auto-calibración la cónica de interés es la absoluta, por lo tanto, la homografía inducida respectiva es la homografía al infinito, ya que la cónica absoluta se encuentra en el plano al infinito. Tomando estas consideraciones y recordando que estamos trabajando con las imágenes de las cónicas, reescribimos la ecuación (5.10) obteniendo:

$$\widehat{\mathbf{e}}' \boldsymbol{\Omega}^* \widehat{\mathbf{e}}' = \mathbf{F} \boldsymbol{\Omega}^* \mathbf{F}^T. \quad (5.11)$$

Finalmente, si consideramos que los parámetros intrínsecos son fijos para ambas vistas, obtenemos

$$\widehat{\mathbf{e}}' \boldsymbol{\Omega}^* \widehat{\mathbf{e}}' = \mathbf{F} \boldsymbol{\Omega}^* \mathbf{F}^T, \quad (5.12)$$

a las que son conocidas como las ecuaciones de Kruppa en su forma original. La ecuación (5.11) al ser complicada de aplicar, se desarrolló su forma equivalente presentada en la ecuación (5.13), y consiste en tres ecuaciones cuadráticas en los elemento de $\boldsymbol{\Omega}^*$, pero sólo dos de ellas son independientes.

$$\begin{bmatrix} \mathbf{c}_2^T \boldsymbol{\Omega}^* \mathbf{c}_2 \\ -\mathbf{c}_1^T \boldsymbol{\Omega}^* \mathbf{c}_2 \\ \mathbf{c}_1^T \boldsymbol{\Omega}^* \mathbf{c}_1 \end{bmatrix} \times \begin{bmatrix} \sigma_1^2 \mathbf{d}_1^T \boldsymbol{\Omega}^* \mathbf{d}_1 \\ \sigma_1 \sigma_2 \mathbf{d}_1^T \boldsymbol{\Omega}^* \mathbf{d}_2 \\ \sigma_2^2 \mathbf{d}_2^T \boldsymbol{\Omega}^* \mathbf{d}_2 \end{bmatrix} = \mathbf{0}, \quad (5.13)$$

donde, \mathbf{c}_i , \mathbf{d}_i y σ_i son los vectores columna de las matrices ortogonales y los valores singulares de la SVD.

Una forma de solucionar las ecuaciones de Kruppa es mediante su simplificación. Para ésto, suponemos que conocemos el punto principal y la razón de aspecto, y que la asimetría (*skew*) es igual a cero para ambas cámaras. Las distancias focales son consideradas desconocidas y diferentes en las dos cámaras. Si realizamos un cambio de coordenadas en el punto principal (que después será restaurado) obtenemos:

$$\boldsymbol{\Omega}^* = \text{diag}(f_u^2, f_u^2, 1), \quad \boldsymbol{\Omega}^{*'} = \text{diag}(f_u'^2, f_u'^2, 1) \quad (5.14)$$

donde f_u y f'_u son las distancias focales desconocidas en unidades de píxeles para la primera y segunda cámara, respectivamente. Como los vectores de la ecuación (5.12) son colineales, la razón entre sus componentes es igual, con lo que obtenemos la siguiente ecuación,

$$\frac{\mathbf{c}_2^T \boldsymbol{\Omega}^* \mathbf{c}_2}{\sigma_1^2 \mathbf{d}_1^T \boldsymbol{\Omega}^* \mathbf{d}_1} = -\frac{-\mathbf{c}_1^T \boldsymbol{\Omega}^* \mathbf{c}_2}{\sigma_1 \sigma_2 \mathbf{d}_1^T \boldsymbol{\Omega}^* \mathbf{d}_2} = \frac{\mathbf{c}_1^T \boldsymbol{\Omega}^* \mathbf{c}_1}{\sigma_2^2 \mathbf{d}_2^T \boldsymbol{\Omega}^* \mathbf{d}_2}. \quad (5.15)$$

Finalmente, multiplicando de forma cruzada en la ecuación anterior obtenemos dos ecuaciones cuadráticas en f_u y f'_u que pueden resolverse fácilmente. Es importante mencionar que las ecuaciones de Kruppa presentan ambigüedades en su solución cuando entre las vistas no existe rotación alguna. Ésto se debe a que ninguna restricción se establece en $\boldsymbol{\Omega}^*$ y, por lo tanto, la ecuación (5.12) se puede reescribir como la ecuación (5.16).

$$\widehat{\mathbf{e}}' \boldsymbol{\Omega}^* \widehat{\mathbf{e}}' = \widehat{\mathbf{e}}' \boldsymbol{\Omega}^* \widehat{\mathbf{e}}' \quad (5.16)$$

Ya que la matriz fundamental se reduce a $\mathbf{F} = \widehat{\mathbf{e}}'$ al obtener las mismas líneas epipolares en ambas cámaras, ocasionando que la homografía al infinito se convierta en la matriz identidad. Es por eso que al momento de decidir el método para auto-calibrar el sistema *cámara-ojo* tenemos que probar a las ecuaciones de Kruppa en desempeño bajo estas condiciones entre vistas, como se verá en la sección 5.5. Para mayor detalle de las condiciones impuestas para la simplificación de las ecuaciones de Kruppa refiérase a [Lourakis *et al.*, 1999].

5.4. Auto-calibración a partir de la matriz esencial

El segundo método de auto-calibración analizado en esta tesis es un mejoramiento al algoritmo propuesto en [Hartley, 1992], el cual está basado en propiedades de la matriz esencial demostradas en [Huang y Faugeras, 1989]. Al descomponer por valores singulares a la matriz esencial, obtenemos en su matriz diagonal los primeros dos valores singulares iguales y el tercero es cero (en orden descendente), debido al hecho de que dicha matriz es de rango dos. La igualdad en los valores singulares nos genera dos restricciones en las entradas de la matriz esencial, siendo estas condiciones la base del algoritmo de Hartley para auto-calibrar cámaras. La desventaja de este algoritmo es que únicamente se pueden estimar cualquiera dos de los parámetros de la cámara y que es muy poco robusto e inestable [Mendonça y Cipolla, 1999].

La mejora al algoritmo de Hartley propuesta en [Mendonça y Cipolla, 1999], consiste en utilizar un número mayor de imágenes de la misma escena, para calcular las matrices fundamentales entre pares de ellas. Lo anterior nos permite estimar un mayor número de parámetros internos, por medio de un método numérico de minimización sobre las restricciones

parametrizadas en los parámetros intrínsecos de las cámaras. Así obtenemos de manera directa dichos parámetros en vez de encontrar primero una cónica y después tener que extraer los parámetros de esta última, como en las ecuaciones de Kruppa. A continuación presentamos la descripción del algoritmo.

Para determinar el número mínimo de imágenes para poder realizar el proceso de auto-calibración seguimos la ecuación (5.17), a lo que se conoce como argumento de conteo (*counting argument*).

$$nc + (n - 1)d \geq 8, \quad (5.17)$$

donde, n es el número de imágenes. De todos los parámetros intrínsecos c de ellos son conocidos en todas las vistas y d de ellos son fijos en todas las vistas pero son desconocidos, por lo tanto, $c + d \geq 5$. Necesitamos 8 restricciones como mínimo, porque al auto-calibrar lo que buscamos es restringir la transformación proyectiva (15 grados de libertad) a una métrica (3 grados de libertad de la rotación, 3 de la traslación y 1 del factor de escala, en total 7), es decir, necesitamos fijar $15 - 7 = 8$ entradas por medio de restricciones.

Contando con n imágenes de la escena y con los punto en correspondencia entre pares de ellas, se pueden calcular $n(n - 1)/2$ matrices fundamentales \mathbf{F}_{ij} , donde ij son los índices de las cámaras (imágenes) que fueron utilizadas para calcular dicha matriz fundamental. Como siempre mientras mayor sea el número de matrices \mathbf{F}_{ij} , mayor será la robustez y estabilidad del algoritmo por estar sobre-determinado. Si pensamos que cada una de las imágenes fué tomada con una cámara diferente, como ya habíamos mencionado antes en esta tesis, tendremos dos matrices de calibración \mathbf{K}_i y \mathbf{K}_j (junto con la fundamental) involucradas en la matriz esencial \mathbf{E}_{ij} (véase ecuación 6.2) para cada configuración de geometría epipolar, como se muestra en la siguiente ecuación:

$$\mathbf{E}_{ij} = \mathbf{K}_i \mathbf{F}_{ij} \mathbf{K}_j. \quad (5.18)$$

Las matrices de parámetros intrínsecos a estimar son impuestas arbitrariamente a modo de cámaras calibradas débilmente, y se definen como,

$$\mathbf{K} = \begin{bmatrix} \epsilon f_x & \gamma & u_{o1} \\ 0 & f_x & u_{o2} \\ 0 & 0 & 1 \end{bmatrix}, \quad (5.19)$$

donde, $f_x = \frac{f}{s_x}$ es la distancia focal en unidades de píxeles, calculado con la dimensión horizontal del píxel (s_x); y los píxeles se consideran cuadrados. La notación es la misma usada

en la ecuación 3.3.

Posteriormente, se calculan los valores singulares de \mathbf{E}_{ij} definida en la ecuación (5.18), a los que denominaremos σ_{ij}^1 y σ_{ij}^2 , donde los superíndices indican su posición en orden descendente. Haciendo uso de la propiedad de los valores de las matrices esenciales, obtenemos una función de costo a ser minimizada para obtener cada una de las matrices de calibración. Básicamente el error cuantificado por la función de costo es la diferencia normalizada entre los dos valores singulares, ya que teóricamente deben ser iguales. La función de costo se presenta en la ecuación (5.20).

$$C(\mathbf{K}_i, i = 1, \dots, n) = \sum_{ij}^n \frac{w_{ij}}{\sum_{kl}^n w_{kl}} \frac{\sigma_{ij}^1 - \sigma_{ij}^2}{\sigma_{ij}^2}, \quad (5.20)$$

donde, w_{ij} son los pesos de la ponderación que representan la confiabilidad en la matriz esencial, por ejemplo, pueden ser inversamente proporcionales al número de pares de puntos en correspondencia empleados para estimar la matriz fundamental, en base a la cual se calcula la matriz esencial.

El promedio realizado se pondera para darle un carácter simétrico a la función de costo, tal como lo mencionan los autores de este método. Pues se evita el sesgo por una determinada imagen que contenga mayor cantidad de ruido, cuente con correspondencias erróneas (*outliers*), en fin, que genere matrices fundamentales en cierto grado "corruptas". Es por ello que los pesos pueden ser los números de puntos en correspondencia utilizados para estimar la determinada matriz \mathbf{F}_{ij} , o bien, de un modo más sofisticado, el inverso de la distancia media de un punto en la imagen a su respectiva línea epipolar.

Si analizamos detenidamente la ecuación (5.20), podemos ver que la estimación de una matriz de calibración por auto-calibración utiliza a las matrices de calibración de las otras cámaras (o vistas) para calcular la matriz esencial. Esto significa que la auto-calibración se hace en base a cámaras débilmente calibradas que permanecen sin alteraciones durante el proceso de minimización. Es por eso, que decidimos implementar el método realizando actualizaciones globales de las matrices de calibración para que al detenerse el proceso de minimización, obtengamos matrices \mathbf{K}_i estimadas en base a matrices que tras cada iteración minimizaron su función de costo. El proceso de minimización no-lineal se realizó, en este trabajo de tesis, con el método del descenso más rápido (*steepest descent method*).

En resumen, este método de auto-calibración está basado en un conjunto de refinamiento

de matrices esenciales, para así obtener por minimización los parámetros intrínsecos que inicialmente fueron introducidos como supuestos. Por lo tanto, es importante evaluar el desempeño de este método con parámetros intrínsecos iniciales alejados de los reales pues en realidad estos son totalmente desconocidos, al menos en el caso de la distancia focal. Esta evaluación se lleva a cabo en la sección 5.5, presentada a continuación.

5.5. Comparación entre los métodos

Para decidir cuál es el método de auto-calibración más adecuado para nuestra aplicación, decidimos realizar un análisis comparativo de los resultados obtenidos en reconstrucción por geometría epipolar. La escena a reconstruir está constituida por dos rejillas rectangulares de puntos con geometría conocida y posicionados sobre planos ortogonales, muy similar a un patrón de calibración clásico, que se muestra en la Figura 5.1. Las matrices de parámetros intrínsecos empleadas fueron obtenidas auto-calibrando las cámaras por los dos métodos vistos anteriormente en las secciones 5.3 y 5.4. El objetivo de utilizar una escena artificial es brindarnos la posibilidad de calificar cuantitativamente el resultado de la reconstrucción con las diferentes matrices de calibración.

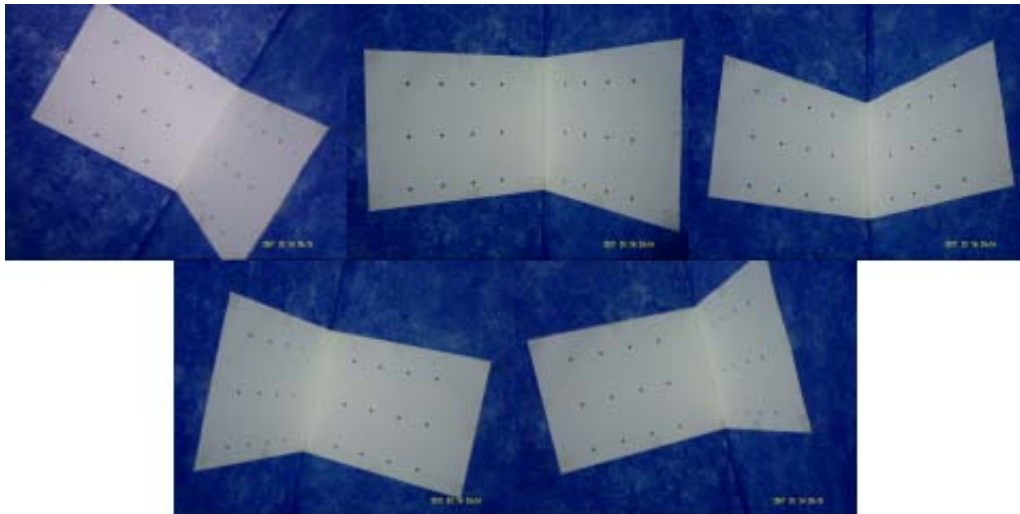


Figura 5.1: Conjunto de imágenes del patrón conocido capturado en diferentes puntos de vista que involucran rotaciones y traslaciones.

Dos secuencias diferentes de imágenes se usaron para auto-calibrar las cámaras. En la

primera, los cambios en la posición de la cámara consistieron en movimientos generales de rotación y traslación, y en la segunda secuencia, la cámara únicamente fue trasladada, es decir, no fue rotada en ninguno de sus ejes. A todas las imágenes se les corrigió la distorsión mediante el sistema *ADICOS*. Los dos conjuntos de imágenes se pueden observar en las Figuras 5.1 y 5.2, respectivamente.



Figura 5.2: Conjunto de imágenes del patrón conocido en puntos de vista cambiados por movimiento únicamente de traslación.

Con movimientos generales en la pose de la cámara proveemos a las matrices fundamentales calculadas con la mayor información posible de la perspectiva de la escena, lo que es deseable para obtener una reconstrucción métrica. Por lo tanto, el motivo para realizar reconstrucción con el segundo conjunto de imágenes es aplicar las ecuaciones de Kruppa en las circunstancias en las cuales se espera que presente ambigüedades en su solución y apreciar la calidad en la reconstrucción. De igual modo aplicamos el método de auto-calibración a partir de la matriz esencial para saber si presenta ventajas en secuencias con movimientos únicamente de traslación.

Otro parámetro a considerar en estas pruebas, es la inicialización con la cual operan los métodos de auto-calibración. Para cada método los autores proponen una matriz de parámetros internos para iniciar los procesos de minimización. Es por eso que decidimos realizar reconstrucciones intercambiando estos parámetros y, además tomar como inicializaciones las matrices de calibración obtenidas por el sistema descrito en la sección 3.5.

Las medidas geométricas de la escena artificial tomadas como parámetros de comparación en las reconstrucciones, cuyo esquema se puede observar en la Figura 5.3, fueron:

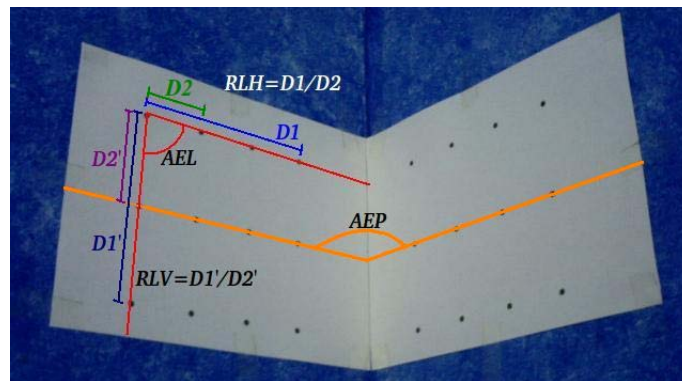


Figura 5.3: Parámetros de comparación elegidos en la rejilla de puntos de la escena artificial.

1. El ángulo entre los dos planos, originalmente ortogonales, donde se sitúan los puntos de las rejillas (*AEP*).
2. El ángulo entre las líneas que pasan por los puntos de la primera fila y la primera columna de la rejillas (*AEL*).
3. La razón de longitud entre el segmento de línea que va del primer al cuarto punto (*D1*) y el segmento que va del primer al segundo punto (*D2*), ambos tomados de la primera fila de puntos (*RLH*).
4. La razón de longitud entre el segmento de línea que va del primer punto de la primera fila al primer punto de la cuarta fila (*D1'*), y el segmento que va del primer punto de la primera fila al primer punto de la segunda fila (*D2'*), es decir, ambas medidas tomadas de la primera columna (*RLV*).

A continuación se presenta el análisis comparativo de los parámetros (*AEP*, *AEL*, *RLH* y *RLV*) cuyos valores fueron medidos directamente de la escena real y de las rejillas de puntos reconstruidos. Adicionalmente se presentan los correspondientes errores relativos (el cociente de la diferencia absoluta del valor medido y el valor exacto, entre el valor exacto) en porcentajes. Empezaremos con las reconstrucciones provenientes de las imágenes tomadas con cambios en la pose de la cámara por movimientos de traslación y rotación. La tabla 5.1

se muestran los errores relativos en la reconstrucción usando como inicialización los parámetros del método de *Kruppa*. En la tabla 5.2 se presentan los valores de los parámetros de comparación inicializando las auto-calibraciones con la matriz \mathbf{K} propuesta para el método Esencial. Usando como inicialización los parámetros internos calculados por *calib-tool* (véase sección 3.5), se obtienen los resultados presentados en la tabla 5.3.

	Valores en reconstrucción		Valores reales	Error relativo (%)	
	Kruppa	Esencial		Kruppa	Esencial
AEP	90.2379°	63.7264°	90°	0.2643	29.1929
AEL	90.4562°	100.4657°	90°	0.5069	11.6286
RLH	2.9512	3.3507	2.9592	0.2720	13.2284
RLV	2.0150	1.9201	2	0.7498	3.9962

Tabla 5.1: Análisis comparativo de las reconstrucciones usando auto-calibración con cambios en la pose de la cámara mediante rotaciones y traslaciones, inicializando con los parámetros genéricos del método de *Kruppa*.

	Valores en reconstrucción		Valores reales	Error relativo (%)	
	Kruppa	Esencial		Kruppa	Esencial
AEP	90.0171°	74.8958°	90°	0.0189	16.7825
AEL	93.0996°	87.3807°	90°	3.4440	2.9103
RLH	2.9389	3.0263	2.9592	0.6850	2.2664
RLV	2.0053	2.0211	2	0.2656	1.0539

Tabla 5.2: Análisis comparativo de las reconstrucciones usando auto-calibración con cambios en la pose de la cámara mediante rotaciones y traslaciones, inicializando con los parámetros genéricos del método *Esencial*.

Los puntos reconstruidos usando los parámetros internos obtenidos en cada una de las pruebas anteriores, es decir, usando cambios en la pose de la cámara por movimientos generales y empleando los tres diferentes conjuntos de parámetros iniciales, se presentan en las Figuras 5.4(a)-(c), respectivamente.

Con estos tres experimentos podemos determinar que el método de *Kruppa* presenta un mejor desempeño porque sus errores relativos son más bajos en los primeros dos casos. En el tercer caso, los errores relativos del método *Esencial* son menores, sin presentar una ventaja importante con respecto al de *Kruppa*. Esto se debe a que el ajuste del método *Esencial* es

	Valores en reconstrucción		Valores reales	Error relativo (%)	
	Kruppa	Esencial		Kruppa	Esencial
AEP	94.5118°	89.9470°	90°	5.0131	0.0589
AEL	93.0402°	87.6830°	90°	3.3780	2.5744
RLH	2.9356	2.9682	2.9592	0.7960	0.3026
RLV	2.0015	1.9861	2	0.0756	0.6932

Tabla 5.3: Análisis comparativo de las reconstrucciones usando auto-calibración con cambios en la pose de la cámara mediante rotaciones y traslaciones, inicializando con los parámetros precalculados por *calib-tool* [Bouguet, 2008].

minúsculo pues estamos partiendo de una matriz de calibración que cumple la propiedad de igualdad en sus valores singulares. Ahora pasamos a las pruebas donde el cambio en la pose de la cámara es únicamente traslacional. En las tablas 5.4, 5.5 y 5.6 se muestran los resultados las reconstrucciones inicializando los procesos de auto-calibración con los parámetros genéricos de Kruppa, de Esencial y los precalculados por *Calib-tool*, respectivamente.

	Valores en reconstrucción		Valores reales	Error relativo (%)	
	Kruppa	Esencial		Kruppa	Esencial
AEP	75.5635°	121.7406°	90°	16.0406	35.2674
AEL	88.4333°	95.9036°	90°	1.7408	6.5595
RLH	3.2270	3.1079	2.9592	9.0513	5.0236
RLV	2.0351	2.0239	2	1.7531	1.1942

Tabla 5.4: Análisis comparativo de las reconstrucciones usando auto-calibración con cambios en la pose de la cámara únicamente por traslaciones, inicializando con los parámetros genéricos del método de *Kruppa*.

Las reconstrucciones de las cuales fueron tomadas las medidas presentadas en las tablas anteriores, se muestran en las Figuras 5.5(a)-(c), según los parámetros de inicialización usados en la etapa de auto-calibración.

En estos tres experimentos ambos métodos presentan errores relativos importantes, sobre todo en el ángulo entre los planos ortogonales (AEP). Sin embargo podemos observar que nuevamente el método de Kruppa nos permite recuperar con mayor precisión la geometría de la escena. Ya que sus errores relativos son menores que usando el método Esencial, excepto en el tercer caso. Esto se debe a que la calibración previa mediante *Calib-tool* usa imágenes

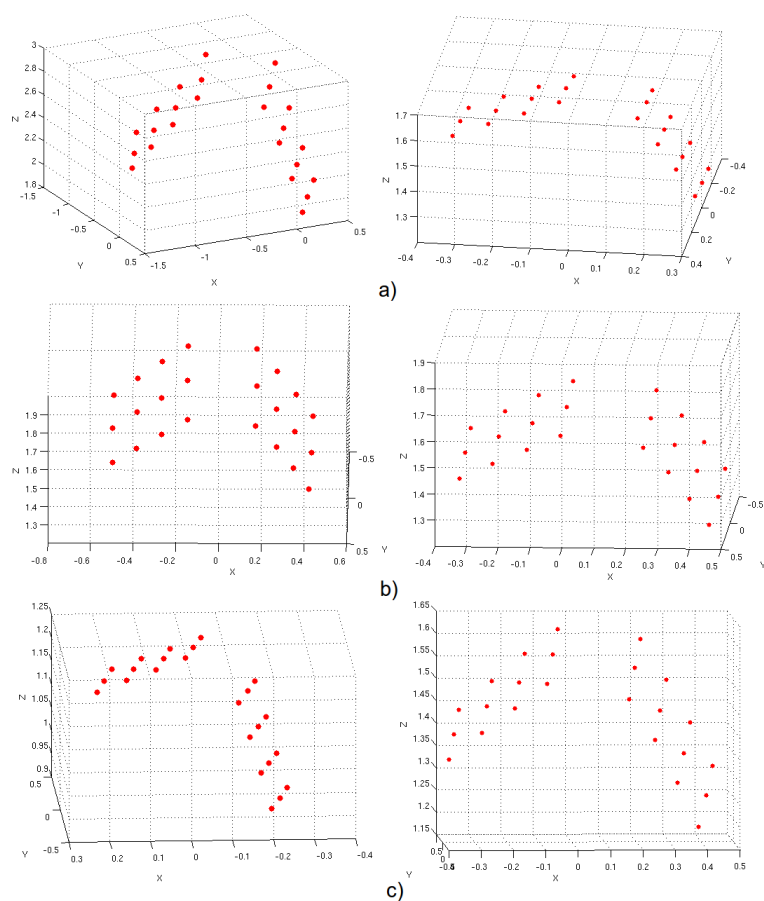


Figura 5.4: Reconstrucciones de la escena controlada de puntos localizados en planos perpendiculares. A la izquierda de cada pareja se usó el método de la matriz esencial y a la derecha el de Kruppa. Con respecto a la pose cambiada por traslación y rotación tenemos en (a) la inicialización con parámetros genéricos de Kruppa, en (b) inicialización con los parámetros genéricos de esencial y en (c) la inicialización por parámetros calculados por *Calib-tool* [Bouquet, 2008].

en diversas poses de la cámara, cambiadas por movimientos generales. De nuevo el ajuste realizado por el método Esencial a una matriz fundamental es mínimo y los errores relativos aunque son menores, son aproximados a los obtenidos con Kruppa.

En base a los datos presentados en las tablas es fácil darse cuenta de que la calidad de las reconstrucciones es mucho mejor cuando el cambio en la pose de las cámaras cambia

	Valores en reconstrucción		Valores reales	Error relativo (%)	
	Kruppa	Esencial		Kruppa	Esencial
AEP	75.3747°	127.6826°	90°	16.2503	41.8696
AEL	88.6481°	95.3754°	90°	1.5021	5.9727
RLH	3.2248	3.2129	2.9592	8.9738	8.5739
RLV	2.0377	2.0154	2	1.8851	0.7685

Tabla 5.5: Análisis comparativo de las reconstrucciones usando auto-calibración con cambios en la pose de la cámara únicamente por traslaciones, inicializando con los parámetros genéricos del método *Esencial*.

	Valores en reconstrucción		Valores reales	Error relativo (%)	
	Kruppa	Esencial		Kruppa	Esencial
AEP	75.3887°	103.5630°	90°	16.2348	15.0700
AEL	86.9372°	92.9068°	90°	3.4031	3.2298
RLH	3.2308	3.1619	2.9592	9.1798	6.8510
RLV	2.0299	2.0183	2	1.4975	0.9165

Tabla 5.6: Análisis comparativo de las reconstrucciones usando auto-calibración con cambios en la pose de la cámara únicamente por traslaciones, inicializando con los parámetros precalculados por *Calib-tool*.

por traslación y rotación, a diferencia de cambiar por un movimiento únicamente traslacional. Esto se debe a que las matrices fundamentales caen de rango al no tener la suficiente información provista por los pares de imágenes y por falta de independencia lineal en las restricciones del sistema sobre determinado con el cual son calculadas, es decir, las diferentes vistas no contienen la suficiente información de perspectiva.

El mejor desempeño fue presentado por las ecuaciones de Kruppa cuando se usan los parámetros genéricos propuestos en [Lourakis *et al.*, 1999], salvo en las pruebas realizadas con traslación únicamente. La razón de esto se describe en [Hartley y Zisserman, 2004], la cual se fundamenta en el hecho de que la matriz fundamental al provenir de la homografía al infinito, la podemos representar en términos de los epipolos y líneas epipolares. Estas últimas al haber un cambio únicamente por traslación en una de las vistas de la geometría epipolar, ocasiona que la matriz fundamental obtenida sea igual a la forma anti-simétrica del epipolo, lo cual genera una ambigüedad en la solución. Lo anterior se muestra en la ecuación (5.16).

Después del análisis anterior decidimos emplear el método de las ecuaciones de Kruppa

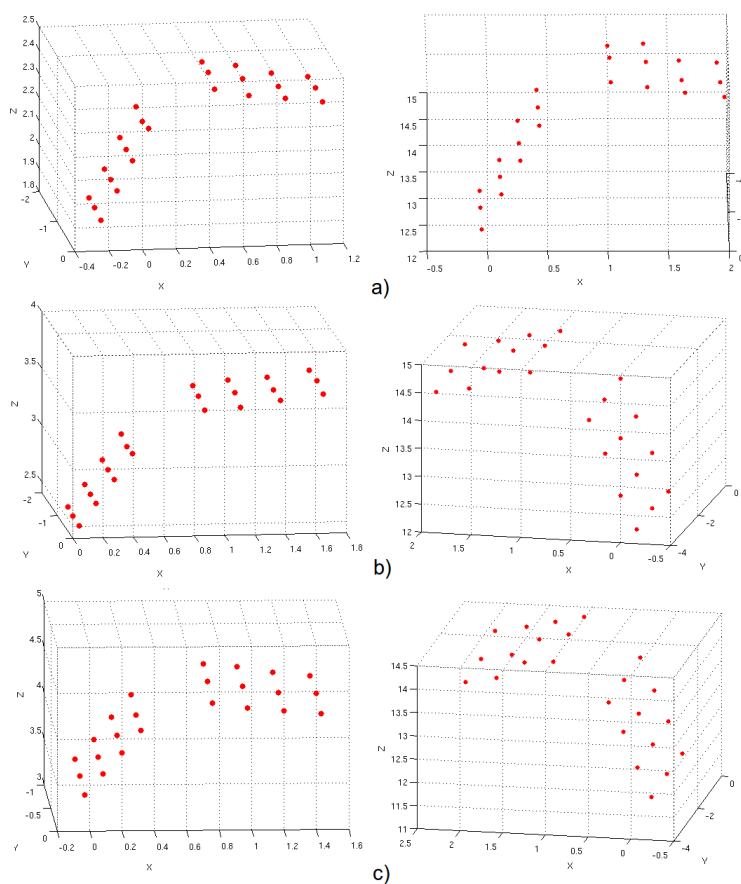


Figura 5.5: Reconstrucciones de la escena controlada de puntos localizados en planos perpendiculares. A la izquierda de cada pareja se usó el método de la matriz esencial y a la derecha el de Kruppa. Con respecto a la pose cambiada únicamente por traslación tenemos en (a) la inicialización con parámetros genéricos de Kruppa, en (b) inicialización con los parámetros genéricos de esencial y en (c) la inicialización por parámetros calculados por *Calib-tool* [Bouguet, 2008].

para auto-calibrar el sistema *cámara-ojo*. Bajo la consideración de que las imágenes usadas para auto-calibrar y reconstruir contarán en sus cambios de pose con rotaciones y traslaciones, y que los parámetros de inicialización sean los propuestos en [Lourakis *et al.*, 1999] para una cámara genérica. Otro justificante para esta decisión fue que en las pruebas realizadas con la segunda secuencia de imágenes, usando el método de matriz esencial hubieron casos en donde los puntos reconstruidos resultaron colapsados y sin ningún parecido a la escena

real (véase Figura 5.6), a diferencia que con Kruppa nunca ocurrió lo anterior.

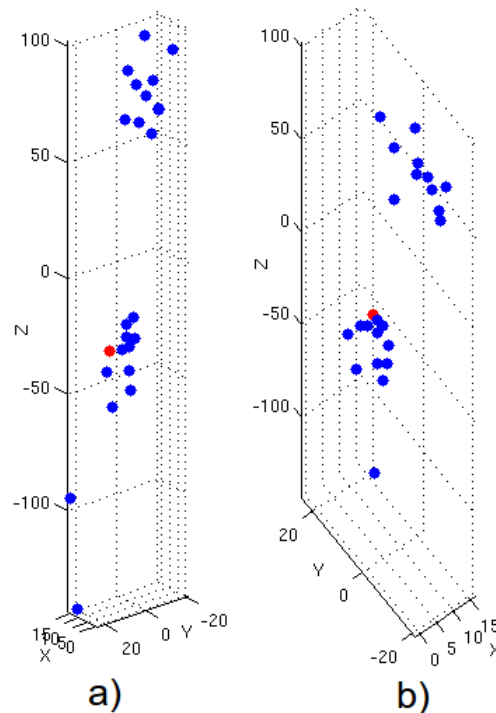


Figura 5.6: (a) y (b) son dos vistas de la reconstrucción del patrón de puntos en planos perpendiculares por el método Esencial, usando las imágenes donde la pose de la cámara cambia por traslación. Los puntos aparecen colapsados y sin similitud alguna con la escena real.

En concreto, consideramos que las mejores reconstrucciones (en la secuencia sin rotaciones) por ambos métodos resultó ser mejor el método de matriz esencial pero lo que nos lleva a desistir de usar este método es que en las peores reconstrucciones resultó ser mucho mejor el método de las ecuaciones Kruppa, esto nos indica que en una escena no controlada y compleja como la retina es muy riesgoso emplear el método de la matriz esencial ya que la geometría de las estructuras no es intuitiva.

Una vez obtenida la matriz de parámetros internos, por calibración clásica o por auto-calibración, sólo nos resta encontrar la posición de los puntos característicos de las imágenes, en el espacio 3D bajo el sistema de referencia del mundo. Por lo tanto, es importante entender el principio de funcionamiento de este método de reconstrucción conocido como

geometría epipolar y la herramienta principal para calcular las coordenadas 3D de los puntos a reconstruir, conocida como *triangulación lineal*. En el siguiente capítulo se explican los fundamentos de los métodos antes mencionados, así como también el algoritmo utilizado para la representación de volúmenes.

Capítulo 6

Procesos posteriores a la calibración

Este capítulo lo empezaremos introduciendo el método de reconstrucción usado en este trabajo de tesis, llamado geometría epipolar, el cual fue descrito someramente en la sección 1.2. Posteriormente, presentaremos breves descripciones de los procesos utilizados en la última etapa de la reconstrucción, para obtener finalmente la visualización de las estructuras retinales en un espacio virtual de 3D. Dichos métodos son la triangulación lineal y los cilindros generalizados, los resultados de estas etapas se muestran en las secciones 7.2-7.5.

6.1. Geometría epipolar

Recordemos que el término de geometría epipolar se refiere a la estructura geométrica formada por dos vistas de una misma escena, es decir, contamos con dos imágenes de la misma escena tomadas desde puntos de vista distintos. La Figura 6.1 muestra esta configuración de cámara, donde \mathbf{x}_w es un punto en la escena observado por ambas cámaras, \mathbf{x}_o y \mathbf{x}'_o son los centros de la primera y segunda cámara, \mathbf{e} y \mathbf{e}' son los epipolos de la primera y segunda imagen.

Las coordenadas de dos puntos en correspondencia \mathbf{x}_c y \mathbf{x}'_c , cuyos sistemas de referencia son los de sus respectivas cámaras, se encuentran ligadas por una relación geométrica muy precisa, conocida como restricción epipolar, la cual se enuncia en el Teorema 6.1.

Teorema 6.1 *Considere dos puntos imagen $\mathbf{x}_c, \mathbf{x}'_c \in \mathbb{R}^3$ correspondientes al mismo punto \mathbf{x}_w en la escena capturados en dos posiciones de cámara con pose relativa $(\mathbf{R}$ y \mathbf{t}), donde $\mathbf{R} \in SO(3)$ es la orientación relativa y $\mathbf{t} \in \mathbb{R}^3$ es la posición relativa [Ma et al., 2005]. Entonces $\mathbf{x}_c, \mathbf{x}'_c$ satisfacen:*

$$\mathbf{x}'_c{}^T \widehat{\mathbf{t}} \mathbf{R} \mathbf{x}_c = 0. \quad (6.1)$$

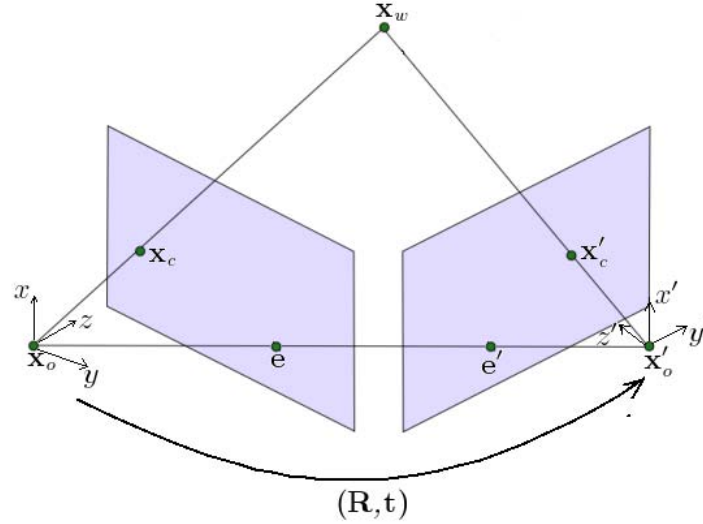


Figura 6.1: Configuración de dos vistas. Geometría epipolar [Ma *et al.*, 2005].

Del Teorema 6.1, es importante aclarar que la matriz $\hat{\mathbf{t}}$ es la versión anti-simétrica del vector \mathbf{t} , y que $SO(3)$ es el subgrupo ortogonal especial de transformaciones lineales de 3×3 . El adjetivo especial indica que el determinante de la transformación es igual a $+1$.

La matriz

$$\mathbf{E} \doteq \hat{\mathbf{t}}\mathbf{R} \in \mathbb{R}^{3 \times 3}, \quad (6.2)$$

de la ecuación (6.1) que expresa la restricción epipolar y se le conoce como matriz esencial. Dentro de la matriz esencial se encuentra codificado la pose relativa entre las dos cámaras. Como se mencionó anteriormente, las coordenadas de los puntos en correspondencia \mathbf{x}_c y \mathbf{x}'_c están dadas bajo el sistema de referencia de las cámaras, datos con los cuales no se cuenta cuando se extraen los puntos característicos directamente de las imágenes.

Es por eso que para describir la restricción epipolar, en los sistemas de referencia de las imágenes, las matrices de parámetros intrínsecos se aplican en la ecuación (6.1), como se muestra en la siguiente ecuación:

$$\mathbf{u}'^T \mathbf{K}_2^{-T} \hat{\mathbf{t}} \mathbf{R} \mathbf{K}_1^{-1} \mathbf{u} = 0, \quad (6.3)$$

donde, la matriz

$$\mathbf{F} = \mathbf{K}_2^{-T} \hat{\mathbf{t}} \mathbf{R} \mathbf{K}_1^{-1}, \quad (6.4)$$

se le conoce como matriz fundamental, la cual contiene internamente los parámetros intrínsecos y extrínsecos de las cámaras, \mathbf{u} y \mathbf{u}' son las coordenadas de los puntos en correspondencia bajo el sistema de referencia de la imagen, y \mathbf{K}_1 y \mathbf{K}_2 son las matrices de calibración de las cámaras. De la ecuación (6.4), se puede deducir fácilmente la ecuación (5.18).

El algoritmo de los 8 puntos normalizados [Hartley, 1997], estima la matriz fundamental de un conjunto de 8 o más pares de puntos en correspondencia con un sistema sobredeterminado de ecuaciones. Mediante la ecuación (5.18) obtenemos la matriz esencial. Posteriormente se extraen los parámetros extrínsecos de la segunda cámara respecto a la primera mediante la factorización de \mathbf{E} , descomponiéndola por SVD y eligiendo una de las cuatro soluciones posibles (debido a las ambigüedades proyectivas) según la interpretación geométrica de la reconstrucción. Para mayor detalle de lo anterior consultar [Hartley y Zisserman, 2004].

Finalmente, con los parámetros internos y externos de ambas cámaras, podemos construir las matrices $\mathbf{\Pi}_i$ (véase ecuación (3.5)), las cuales son utilizadas por el método de triangulación lineal para obtener la reconstrucción. Este último método se describe a continuación.

6.2. Triangulación lineal

Este algoritmo supone que ya contamos con las matrices de proyección en perspectiva de ambas cámaras ($\mathbf{\Pi}$ y $\mathbf{\Pi}'$). Lo que se pretende encontrar es un punto $\mathbf{x}_w = (x_{w1}, x_{w2}, x_{w3}, 1)^T$ en la escena 3D con respecto al sistema de coordenadas del mundo teniendo como datos de entrada sus proyecciones en el plano de las imágenes, bajo el sistema de coordenadas de la imagen, $\mathbf{u} = (u_1, u_2, 1)^T$ y $\mathbf{u}' = (u'_1, u'_2, 1)^T$ en ambas imágenes. Ahora bien, por definición de la matriz de proyección tenemos que:

$$\mathbf{u} = \mathbf{\Pi}\mathbf{x}_w, \quad (6.5)$$

$$\mathbf{u}' = \mathbf{\Pi}'\mathbf{x}_w. \quad (6.6)$$

Ya que estamos trabajando con coordenadas homogéneas, las ecuaciones (6.5) y (6.6) se cumplen para cualquier factor de escala, por lo tanto, resulta más conveniente usar la expresión de colinealidad, como la siguiente:

$$\mathbf{u} \times \mathbf{\Pi}\mathbf{x}_w = 0, \quad (6.7)$$

$$\mathbf{u}' \times \mathbf{\Pi}'\mathbf{x}_w = 0. \quad (6.8)$$

Desarrollando las ecuaciones 6.7 y 6.8, despejando \mathbf{x}_w en cada una de ellas y eliminando la fila que resulte ser linealmente dependiente, obtenemos un sistema de cuatro ecuaciones

como el mostrado a continuación:

$$\begin{bmatrix} u_2 \mathbf{\Pi}_3^T - \mathbf{\Pi}_2^T \\ \mathbf{\Pi}_1^T - u_1 \mathbf{\Pi}_3^T \\ u'_2 \mathbf{\Pi}'_3{}^T - \mathbf{\Pi}'_2{}^T \\ \mathbf{\Pi}'_1{}^T - u'_1 \mathbf{\Pi}'_3{}^T \end{bmatrix} \mathbf{x}_w = 0, \quad (6.9)$$

donde, $\mathbf{\Pi}_j$ y $\mathbf{\Pi}'_j$ son los j -ésimos renglones de $\mathbf{\Pi}$ y $\mathbf{\Pi}'$. Como paso final resolvemos el sistema de ecuaciones para encontrar \mathbf{x}_w por SVD. De esta forma recuperamos la posición del punto en el espacio bajo el sistema de referencia del mundo a partir de las coordenadas en las que fue proyectado el punto en el plano de las imágenes.

6.3. Modelo de superficies

Como hemos visto hasta ahora, el proceso de reconstrucción tridimensional recupera la posición en el espacio de un punto proyectado en ambas imágenes. Por lo tanto, al estimar la posición 3D de todos los puntos que conforman a la escena, en este caso los vasos retinales, se obtiene la reconstrucción de los esqueletos.

Debido a que RISA, además de segmentar y calcular los puntos característicos, mide el grosor de los vasos sanguíneos, es posible generar una representación de las superficies mediante un modelo conocido como *cilindros generalizados*, cuyos radios serán determinados por las mediciones hechas por RISA. El método anterior fue empleado en trabajos previos [Martinez-Perez y Espinosa-Romero, 2004; Espinosa-Romero y Martinez-Perez, 2005], en los cuales esta basada esta tesis.

En las representaciones por cilindros generalizados, el volumen de los objetos se describe como un volumen barrido por un conjunto de superficies bidimensionales que se mueven a lo largo de la trayectoria descrita por una curva tridimensional arbitraria. Los vectores normales de la superficies son tangentes a dicha curva que representa el esqueleto del objeto. El conjunto de superficies pueden cambiar de forma paramétricamente a lo largo del barrido. A este tipo barrido se le conoce como *barrido general* (véase Figura 6.2) y es común su uso en visión computacional para la representación de superficies. Otro nombre con el cual se conoce a este método es *conos generalizados* [Ballard y Brown, 1982].

Podemos considerar a un cilindro generalizado como un cuerpo sólido cuyo eje es una curva en el espacio 3D, y para cada punto de esa curva corresponde una sección transversal (superficie bidimensional de barrido) ortogonal al eje. Continuando con la idea del barrido

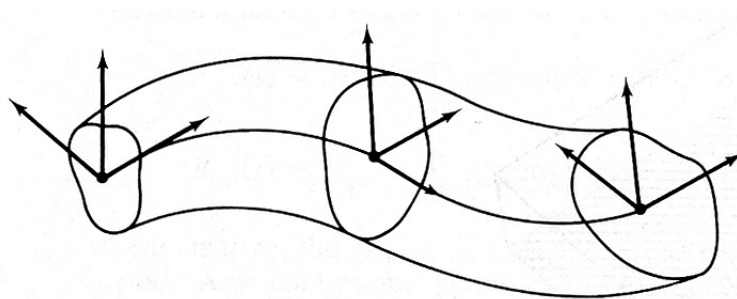


Figura 6.2: Ejemplo ilustrativo del barrido general, representación de cilindros generalizados. Imagen tomada de [Ballard y Brown, 1982]

general, se puede pensar que tanto la curva espacial (eje) y la forma de los contornos de las secciones transversales son funciones parametrizadas en la longitud de arco de dicha curva axial.

En nuestra aplicación decidimos utilizar un caso más particular de barrido conocido como *barrido traslacional*, en el cual las secciones transversales son de forma constante a lo largo de la curva axial. La forma elegida para la representación fue circular (una buena aproximación a la morfología real de los vasos sanguíneos) con un radio proporcional a la medición hecha por RISA al grosor de los vasos. Los círculos son de radio constante a lo largo de un mismo segmento de vaso. Entiéndase como segmento de un vaso a una de sus secciones comprendidas entre dos bifurcaciones, la sección entre la raíz y la primera bifurcación, y las ramas finales del vaso.

Por lo anterior, tenemos que cada segmento de vaso se modela como un cilindro generalizado de radio constante y cuyo eje es el esqueleto reconstruido anteriormente. Debido a la pérdida de la escala en los puntos reconstruidos, es necesario escalar los cilindros usando los parámetros de la matriz \mathbf{K} .

En cuestiones de implementación, las superficies de los vasos son mallas de polígonos cuadrangulares. La generación de dichas mallas consiste en que, para cada punto en 3D del esqueleto reconstruido (la curva axial), se le asocia un polígono regular que representa su respectiva sección transversal.

El polígono se centra en la posición del respectivo punto del esqueleto, es decir, el polígono se traslada de tal forma que la posición de su centro coincida con el punto de la curva

axial. Su orientación es determinada de tal forma que el vector tangente al esqueleto sea normal al plano que contiene al polígono. Los vértices del polígono funcionan como vértices de la malla. Las aristas se definen con la unión de cada vértice con sus dos vecinos adyacentes de la misma sección y con sus dos correspondientes vértices de secciones transversales vecinas, es decir, los vértices del mismo índice en polígonos adyacentes. En el caso de los puntos extremos del vaso, las aristas son determinadas de la misma forma, con la única excepción de que sólo se conectarán con un vértice de secciones trasversales vecinas. Como resultado de este algoritmo en una curva axial de prueba, se presenta la Figura 6.3.

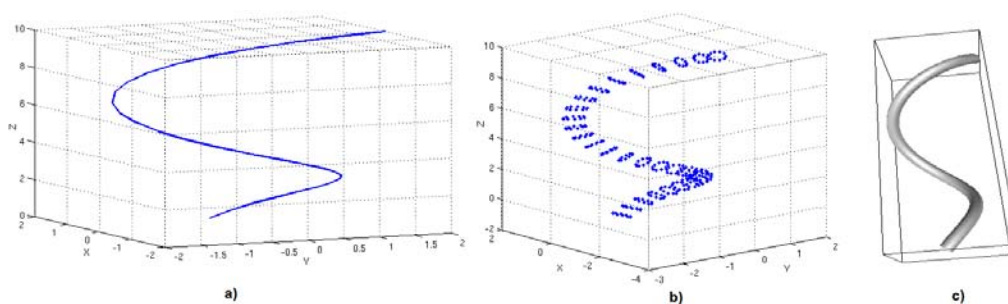


Figura 6.3: Modelo de superficie de un esqueleto prueba obtenido mediante el método de cilindros generalizados. (a) Esqueleto del objeto, (b) vértices de la malla cuadriculada y (c) modelo renderizado.

Una vez analizados los métodos de auto-calibración, realizar pruebas experimentales y haber elegido a uno de ellos por la calidad de sus resultados en reconstrucción, conocer los métodos de reconstrucción y el modelo de superficies a utilizar; sólo nos resta aplicar todas las etapas del sistema descrito en los capítulos anteriores a imágenes de fondo de ojo y reconstruir algunas estructuras retinianas. Es por ello que se presenta el siguiente capítulo.

Capítulo 7

Experimentación en la reconstrucción 3D de estructuras de retina

En este capítulo se presentan las pruebas realizadas a las diferentes partes del proceso de reconstrucción. Las imágenes utilizadas para estas pruebas fueron adquiridas con una cámara de fondo de ojo de la marca *Carl Zeiss*, modelo *FF450 plus* que cuenta con tres opciones de campo visual (FOV por sus siglas en inglés, *Field Of View*), las cuales son de 20°, 30° y 50°, sin embargo, las reconstrucciones hechas fueron con imágenes de 30° y 50° pues el área retinal capturada con 20° era insuficiente para encontrar un número aceptable de correspondencias al cambiar la pose de la cámara. Para digitalizar las imágenes capturadas por la cámara *fundus* se acopló una cámara *reflex* digital de la marca *Canon*, modelo *EOS 5D Mark II*. El tamaño de las imágenes a procesar fue de 720 x 480 píxeles. En las imágenes de retina sólo se utilizó la banda verde por ser esta la que caracteriza mejor a los vasos sanguíneos.

Además de las estructuras retinales, se decidió reconstruir dos escenas artificiales con el fin de cuantificar la calidad de los resultados en reconstrucción. Dichas escenas consisten, primeramente, en una rejilla regular de puntos que se encuentra fija a una superficie cilíndrica cuyas dimensiones son conocidas. Y en segundo lugar, un maniquí anatómico del ojo humano y en específico de la retina cuyas dimensiones fueron tomadas directamente del objeto. En ambos casos, las relaciones de distancias y los ángulos fueron calculados para comparar con las reconstrucciones.

Es importante señalar que la evaluación numéricas de las reconstrucciones se presenta en la sección 7.4 y que en las secciones anteriores a esta solo se presentarán las reconstrucciones de las diferentes escenas.

7.1. Preprocesamiento a las imágenes fundus

Corrección de distorsión

El primer proceso aplicado a las imágenes provenientes de la cámara *fundus*, es la corrección de distorsión por medio de *ADICOS* (véase sección 2.2). En la Figura 7.1 se presentan las imágenes del patrón de referencia (para modelar la distorsión) tomados con ambos campos visuales en las Figuras 7.1(a)-(c) a 30° y en las Figuras 7.1(d)-(e) a 50° , y el resultado de su corrección. Las Figuras 7.1(a) y (d) son las imágenes originales capturadas con la cámara *fundus*. Las Figuras 7.1(b) y (e) son los resultados de corregir la distorsión mediante el mapeo directo, de ahí la presencia de artefactos en las imágenes. Las Figuras 7.1(c) y (f) se obtienen mediante la corrección de distorsión mediante mapeo inverso con interpolación.

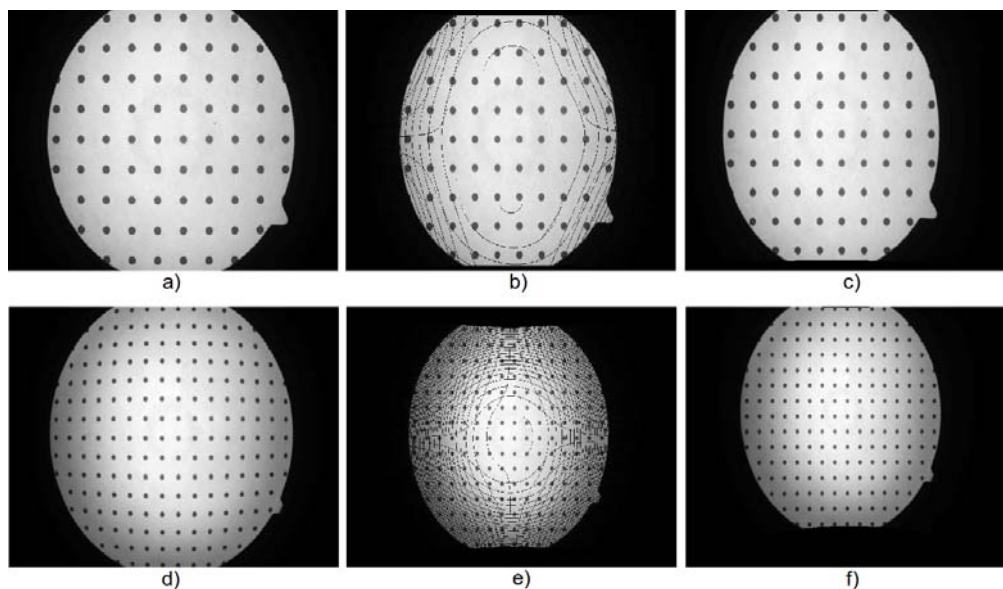


Figura 7.1: Resultados de corrección de distorsión a las imágenes de una rejilla regular de puntos. (a) Imagen original usando 30° de campo visual, (b) su resultado con mapeo directo y (c) su corrección con mapeo inverso. Igualmente, (d) la imagen de la rejilla original con 50° de campo visual, (e) su corrección con mapeo directo e (f) inverso.

Adicionalmente, en la Figura 7.2 se muestra la corrección hecha a una imagen generada artificialmente que consiste en una cuadrícula regular (Figura 7.2(a)), que por su naturaleza está libre de distorsión, con las funciones de transferencia de mapeo previamente almacenadas de cada campo visual (Figura 7.2(b) a 30° y Figura 7.2(c) a 50°). Lo anterior nos

muestra el efecto que tienen los mapas sobre las imágenes al compensar la distorsión y así eliminarla.

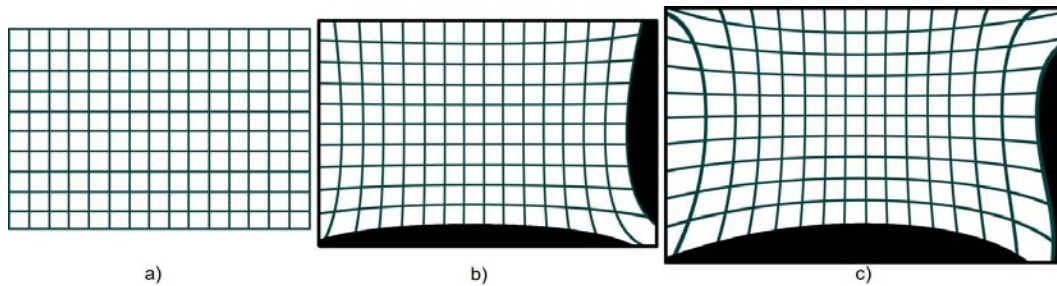


Figura 7.2: Aplicación de las funciones de transformación de mapeo a una (a) cuadrícula artificial originalmente libre de distorsión. Con el campo visual de (b) 30° y (c) 50° .

Ahora presentamos ejemplos de corrección de distorsión en imágenes de retina. Las Figuras 7.3(a) y (c) muestran las imágenes originales capturadas con la cámara de fondo de ojo para los campos visuales de 30° y 50° , respectivamente. En tanto que las Figuras 7.3(b) y (d) muestran las imágenes libres de distorsión por mapeo inverso con interpolación, para los campos visuales de 30° y 50° , respectivamente. Podemos observar que cuando la geometría de la escena es compleja o no es conocida, resulta difícil percibir la eliminación de la aberración de distorsión. De ahí la importancia de corroborar, en escenas con geometría simple o conocida, que la distorsión fue suprimida completamente.

Segmentación de los vasos sanguíneos retinales

Esta etapa de preprocesamiento tan importante se lleva a cabo con la ayuda del sistema RISA (*Retinal Image multiScale Analysis*) [Martinez-Perez *et al.*, 2007]. Debido a las características de las imágenes retinales, como son la gran variedad de grosores en los vasos sanguíneos, lo poco homogéneo del fondo de la retina y las complicaciones en la toma de las imágenes, no es posible aplicar métodos simples de segmentación. El método implementado en RISA se basa en el análisis multi-escala de la primera y segunda derivada espacial de las imágenes retinales, las cuales nos proveen de información topológica de los vasos sanguíneos. Gracias a las propiedades de la derivada, las variaciones en contraste del fondo no afectan al proceso de segmentación. Por otra parte, el análisis multi-escala otorga robustez al sistema en la detección de los vasos sin importar su grosor, longitud y orientación.

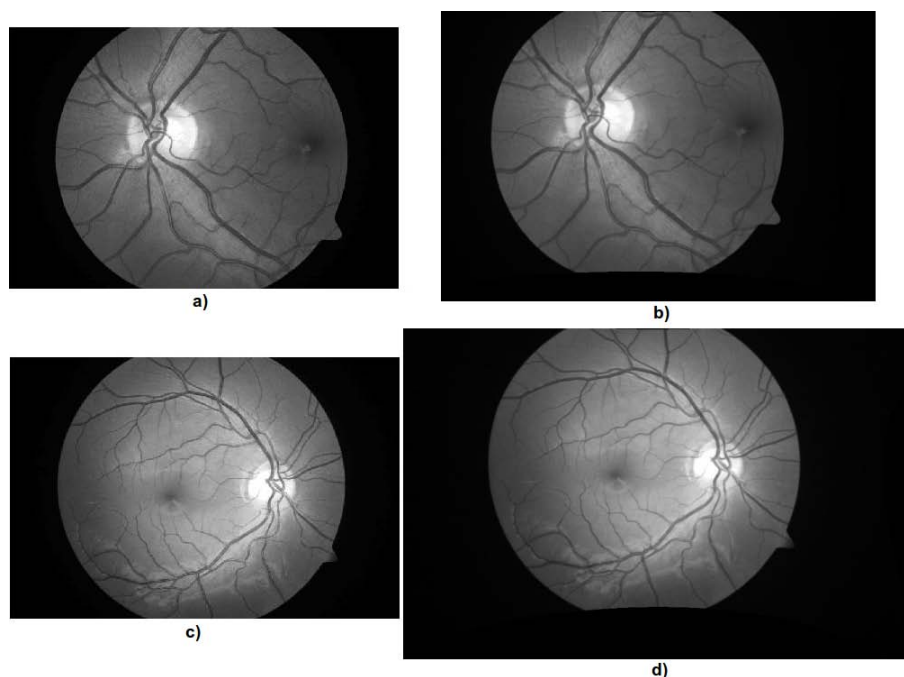


Figura 7.3: Imágenes retinales. (a) Imagen original a 30° , (b) imagen libre de distorsión a 30° , (c) imagen original a 50° y (d) la imagen libre de distorsión a 50° . Mapas obtenidos en el proceso de corrección de distorsión de la cámara de fondo de ojo.

Posteriormente, se aplica un procedimiento de dos etapas de crecimiento de regiones, las cuales emplean como parámetros: el máximo local de la magnitud del gradiente sobre las escalas y la curvatura principal máxima del tensor Hessiano. La clasificación de los píxeles se realiza sólo en clases distintas: fondo y vaso. En la primera etapa, el crecimiento es restringido a regiones con bajo gradiente, permitiendo un rápido crecimiento de las regiones fuera de los bordes. En la segunda etapa, el algoritmo hace crecer a las clases vaso y fondo simultáneamente sin restricciones del gradiente. Para mayores detalles del algoritmo de segmentación y descripción del sistema completo referirse a [Martínez-Pérez *et al.*, 2007].

En la Figura 7.4 se presentan resultados de la segmentación hecha por RISA, en imágenes tomadas con la cámara fundus antes mencionada y con otra de la misma compañía pero del modelo *Visucam Lite*, la cual ofrece una calidad mucho más baja en la captura, pues su resolución es mucho menor y la iluminación es deficiente. Sin embargo, se puede apreciar que la segmentación es correcta para ambos casos. En las Figuras 7.4(a) y (c) se presentan las imágenes de retina libres de distorsión y en las Figuras 7.4(b) y (d) se presentan las imágenes

binarias obtenidas por el sistema de segmentación, para los campos visuales de 35° (*Visucam Lite*) y 50° (*FF450 Plus*), respectivamente

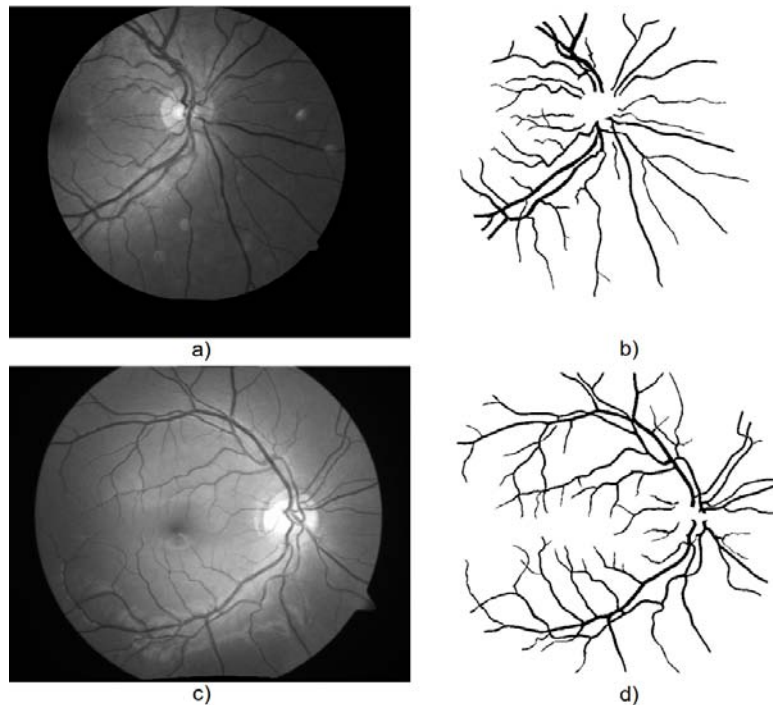


Figura 7.4: Segmentación realizada por RISA. (a) Imagen original tomada a 35° de campo visual con la cámara *Visucam Lite* y (b) su imagen binaria. (c) Imagen original tomada a 50° de campo visual con la cámara *FF450 plus* y (d) su respectiva imagen segmentada.

Adicionalmente, el sistema RISA nos provee de los puntos característicos a considerar en el cálculo de correspondencias. Un punto en la imagen es considerado significativo si es un punto de bifurcación, un punto de cruce entre dos vasos o si es el extremo final de un vaso. Como ejemplo de puntos característicos presentamos los marcados con una cruz roja en la imagen de retina de la Figura 7.5. A partir de las coordenadas de estos puntos se contruyen los descriptores para el cálculo de las correspondencias.

Cálculo de correspondencias

Como se describió en la sección 2.3, el algoritmo de GTM [Aguilar, 2006] necesita un apareamiento inicial entre los puntos característicos (provistos por RISA, véase sección 7.1)

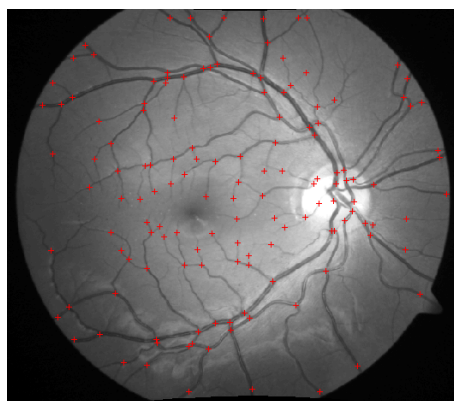


Figura 7.5: Puntos característicos seleccionados por el sistema RISA, los cuales son puntos de bifurcación, cruces entre vasos y puntos terminales.

de las imágenes para eliminar de ellos las correspondencias erróneas. Dicho apareamiento inicial fue calculado por el método de correlación entre los descriptores, los cuales consistían en parches o ventanas de 25×25 píxeles con el punto característico como pixel central. Las comparaciones entre los puntos fueron exhaustivas (todos contra todos). En la Figura 7.6(a) se muestran las correspondencias encontradas por correlación para esas imágenes. Los puntos característicos de la primera y segunda imagen están marcados con cruces rojas y verdes, respectivamente. Las líneas azules unen los puntos característicos de ambas imágenes, indicando el apareamiento calculado. Se puede observar que las correspondencias correctas son las líneas azules que tiene la misma inclinación y longitud. Los valores atípicos son las líneas con inclinación y longitud notablemente diferentes a la mayoría de las correspondencias mostradas.

Aplicando el algoritmo de GTM obtenemos los pares de puntos en correspondencia sin la presencia de valores atípicos, lo cual se muestra en la Figura 7.6(b). Los grafos (de ambas imágenes) construidos en la última iteración del algoritmo, con los vértices y aristas finales, se presentan en la Figura 7.6(c).

El cálculo de correspondencias se realiza sobre todas las parejas posibles de imágenes a utilizar en el proceso de auto-calibración. El número de parejas posibles es igual a $n(n-1)/2$ donde n es el número de imágenes. Como se vió en la sección 5.3 cada par de imágenes nos sirve para fijar restricciones sobre la imagen dual de la cónica absoluta y, con un número suficiente de imágenes, construir un sistema sobre determinado de ecuaciones con el cual hallaremos los coeficientes de la matriz de calibración. Con los datos extraídos de las imágenes pre-procesadas, procedemos a realizar las reconstrucciones de tres escenas diferentes,

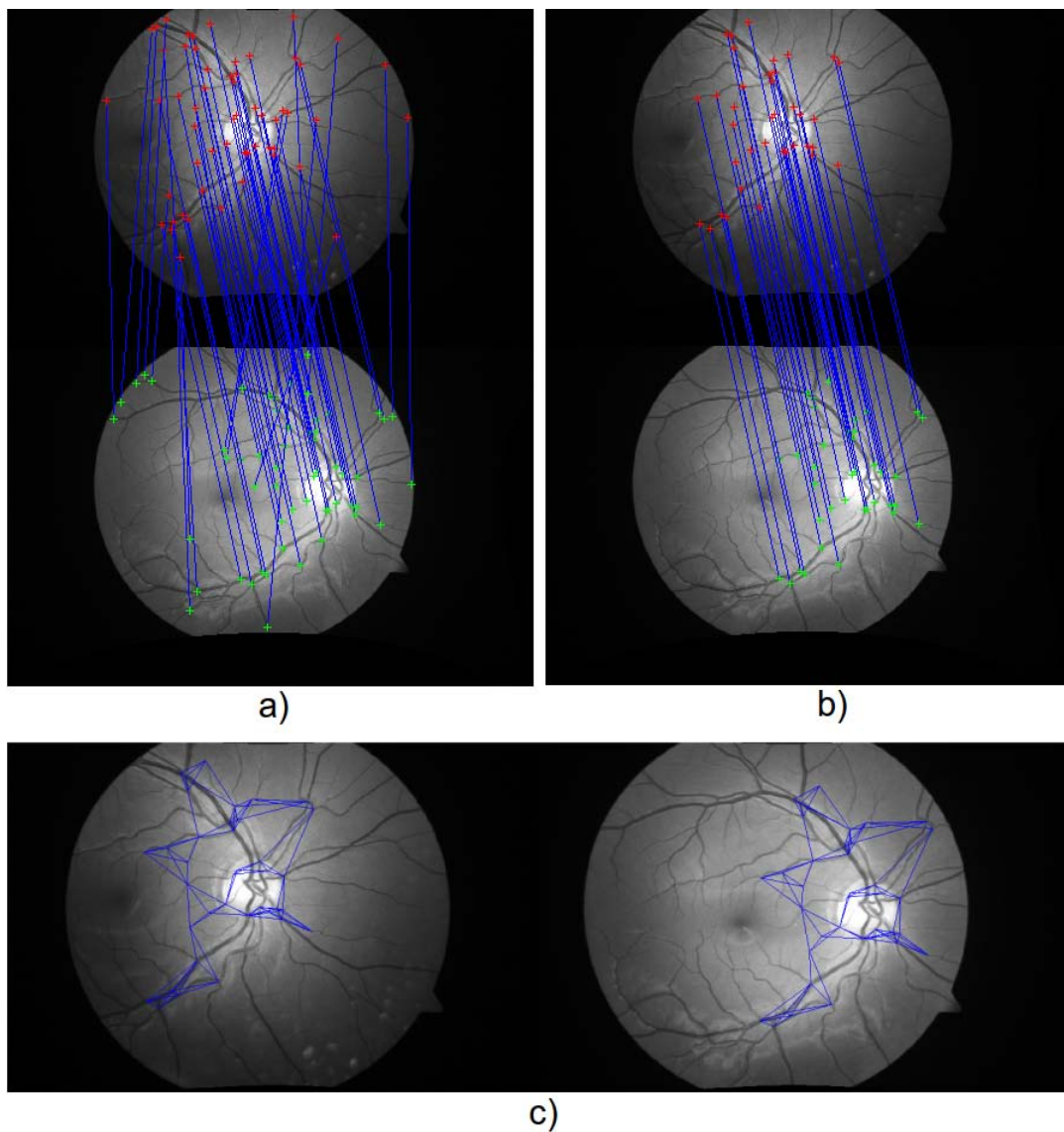


Figura 7.6: Ejemplo de las correspondencias calculadas por GTM. Los puntos característicos de la primera y segunda imagen están marcados con cruces rojas y azules, respectivamente. El apareamiento de puntos se indica con las líneas azules. (a) correspondencias iniciales calculadas por correlación con la presencia de *outliers* y (b) correspondencias finales sin errores. (c) las gráficas de la última iteración del algoritmo de GTM, las aristas se muestran con líneas azules.

descritas en la siguiente sección.

7.2. Estructuras reconstruidas calibrando únicamente la cámara

En esta sección se muestran los resultados de las reconstrucciones de tres diferentes escenas, para lo cual dividimos la presente sección en tres experimentos, uno para cada escena. En el primer experimento se captura un cilindro con una rejilla regular de puntos adherida a él (Fig. 7.7(a)), en el segundo experimento la escena consiste en un modelo anatómico del ojo humano (Fig. 7.7(b)) y en el último experimento, la escena es el fondo de ojo (Fig. 7.7(c)), capturada en imágenes con el filtro verde conocidas en el idioma inglés como *red-free* (libres de color rojo). Las secuencias de imágenes fueron adquiridas bajo las consideración para minimizar las aberraciones del sistema *cámara-ojo*, descritas en la sección 4.5, y usando los parámetros intrínsecos estimados mediante *Calib-tool* (véase sección 3.5).

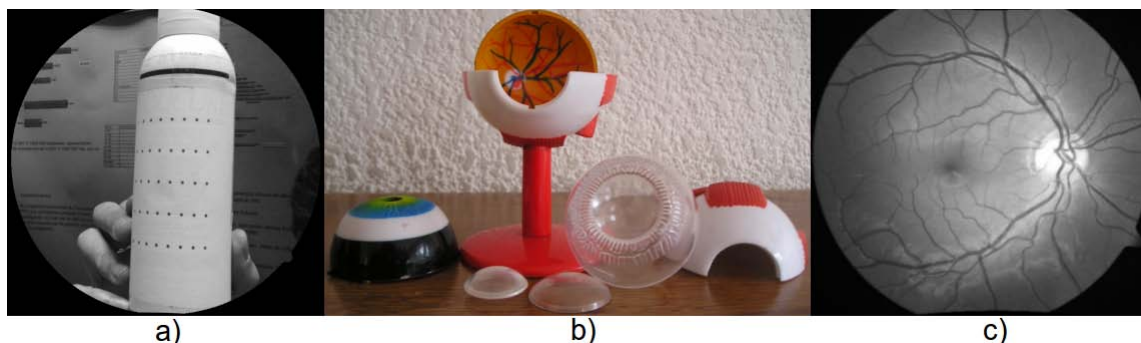


Figura 7.7: Escenas elegidas para la etapa de experimentación: (a) patrón regular de puntos adherido a una superficie cilíndrica, (b) modelo anatómico del ojo y (c) el fondo de ojo.

Experimento 1 - Rejilla cilíndrica de puntos

Las imágenes del patrón cilíndrico fueron capturadas, a diferencia de las otras dos escenas, con una cámara de la marca *Carl Zeiss*, modelo *Visucam Lite*. El principal motivo de realizar pruebas en reconstrucción de esta escena, es demostrar que el método propuesto anteriormente permite recuperar la curvatura de la superficie donde yacen los puntos característicos.

Lo cual no depende del equipo con el cual se adquieren las imágenes.

Se cuenta con dos conjuntos de imágenes del patrón cilíndrico, el primero fue tomado con un campo visual de 35° y el segundo con 45° . El motivo para usar dos campos visuales diferentes es averiguar si existen diferencias en el resultado de la reconstrucción debido al cambio en esta variable. A continuación presentamos el análisis completo del campo visual de 35° . En la Figura 7.8 se muestran las cinco imágenes con las cuales se formaron 10 parejas diferentes para obtener reconstrucciones de las cuales elegir la mejor, y minimizar la posibilidad de caer en una configuración crítica de cámaras [Hartley y Zisserman, 2004]. Adicionalmente, estas parejas de imágenes de la escena se usan para estimar los parámetros internos de la cámara por auto-calibración. El tamaño original de las imágenes es de 720×576 píxeles y se encuentran en escala de grises desde la adquisición.

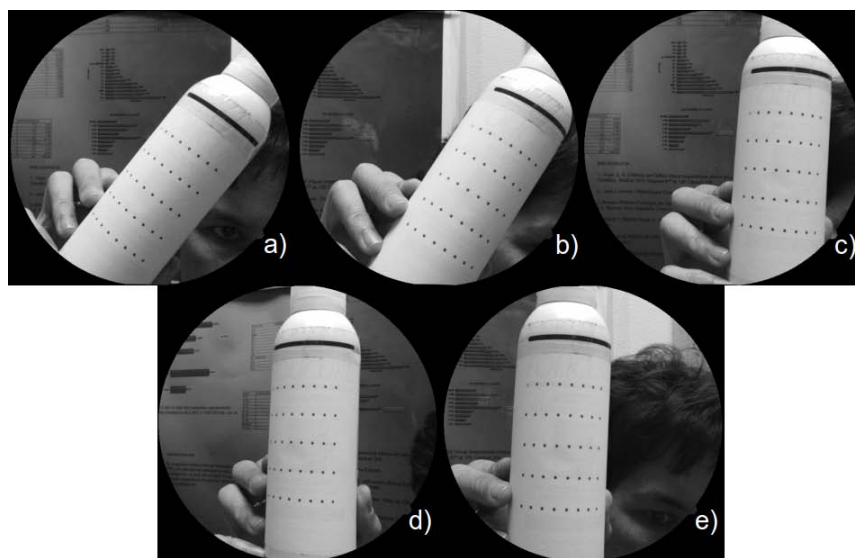


Figura 7.8: (a)-(e) Conjunto de imágenes del patrón cilíndrico a 35° de campo visual, en escala de grises y libres de distorsión por *ADICOS*.

La matriz de calibración correspondiente a 35° con la cámara *Visucam Lite* obtenidas por el método clásico (*Calib-tool*) se presenta en la ecuación (7.1), en la cual podemos ver los parámetros internos de la cámara, tal como se describieron la sección 3.1. De las matrices de calibración podemos ver que el punto principal se encuentra muy cercano al centro de la imagen y que la relación de aspecto es 0.9983, prácticamente 1. Estos datos son congruentes con lo esperado de una cámara de uso profesional, con un sensor CCD de píxeles cuadrados

y un sistema de lentes montados con alta precisión.

La reconstrucción de los puntos de la rejilla con la pareja de imágenes formada por la Fig. 7.8(d) y (e) se muestra desde diferentes puntos de vista en la Figura 7.9.

$$\mathbf{K}_{v35c} = \begin{bmatrix} 971.1898 & 0 & 349.5883 \\ 0 & 969.6030 & 307.7361 \\ 0 & 0 & 1 \end{bmatrix} \quad (7.1)$$

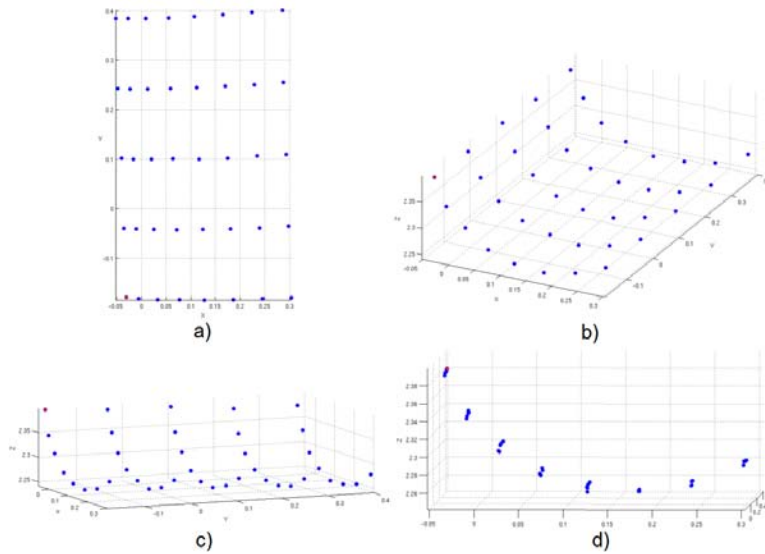


Figura 7.9: (a)-(d) Vistas del patrón de puntos cilíndrico reconstruido usando las imágenes con campo visual de 35° , estimando los parámetros internos con el método clásico de calibración.

Ahora bien, pasemos al caso del campo visual de 45° . El conjunto de 5 imágenes se presentan en la Figura 7.10, recordemos que el equipo utilizado para la captura es el *Visucam Lite* y que el tamaño de las imágenes es de 720×576 píxeles. La matriz de calibración correspondiente a 45° obtenidas por *Calib-tool*, se presenta en la ecuación (7.2). La reconstrucción de los puntos de la rejilla con la pareja de imágenes formada por la Fig. 7.10(b) y (c) se muestra desde diferentes puntos de vista en la Figura 7.11.

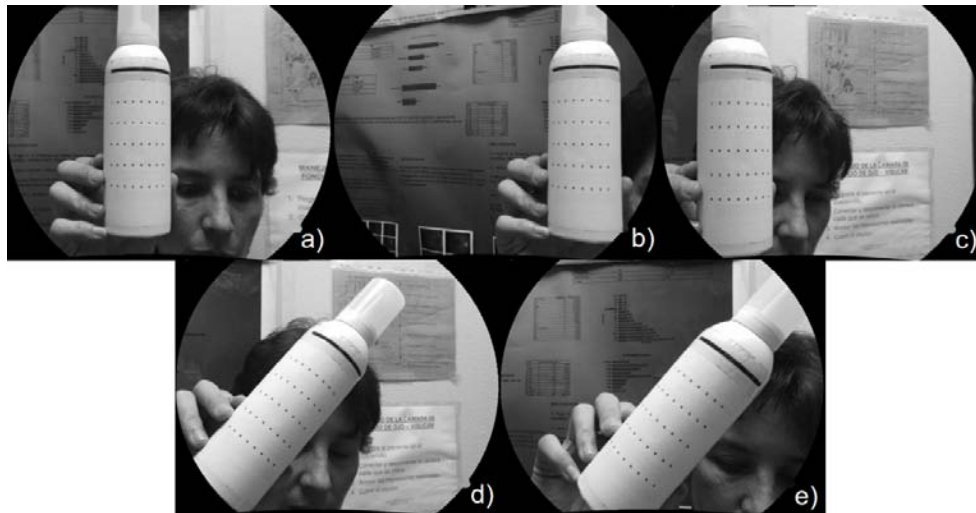


Figura 7.10: (a)-(e) Conjunto de imágenes del patrón cilíndrico a 45° de campo visual, en escala de grises y libres de distorsión por *ADICOS*.

$$\mathbf{K}_{v45c} = \begin{bmatrix} 973.8952 & 0 & 327.2782 \\ 0 & 975.5619 & 285.9094 \\ 0 & 0 & 1 \end{bmatrix} \quad (7.2)$$

Para esta escena en particular, fue posible realizar un análisis cuantitativo de la calidad de la reconstrucción, la cual se presenta en la sección 7.4.

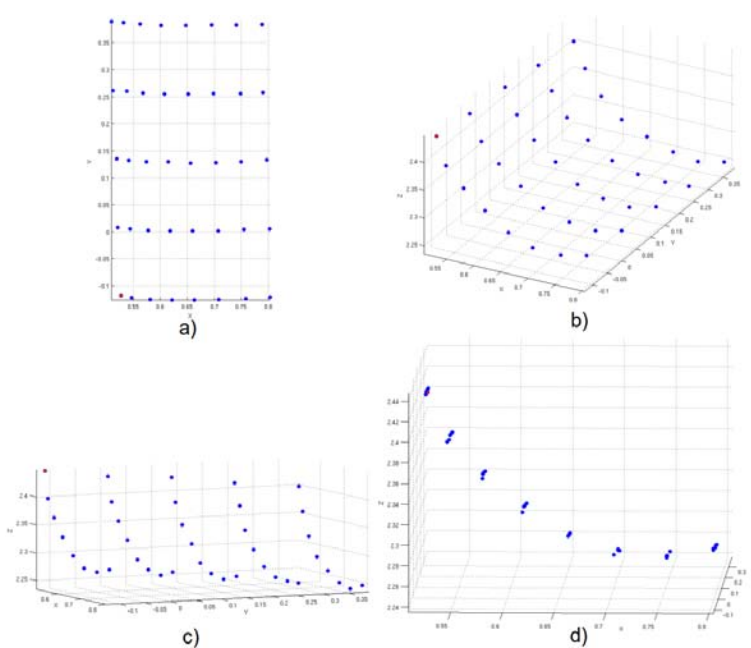


Figura 7.11: (a)-(d) Vistas del patrón de puntos cilíndrico reconstruido usando las imágenes con campo visual de 45° , estimando los parámetros internos con el método clásico de calibración.

Experimento 2 - Modelo anatómico del ojo

La segunda escena a reconstruir (Fig. 7.7(b)) se eligió con el fin de contar con estructuras de morfología similar a la retina (exceptuando el efecto óptico de ojo) y así corroborar la reconstrucción de modo cualitativo pues podemos acceder directamente a la escena. En la Figura 7.12 se muestran las 5 imágenes tomadas al modelo anatómico del ojo que fueron usadas para el proceso de reconstrucción. Estas imágenes se tomaron con la cámara de fondo de ojo *Carl Zeiss FF450plus* sincronizada con la cámara digital *Canon 5D Mark II*, y usando únicamente el campo visual de 50° . La cámara *fundus* cuenta adicionalmente con campos visuales de 30° y 20° , los cuales tienen un factor de magnificación mayor porque se usan para inspeccionar a detalle ciertas regiones de la retina.

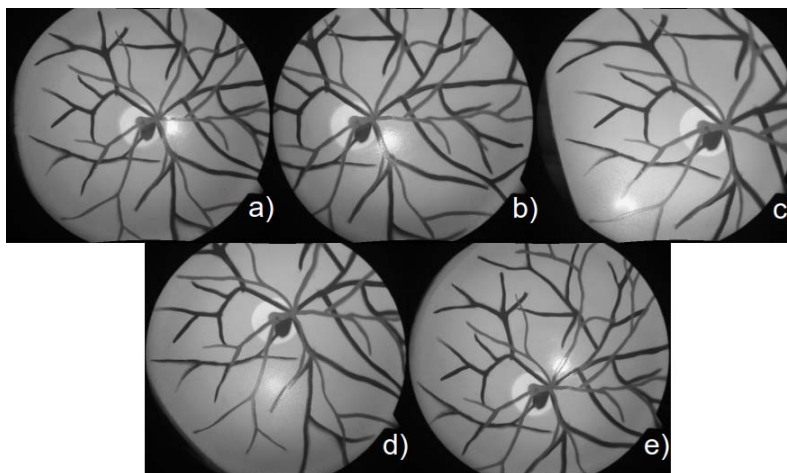


Figura 7.12: (a)-(e) Conjunto de imágenes del modelo anatómico del ojo a 50° de campo visual, en escala de grises y libres de distorsión por *ADICOS*.

A mayor magnificación, el área de la región capturada es menor, lo cual implica en esta aplicación que: la superficie donde se encuentran los puntos a reconstruir aparenta ser plana (pues la sección de la esfera es muy pequeña), un número reducido de pares de puntos característicos en correspondencia pueden observarse a lo largo de la secuencia de imágenes, y se cuenta con poca libertad en los movimientos para cambiar la pose de la cámara. Las implicaciones anteriores ocasionan que los parámetros intrínsecos, la pose relativa de la segunda cámara respecto a la primera y por consiguiente la recuperación de la posición 3D de los puntos por triangulación lineal sea calculada de manera imprecisa cayendo en las ambigüedades del método de Kruppa y en una correcta interpretación de la *quiralidad* en \mathbb{P}^3 .

Es por lo anterior que decidimos reconstruir usando sólo imágenes de 50° de campo visual, para las escenas del maniquí y la retina.

La matriz de calibración de la cámara *FF450plus* usando el campo visual de 50°, calculado por el método clásico de calibración (*Calib-tool*), se presenta en la ecuación (7.3).

$$\mathbf{K}_{f50c} = \begin{bmatrix} 639.7870 & 0 & 460.7453 \\ 0 & 639.3892 & 247.0214 \\ 0 & 0 & 1 \end{bmatrix} \quad (7.3)$$

A modo de referencia, presentamos en la Figura 7.13 los esqueletos de los vasos seleccionados en la etapa de medición del sistema *RISA*. La asignación de colores permanece constante en las gráficas presentadas más adelante, con el objetivo de identificar con facilidad a cada uno de los vasos.

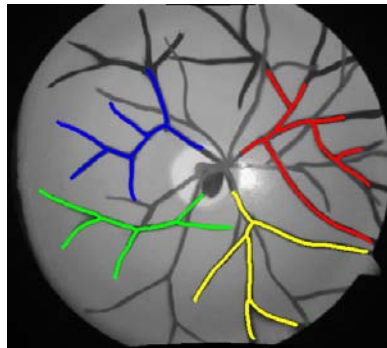


Figura 7.13: Esqueletos de los árboles seleccionados, y medidos por *RISA*, del modelo anatómico del ojo para la reconstrucción.

Diferentes vistas de la reconstrucción 3D de los esqueletos del modelo, empleando la matriz de calibración de la ecuación (7.3), se muestran en la Figura 7.14(a)-(d). Es importante mencionar que los esqueletos (en el modelo y la retina) son pre-procesados para suavizar la curva que describen en el espacio y para aumentar su densidad de puntos, para mejorar la visualización de sus respectivos modelos de superficies. La suavización se realizó con un filtro promedio cuya ventana o *kernel* de 11 píxeles centrado y la densidad de puntos se aumentó en un 200 % por medio del ajuste de un spline bicúbico a los puntos posicionados en el centro de cada pixel que pertenece al esqueleto.

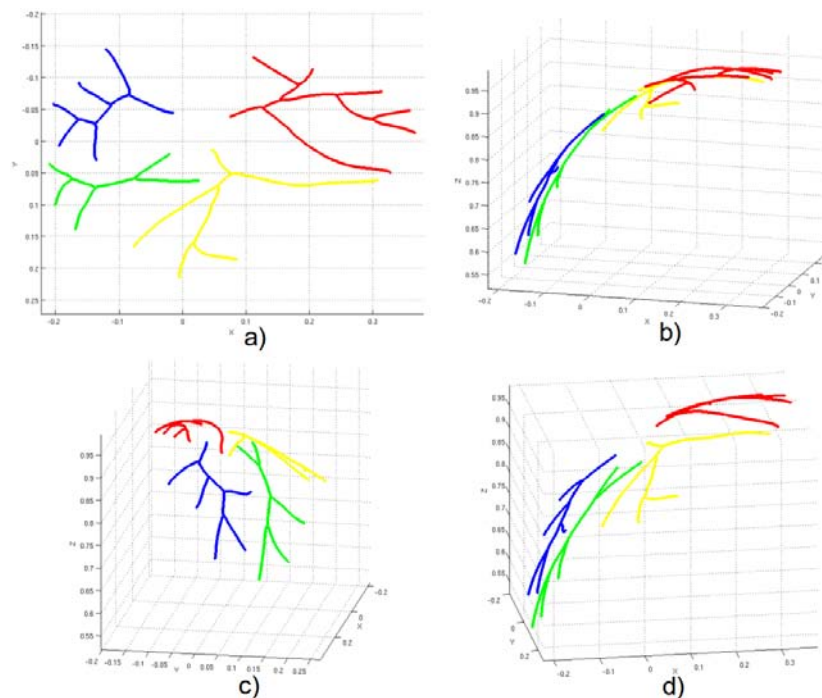


Figura 7.14: (a)-(d) Esqueletos del modelo anatómico del ojo reconstruidos en 3D usando calibración clásica (*Calib-tool*). Los colores de los esqueletos corresponden a los mostrados en la Figura 7.13.

Experimento 3 - Imágenes retinales

Por parte de las reconstrucciones haciendo uso de la calibración clásica (*Calib-tool*), tenemos por último las imágenes de fondo de ojo. Las cuales fueron capturadas, como se mencionó con anterioridad, usando el filtro verde para brindandonos un mejor contraste que el obtenido mediante la conversión de las imágenes a color (RGB) a escala de grises, introduciendo al sistema de reconstrucción únicamente el canal verde. Las 5 imágenes retinales usadas se muestran en la Figura 7.15.

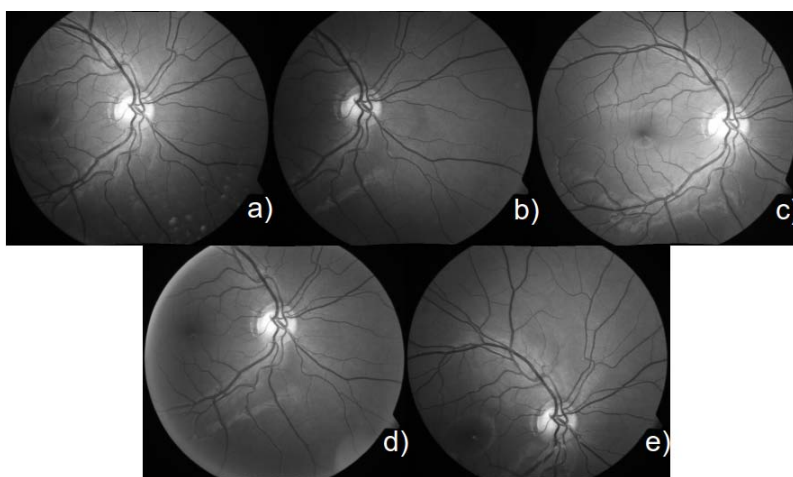


Figura 7.15: (a)-(e) Conjunto de imágenes retinales a 50° de campo visual, tomando únicamente el canal verde de las imágenes digitales, debido al uso del filtro *red-free* y libres de distorsión por *ADICOS*.

Los parámetros internos usados en este caso son los mismos que en la reconstrucción de modelo anatómico (véase la ecuación (7.3)), pues las imágenes fueron capturadas bajo las condiciones predefinidas para minimizar la presencia de las aberraciones introducidas por el globo ocular. La Figura 7.16 muestra los esqueletos de los vasos retinales elegidos y medidos para ser reconstruidos usando como identificador su color.

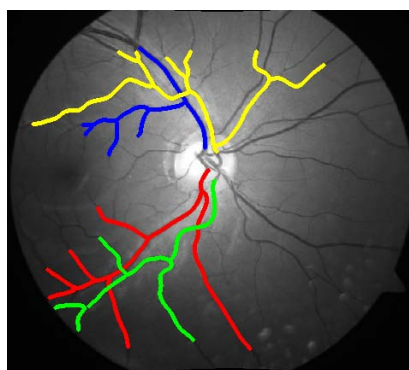


Figura 7.16: Esqueletos de los árboles seleccionados, y medidos por *RISA*, en las imágenes *fundus* para la reconstrucción, marcados por colores.

La reconstrucción de las estructuras en la retina correspondiente a las anteriores condiciones y parámetros se presentan en la Figura 7.17.

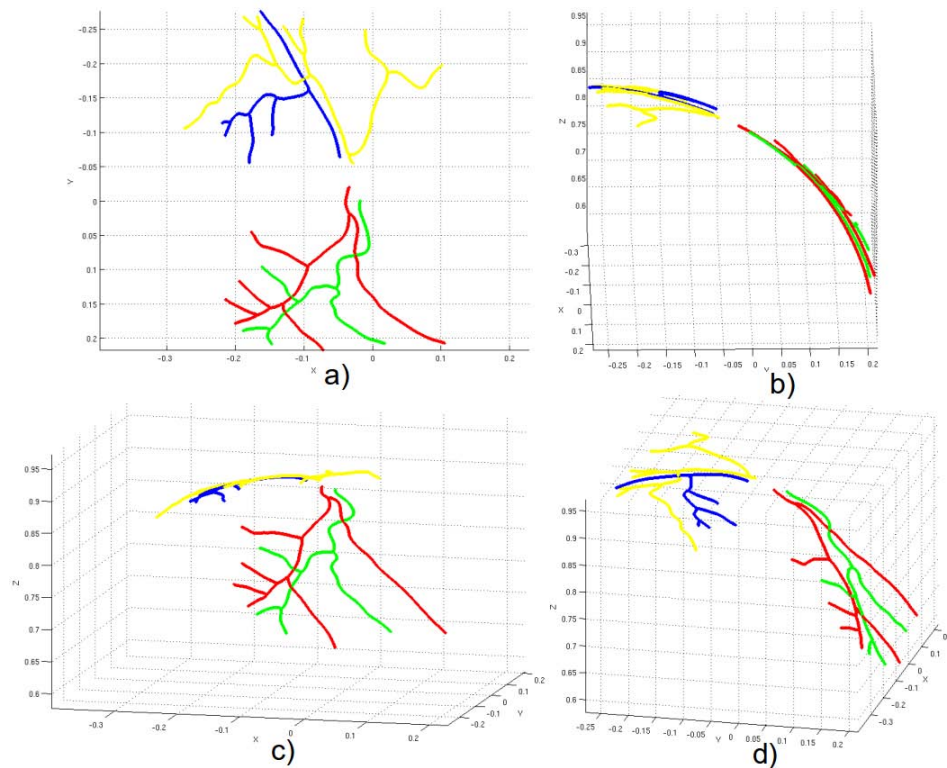


Figura 7.17: (a)-(d) Esqueletos en las imágenes retinianas reconstruidos en 3D usando calibración clásica. Los colores de los esqueletos corresponden a los mostrados en la Figura 7.16.

En las Figuras 7.14 y 7.17 puede verse que la curvatura de las superficies en las cuales se encuentran los puntos reconstruidos, se recuperó notablemente. La morfología de las estructuras concuerdan con lo observado en las imágenes segmentadas, sin embargo, los vasos (del modelo anatómico y de la retina) más alejados del nervio óptico cambiaron ligeramente de forma. Esto último difícilmente se observa en las gráficas presentadas.

Habiendo mostrados las reconstrucciones de las tres escenas propuestas, calibrando únicamente la cámara *fundus* y despreciando los efectos de la presencia de la óptica del ojo (en las imágenes retinianas) por la aplicación de las consideraciones de minimización de aberraciones

en la configuración de las cámaras *fundus* y digital, pasamos a presentar las reconstrucciones usando un modelo de cámara estimado por auto-calibración.

7.3. Estructuras reconstruidas usando auto-calibración

Ahora presentamos las reconstrucciones que emplean los parámetros intrínsecos de la cámara estimados por las ecuaciones de Kruppa (véase sección 5.4). Debemos recordar que este método de auto-calibración emplea pares de imágenes para calcular matrices fundamentales con los puntos en correspondencia y, a partir de dichas matrices, se calculan las matrices esenciales, finalmente después del método obtenemos los parámetros internos de la cámara. Es por eso que decidimos emplear 5 imágenes con las cuales formamos 10 parejas, a pesar de sólo ser necesarias dos para recuperar las poses de las cámaras y reconstruir. Por cada conjunto de imágenes se obtiene una diferente \mathbf{K} . Al igual que en la sección anterior, presentamos las pruebas divididas en experimentos según la escena a reconstruir.

Experimento 1 - Rejilla cilíndrica de puntos

Empecemos con la reconstrucción de la rejilla regular de puntos en la superficie cilíndrica usando las imágenes capturadas con el campo visual de 35° mostrada en las Figura 7.18 en diferentes puntos de vista.

La matriz de calibración empleada para obtener esta reconstrucción (campo visual de 35°) se presenta en la ecuación (7.4).

$$\mathbf{K}_{v35a} = \begin{bmatrix} 973.8948 & 0 & 327.2776 \\ 0 & 985.4284 & 285.9113 \\ 0 & 0 & 1 \end{bmatrix}, \quad (7.4)$$

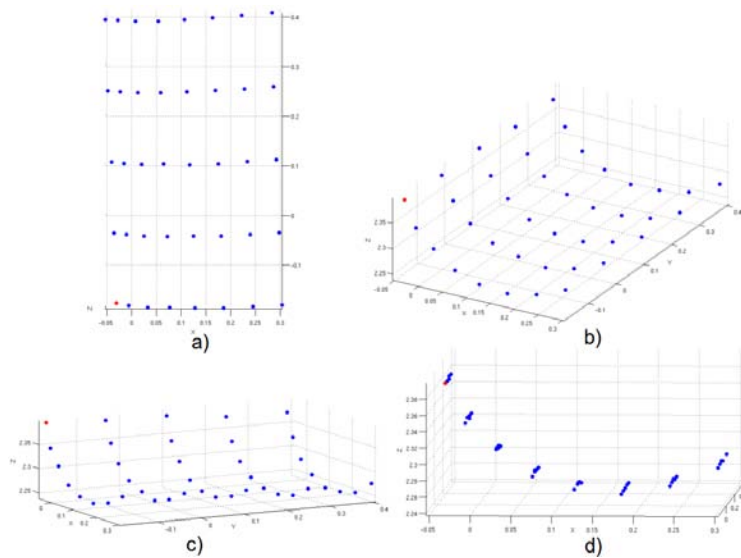


Figura 7.18: (a)-(d) Vistas del patrón de puntos cilíndrico reconstruido usando las imágenes con campo visual de 35° , estimando los parámetros internos con auto-calibración.

Ahora bien, presentemos la reconstrucción del cilindro con el campo visual de 45° . La reconstrucción se muestra desde diferentes puntos de vista en la Figura 7.19.

La matriz de calibración empleada para obtener esta reconstrucción (campo visual de 45°) se presenta en la ecuación (7.5).

$$\mathbf{K}_{v45a} = \begin{bmatrix} 973.896 & 0 & 327.2767 \\ 0 & 1060.6838 & 285.9081 \\ 0 & 0 & 1 \end{bmatrix}. \quad (7.5)$$

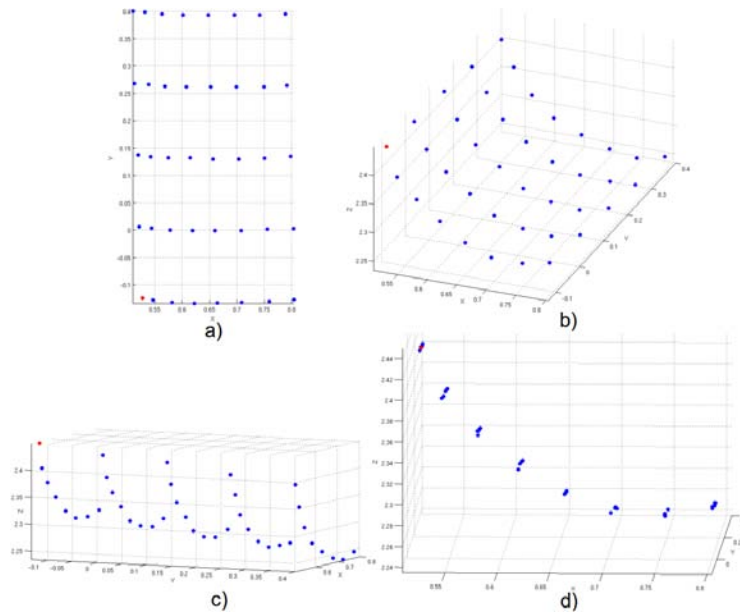


Figura 7.19: (a)-(d) Vistas del patrón de puntos cilíndrico reconstruido usando las imágenes con campo visual de 45° , estimando los parámetros internos con auto-calibración.

Experimento 2 - Modelo anatómico del ojo

La matriz de calibración obtenida con el conjunto de imágenes del modelo con un campo visual de 50° (*FF450 Plus*) se muestra en la ecuación (7.6).

$$\mathbf{K}_{f50am} = \begin{bmatrix} 639.7891 & 0 & 460.7410 \\ 0 & 589.8010 & 247.0209 \\ 0 & 0 & 1 \end{bmatrix} \quad (7.6)$$

Las estructuras reconstruidas se pueden observar en la Figura 7.20.

Experimento 3 - Imágenes retinales

Finalmente, con la secuencia de imágenes de la retina real se obtiene la matriz de parámetros internos presentada en la ecuación (7.7).

$$\mathbf{K}_{f50ar} = \begin{bmatrix} 616.6744 & 1 & 483.3050 \\ 0 & 759.3333 & 226.5309 \\ 0 & 0 & 1 \end{bmatrix} \quad (7.7)$$

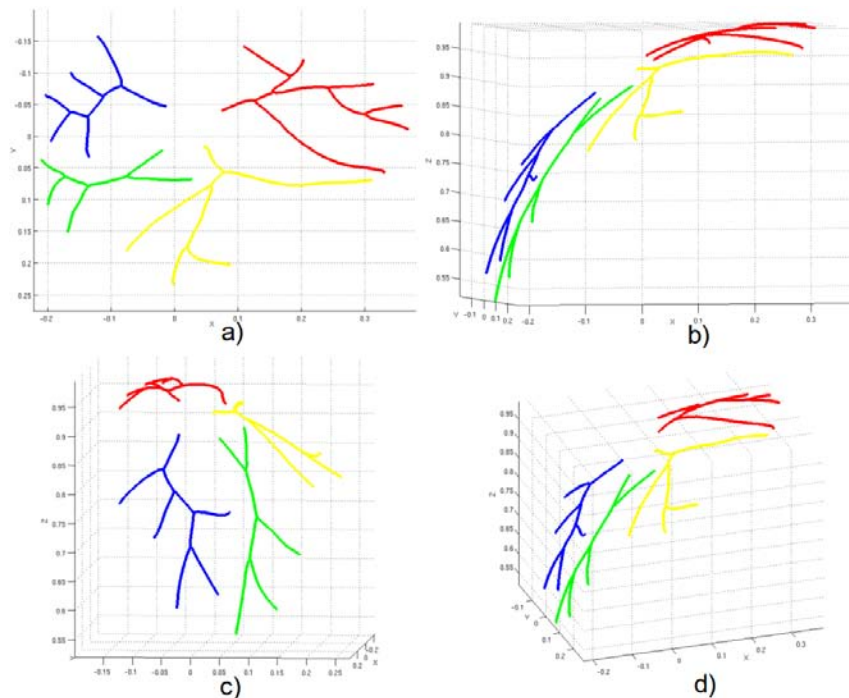


Figura 7.20: (a)-(d) Esqueletos del modelo anatómico del ojo reconstruidos en 3D usando auto-calibración. Los colores de los esqueletos corresponden a los mostrados en la Figura 7.13.

Las reconstrucciones de los vasos retinales calculadas en base a la anterior matriz de calibración se muestran en la Figura 7.21.

En las Figuras 7.20 y 7.21 puede verse que la curvatura de las superficies en las cuales se encuentran los puntos reconstruidos, se recuperó correctamente pues presenta simetría de revolución, ya que estamos suponiendo que tiene forma esférica. Las estructuras de toda la escena reconstruida concuerdan en morfología con las observadas en las imágenes segmentadas, no importando su posición en las imágenes.

Hasta este momento sólo se han presentado las matrices estimadas, ya sea por calibración clásica (*Calib-tool*) o auto-calibración, y las respectivas reconstrucciones. Sin embargo, es necesario establecer una métrica (sí el caso lo permite) para calificar la calidad de las reconstrucciones o bien hacer un análisis cualitativo, lo cual se verá en la siguiente sección.

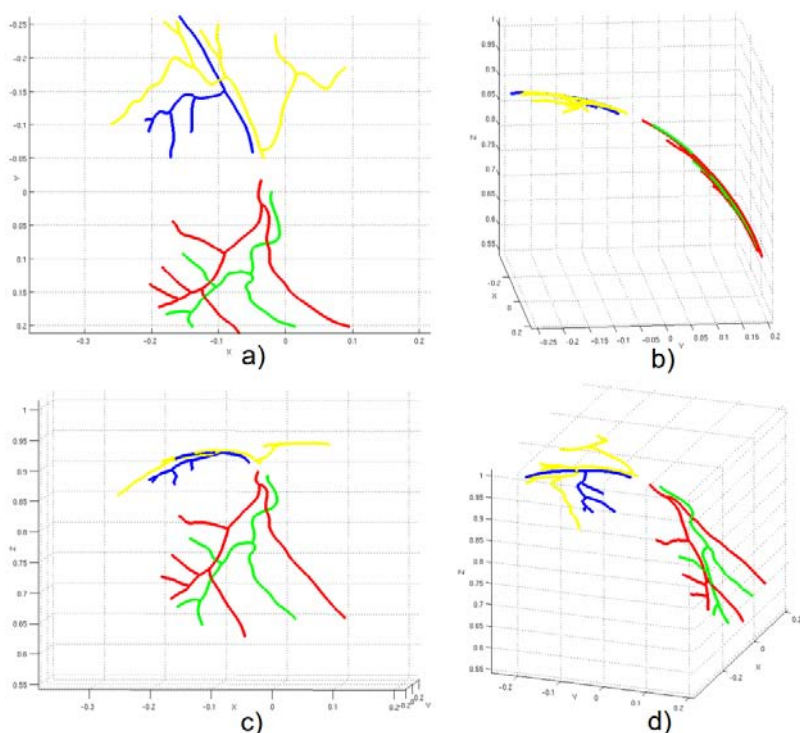


Figura 7.21: (a)-(d) Esqueletos en las imágenes retinianas reconstruidos en 3D usando auto-calibración. Los colores de los esqueletos corresponden a los mostrados en la Figura 7.16.

7.4. Comparación de reconstrucciones por método clásico y auto-calibración

Las reconstrucciones mostradas en las secciones 7.2 y 7.3 fueron seleccionadas como las mejores de cada uno de los conjuntos de 10 reconstrucciones respectivas, es decir, por cada pareja de imágenes obtenemos una reconstrucción de las cuales sólo se selecciona la mejor. El criterio de selección para todas las escenas fue cualitativo salvo en el caso el cilindro, cuya calidad fue medida cuantitativamente en base a un conjunto de parámetros que se midieron tanto en la escena real con en los puntos reconstruidos en 3D.

Los parámetros medidos son: el ángulo superior izquierdo (ASI), el ángulo inferior izquierdo (AII), el ángulo entre aristas (AEA), la sumatoria de ángulos internos (SAI), la razón de longitudes ortogonales (RLO), la curvatura o razón entre la secante y el radio (CUR), la razón de longitudes horizontales (RLH) y la razón de longitudes verticales (RLV). En la

Figura 7.22 se muestra una rejilla de puntos de forma cilíndrica generada artificialmente, haciendo referencia a la escena del patrón cilíndrico capturado en las imágenes de las Figuras 7.8 y 7.10. Se seleccionaron arbitrariamente algunos puntos para tomar medidas de ángulos formados entre líneas de puntos, así como también distancias entre ellos para corroborar que las reconstrucciones recuperen las propiedades métricas de la escena.

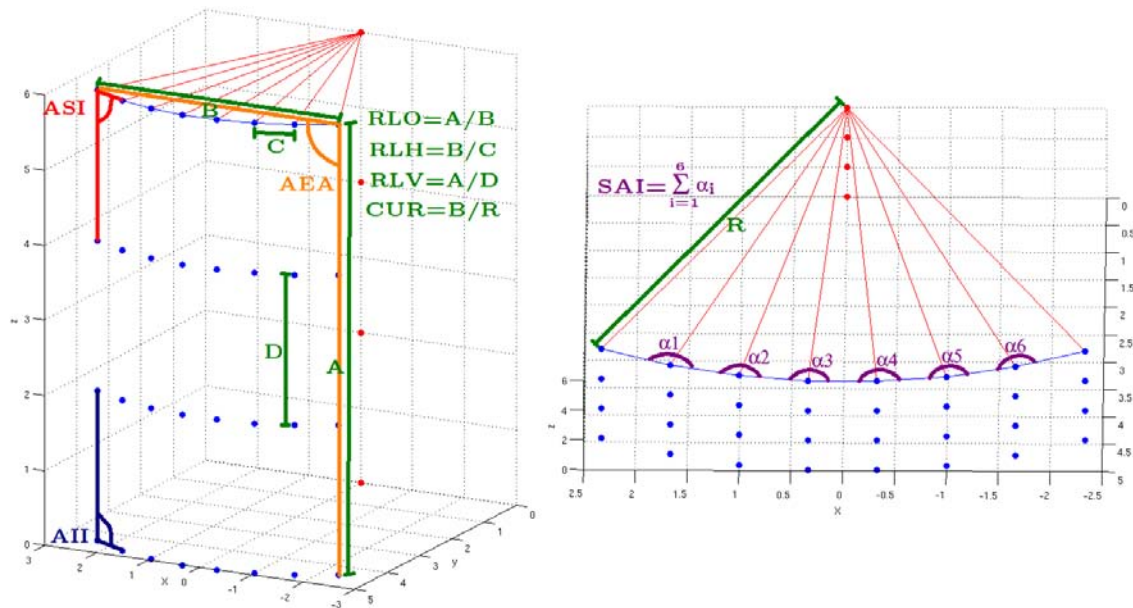


Figura 7.22: Interpretación gráfica de los parámetros elegidos para calificar la calidad de la reconstrucción del patrón de puntos posicionados en una superficie cilíndrica.

La tabla 7.1 presenta las mediciones tomadas sobre la escena *Real*, sobre la escena reconstruida (*Reconstrucción*) y los errores relativos (*Error*) correspondientes al cilindro reconstruido usando la calibración de la cámara *fundus* únicamente (método clásico) para los dos campos visuales empleados con la *visucam*.

Parámetros	Campo visual de 35°			Campo visual de 45°	
	Real	Reconstrucción	Error(%)	Reconstrucción	Error(%)
ASI	90°	91.2404°	1.3782	89.2764°	0.8040
AII	90°	89.9497°	0.0559	90.6555°	0.7283
AEA	90°	88.6133°	1.5408	90.5653°	0.6281
SAI	1000.5218°	17.3651°	0.5570	998.1326°	0.2389
RLO	1.6094	1.6107	0.0807	1.6256	1.0084
CUR	1.4471	1.1996	12.8642	1.2754	10.7263
RLH	6.2756	6.3862	1.7624	6.2324	0.6888
RLV	3.9935	4.0485	1.3764	4.0376	1.1031

Tabla 7.1: Parámetros medidos en la escena real de la rejilla cilíndrica y en la reconstrucción usando los parámetros internos estimados por calibración clásica, usando las imágenes con campos visuales 35° y 45°. Los errores relativos son presentados en porcentajes.

Como podemos apreciar en la tabla 7.1 la reconstrucción es altamente precisa en cuanto a la recuperación de las propiedades métricas de la escena, ya que las diferentes razones de longitud presentan errores relativos pequeños. En tanto a los ángulos elegidos como parámetros, resultan reconstruidos de manera muy precisa (según indican los errores relativos) salvo la curvatura (*CUR*), y esto se debe a que el cálculo del radio en la reconstrucción consiste en aplicar leyes de senos y propiedades de triángulos sobre tres puntos adyacentes de la misma fila, lo cual es de carácter local y, por lo tanto, muy sensible al ruido. Debido a eso, se propone el parámetro *SAI* que estima con mayor robustez la recuperación de la curvatura de la superficie cilíndrica.

En base a los parámetros definidos anteriormente, presentamos en la tabla 7.2 las mediciones hechas sobre las reconstrucciones estimadas con el modelado de la cámara por auto-calibración y los respectivos errores relativos.

De las matrices de calibración obtenidas por el método clásico (en la ecuaciones (7.1) y (7.2)) y por auto-calibración (en la ecuaciones (7.4) y (7.5)) podemos apreciar que las variaciones son mínimas y son ocasionadas por el proceso de minimización no-lineal y la inicialización ingresada al método de auto-calibración, que como se definió en el capítulo 5, se trata de valores correspondientes a un modelo genérico de una cámara *pin-hole*. Es importante notar que la variación mínima en los valores de matrices de calibración es congruente con el hecho de que al sistema óptico de la cámara no se le añade, o interpone entre él y el objeto, algún elemento externo como en el caso de las imágenes *fundus* al añadir la óptica del ojo.

Parámetros	Campo visual de 35°			Campo visual de 45°	
	Real	Reconstrucción	Error(%)	Reconstrucción	Error(%)
ASI	90°	91.2169°	1.3521	88.3582°	1.8242
AII	90°	90.5560°	0.6178	91.1737°	1.3041
AEA	90°	88.6562°	1.4931	90.7648°	0.8498
SAI	17.4624°	994.8323°	0.5689	999.7942°	0.0727
RLO	1.6094	1.6399	1.8955	1.6739	4.0079
CUR	1.4471	1.2248	11.1673	1.2940	6.9958
RLH	6.2756	6.3895	1.8157	6.0952	2.8742
RLV	3.9935	4.0440	1.2643	4.0068	0.3325

Tabla 7.2: Parámetros medidos en la escena real de la rejilla cilíndrica y en la reconstrucción usando los parámetros internos estimados por auto-calibración, usando las imágenes con campos visuales 35° y 45°. Los errores relativos son presentados en porcentajes.

Comparando los errores relativos de las reconstrucciones con diferente valor de campo visual, podemos observar que las diferencias son mínimas y para el caso práctico son despreciables, es decir, el ángulo campo visual no afecta a la calidad de la reconstrucción siempre y cuando no se comprometa la posibilidad de cambiar considerablemente la poses de las cámaras.

En la escena del modelo anatómico del ojo, en primera instancia se realizó una inspección cualitativa para decidir con que método de estimación de parámetros de la cámara (sistema *cámara-ojo*) se obtienen las reconstrucciones más aproximadas a la realidad. Principalmente, se eligieron las estructuras cuyos puntos describan una superficie esférica o pseudo-esférica y que la vista paralela al plano XY luzca lo más parecido posible a la segmentación de los árboles seleccionados sobre la primera imagen de la pareja correspondiente. Para apreciar este aspecto pueden compararse, para el caso el modelo anatómico, las imágenes de las Figuras 7.14(a) obtenidas con la calibración del *Calib-tool* y 7.20(a) obtenida por auto-calibración (Kruppa), con la ilustración de los vasos seleccionados para la reconstrucción mostrados en la Figura 7.13.

Como parte adicional de la evaluación del sistema, realizada en una etapa final del proyecto y con un tiempo limitado para su documentación en esta tesis, se definió un parámetro cualitativo para medir la calidad de la reconstrucción del modelo anatómico del ojo. Este parámetro es la curvatura (CUR) definida como se presenta en la Figura 7.22. Los punto

de los esqueletos reconstruidos se utilizaron para aproximar una esfera en el espacio. Conociendo el radio de la esfera que mejor se ajusta a los puntos reconstruidos, por el criterio de mínimos cuadrados, y la distancia euclidiana entre dos puntos de las estructuras es posible calcular la razón de longitud correspondiente. Por otra parte, las mediciones respectivas se pueden realizar directamente del modelo y calcular la razón de longitud real. Posteriormente, se estimó el error relativo de este parámetro y resultó en un 3.79%. Dicho error nos parece ser aceptable y de suma importancia pues en los trabajos más recientes publicados no se presenta una métrica análoga a la descritas anteriormente.

Finalmente, en las reconstrucciones de la retina *in vivo*, nos vemos obligados a recurrir a un criterio meramente cualitativo para determinar la calidad en la reconstrucción, pues no podemos realizar mediciones dentro del globo ocular. Nuevamente se eligieron como las mejores reconstrucciones a las que sus puntos describían una superficie esférica o pseudo-esférica y que la vista paralela al plano XY lucía lo más parecido posible a la segmentación de los árboles seleccionados que presentaron estructuras. Pueden compararse las Figuras 7.17(a) obtenidas con la calibración del *Calib-tool* y 7.21(a) obtenida por auto-calibración (Kruppa), con los vasos indicados en la Figura 7.16.

7.5. Estructuras representadas con volúmenes

Aplicando el algoritmo de cilindros generalizados (véase sección 6.3) a los esqueletos obtenidos por auto-calibración con el modelo anatómico y la retina, obtenemos la representación con superficies de las estructuras en 3D. El diámetro de los cilindros es proporcional a las mediciones hechas por *RISA* en las imágenes retinianas. No es posible aplicar directamente los valores medidos pues, como ya se mencionó anteriormente, la información de escala se pierde en el proceso de reconstrucción y no es posible recuperarlo, es por eso que sólo se respeta la proporcionalidad en el tamaño de los vasos. En la Figuras 7.23 y 7.24 se presentan diferentes vistas de la visualización de las superficies que describen a los vasos reconstruidos del modelo y de la retina, respectivamente.

Habiendo presentado los resultados en reconstrucción bajo los diferentes criterios y consideraciones, solo nos resta presentar nuestras conclusiones en el siguiente capítulo. Finalmente presentaremos, las tareas que a nuestra consideración quedan fuera de los alcances de este trabajo de tesis y por lo tanto, se presenta como trabajo futuro.

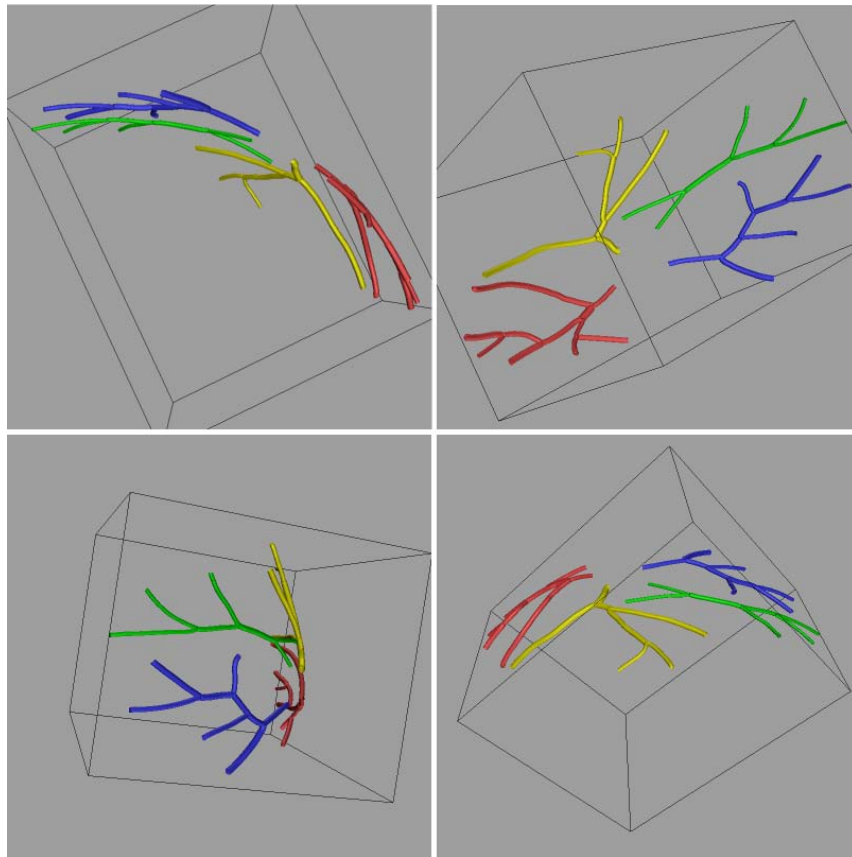


Figura 7.23: Representación de superficies de las estructuras reconstruidas para el modelo anatómico del ojo.

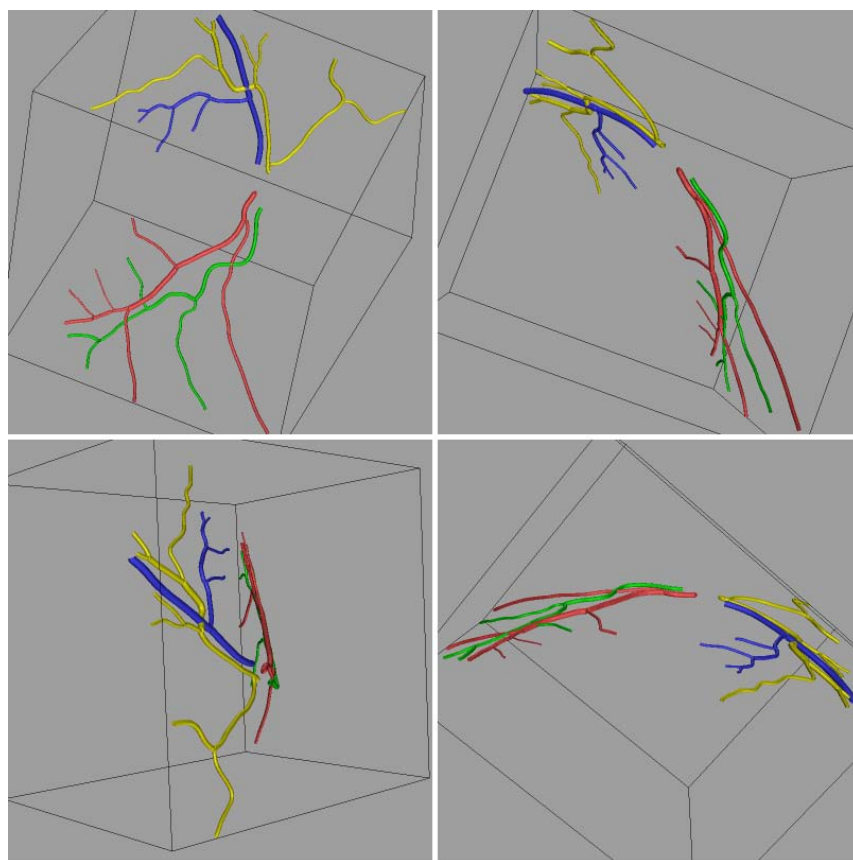


Figura 7.24: Representación de superficies de las estructuras reconstruidas para las imágenes de fondo de ojo.

Capítulo 8

Conclusiones y trabajo futuro

En este trabajo de tesis se presentan reconstrucciones por geometría epipolar de estructuras en imágenes de retina. Adicionalmente se realizan reconstrucciones de escenas artificiales propicias para tomar medidas cuantitativas que nos permitan identificar la reconstrucción de mejor calidad y por ende, el método más preciso y robusto. El núcleo del método de reconstrucción es la calibración de la cámara y del sistema *cámara-ojo*. Por la naturaleza del problema proponemos dos métodos para considerar implícitamente el efecto de la añadidura del sistema óptico del ojo en los valores de la calibración clásica. Por una parte, realizamos la calibración de la cámara *fundus* independientemente y se minimizan las aberraciones del ojo tomando consideraciones en la toma de las imágenes de retina, con el objetivo de suponer que son despreciables. Por otra parte, empleamos un método de auto-calibración que en teoría engloba todo el sistema óptico del sistema de adquisición en los valores de la matriz de calibración que estima. Finalmente decidimos qué método nos permite recuperar con un menor índice de error las propiedades métricas o euclidianas de la escena, en base a los errores relativos calculados y al análisis cualitativo de las reconstrucciones de la rejilla de puntos posicionada en dos planos ortogonales (véase sección 5.5). Después de realizar los resultados de las pruebas realizadas presentamos nuestras conclusiones en la sección siguiente.

8.1. Conclusiones

En este trabajo de tesis se presento una primera aproximación al modelado de aberraciones en los sistemas ópticos, dando a conocer las aberraciones primarias y los métodos de aberrometría más utilizados; sin embargo, la implementación de uno de ellos para compensar de las aberraciones del ojo queda fuera del alcance de esta tesis. Los defectos de desenfoque o patologías refractivas fueron descritos y analizados, permitiendonos concluir que son com-

pletamente corregidos por la cámara *fundus* al ajustar el foco y observar nitidez las imágenes de la retina.

En cuanto a la adición de la óptica del ojo al sistema de lentes de la cámara, se demostró que ocasiona modificaciones en el modelo inicial de la cámara. Se observa que la distancia focal y el punto principal cambian (véase Secciones 7.2 y 7.3) cuando se auto-calibra la escena del modelo anatómico y la retina real. Es importa decir, que tal como se esperaba los cambios en los valores son pequeños, un promedio del 3.7482° de variación con respecto a los parámetros estimados en el modelo anatómico, lo que se puede esperar cuando el cristalino esta relajado (sin acomodación) y con las aberraciones minimizadas pues se restringe experimentalmente la entrada al sistema óptico de rayos paraxiales.

Los resultados de reconstrucción de la rejilla de puntos en una superficie cilíndrica demostraron que con calibración clásica o auto-calibración se obtienen estructuras con errores relativos del 2.2213° y 2.4022° , respectivamente (véase Tablas 7.1 y 7.2). Los errores anteriores nos indican que la diferencia no es significativa pues no se añaden componentes nuevos al sistema global de lentes. En cuanto a las matrices de calibración obtenidas por calibración clásica y por auto-calibración, para las escenas del cilindro y del modelo anatómico del ojo, resultaron mínimamente diferentes en sus valores. Lo cual es congruente con lo que nos sugiere la naturaleza del experimento, es decir, al no presentar cambios en la configuración del sistema de lentes y el sensor del sistema de adquisición, es de esperarse que la solución en la que converge el método de auto-calibración debe ser la misma que la obtenida mediante la estimada por el método clásico, o en su defecto, muy similar. Sin embargo por cuestiones numéricas es mejor calibrar clásicamente sí uno cuenta con el sistema de adquisición y sí no se añadirá un elemento desconocido al sistema óptico de la cámara o realizaran cambios en la distancia focal de la misma. Es importante mencionar que por cuestiones del alcance del trabajo no se incluyo una métrica estadística para estimar un grado de similitud en los parámetros calculados.

Respecto a la calibración del sistema *cámara-ojo*, aunque las matrices obtenidas por los dos métodos propuestos presentan cambios notorios en los parámetros de la distancia focal y la razón de aspecto, los resultados en reconstrucción son muy parecidos en calidad. Sin embargo, se puede apreciar en las reconstrucciones que la morfología de los vasos se recupera mejor usando la matriz obtenida por las ecuaciones de Kruppa (auto-calibración), con movimientos generales en la posición de la cámara. Podemos deducir que las variaciones en los parámetros internos se deben a la presencia de la óptica del ojo, que a pesar de presentar mínimas aberraciones aun afecta al modelos de la cámara. Estas modificaciones son consideradas por el método de auto-calibración, dandonos como resultado reconstrucciones

congruentes con la escena real. En base a lo anterior, elegimos a las reconstrucciones usando auto-calibración como las mejores de todas las obtenidas a lo largo de las pruebas con el modelo y la retina.

Es clave contar con imágenes de la escena a reconstruir, en las cuales la pose de la cámara cambie con rotaciones y traslaciones importantes. Cumpliendo con esta condición, se les provee a los procesos de auto-calibración y recuperación de pose relativa con la mayor información de perspectiva posible. Las ambigüedades en las soluciones estimadas por los procesos anteriores son remediadas con esta consideración.

La curvatura de las superficies sobre las cuales descansaban los puntos reconstruidos (refiriendonos al parámetro cualitativo SAI, véase la Figura 5.3), fue recuperada con una precisión satisfactoria del 2.2826 % de error relativo promedio en los parámetros cuantitativos de la rejilla cilíndrica de puntos, usando auto-calibración con las ecuaciones de Kruppa y la inicialización genérica de los parámetros internos. Aunque es sumamente sensible al ruido en las correspondencias y en específico del error introducido por la discretización de los puntos del esqueleto debido a la digitalización. Es decir, aunque las correspondencias sean correctas es altamente recomendable que el sistema de que extrae los puntos característicos cuente con precisión sub-píxel, ya que el manejar coordenadas enteras introduce perturbaciones severas en la reconstrucción de las estructuras.

El método de segmentación, el de supresión del distorsión y el cálculo de correspondencias funcionan de manera robusta y parecen ser suficientes para esta aplicación en específico. La triangulación lineal recupera la posición 3D de los puntos de manera precisa y correcta (sí se cuenta con los parámetros intrínsecos y extrínsecos correctos) por lo que no consideramos necesario recurrir a métodos más sofisticados de triangulación para esta tarea.

Los experimentos para reconstruir el modelo anatómico del ojo fungieron como una herramienta de depuración para el sistema completo. El refinamiento del sistema se realizó mayormente en esta escena pues contamos con el conocimiento *a priori* de que los puntos reconstruidos deben encontrarse en una superficie regular y de forma semi-esférica. A diferencia del modelo anatómico, en las imágenes de retina no se tiene la seguridad de que sea completamente regular la superficie donde se encuentran las estructuras a reconstruir. Finalmente, los análisis teóricos y experimentales desarrollados en esta tesis, permitieron obtener soluciones al problema de la inestabilidad numérica en las reconstrucciones por el método de Kruppa, así como otras ambigüedades en las soluciones de la configuración de cámaras en geometría epipolar. Las consideraciones de ajustar el foco, el diafragma (para minimizar aberraciones) y cambiar la pose de la cámara en posiciones generales (para redu-

cir el espacio de soluciones de Kruppa y evitar las ambigüedades proyectivas), a pesar de ser muy simples, permiten obtener reconstrucciones que demuestran ser de mejor calidad en terminos cuantitativos o cualitativos (según la escena) que los registrados en el último artículo publicado en congreso internacional [Liu *et al.*, 2008], el cual es uno de los trabajos en los que esta basada esta tesis.

8.2. Trabajo futuro

La identificación de las configuraciones críticas de cámaras es la principal tarea pendiente. Definir métricas para saber si la rotación y traslación bajo la cual cambiamos la pose de la cámara es realmente suficiente para resolver el problemas de las ambigüedades de las ecuaciones de Kruppa y las soluciones erróneas en la interpretación de la quiralidad (*chirality*) en la recuperación de la pose de la cámara.

La robustez en el proceso de la reconstrucción puede aumentarse si se aplica un sistema que calcule la posición de los puntos significativos de la escena con precisión sub-píxel, pues manejar las coordenadas de dichos puntos con cantidades enteras produce deformaciones en las estructuras reconstruidas. El sistema *RISA* provee los puntos característicos en las imágenes de retina, sin embargo, maneja coordenadas enteras (píxeles) que para las tareas que fue diseñado es suficiente, pero en nuestro caso requerimos mayor precisión en el cálculo de la posición de estos puntos.

Será necesario incluir un análisis estadístico de los parámetros cuantitativos, como la prueba de Wilcoxon, para saber que tan parecidos son los resultados obtenidos con diferentes combinaciones de métodos.

En cuanto a la calidad de las estructuras retinales reconstruidas, un primer acercamiento fue la reconstrucción del modelo anatómico del ojo en el cual es posible realizar mediciones directas. Con el parámetro de curvatura que se definió se estimo un error de 3.79 %, que para nosotros es aceptable. Será necesario definir más parámetros mensurables en las estructuras reconstruidas del modelo que nos indiquen con mayor precisión, la calidad con el cual recuperamos la forma de los vasos sanguíneos.

Una mejora a la metodología elegida para medir la precisión en la reconstrucción del cilindro de las secciones 7.2 y 7.3, es medir la curvatura de un modo analítico. A los puntos de toda una fila de la rejilla de puntos se les puede ajustar una curva tridimensional, obteniendo el polinomio que la caracteriza y calculando la curvatura del mismo usando una métrica de geometría diferencial. Esta medición sería mucho mas robusta que la presentada hasta ahora,

pues no se trata de una medición local.

Bibliografía

- Aguilar, W. (2006). *Reconocimiento de objetos basado en la correspondencia estructural de características locales*. Tesis de Maestría, Universidad Nacional Autónoma de México, Ciudad Universitaria, México, D.F.
- Aguilar, W., Martínez-Pérez, M. E., Frauel, Y., Escolano, F., Lozano, M. A., y Espinosa-Romero, A. (2007). Graph-based methods for retinal mosaicing and vascular characterization. En F. Escolano y M. Vento, editores, *Graph-Based Representations in Pattern Recognition*, volumen 4538 de *Lecture Notes in Computer Science*, págs 25–36. Springer Berlin / Heidelberg.
- Asari, K. V., Kumar, S., y Radhakrishnan, D. (1999). A new approach for nonlinear distortion correction in endoscopic images based on least squares estimation. *IEEE Transactions in Medical Imaging*, **18**(4), 345–354.
- Ballard, D. H. y Brown, C. M. (1982). *Computer vision*. Prentice Hall, Inc., Englewood Cliffs, New Jersey 07632.
- Beis, J. S. y Lowe, D. G. (1997). Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. En *IEEE Conference on Computer Vision and Pattern Recognition*, págs 1000–1006. IEEE Computer Society.
- Bentley, J. (1975). Multidimensional binary search trees used for associative searching. *Communications of the Association for Computing Machinery*, **18**(9), 509–517.
- Bonafonte, S. y Bonafonte, E. (2006). *Esquemas clínico-visuales en oftalmología*. Elsevier, España, tercera edición.
- Bouguet, J.-Y. (2008). Camera calibration toolbox for matlab. Disponible en www.vision.caltech.edu/bouguetj/calib_doc/.

- Chanwimaluang, T. y Fan, G. (2007). Constrained optimization for retinal curvature estimation using an affine camera. En *IEEE Conference on Computer Vision and Pattern Recognition*, págs 1–8, Minneapolis, MN.
- Deguchi, K., Mizutani, K., Kawamata, D., Hontani, H., y Wakabayashi, K. (1998). 3d fundus shape reconstruction and display from stereo fundus images. *The Institute of Electronics, Information and Communication Engineers, Transactions on Information and Systems*, **83**(7), 1408–1414.
- Deguchi, K., Hontani, H., y Noami, J. (2000). 3d fundus pattern reconstruction and display from multiple fundus images. En *15th International Conference on Pattern Recognition (ICPR'00)*, volumen 4, págs 4094–4097, Los Alamitos, CA, USA. IEEE Computer Society.
- Díaz Uribe, R. (2009). *Curso de óptica en la oftalmología*. Universidad Nacional Autónoma de México.
- Espinosa-Romero, A. y Martínez-Pérez, M. E. (2005). Optical 3d reconstruction of retinal blood vessels from a sequence of views. En *Eighth International Symposium on Laser Metrology*, volumen 5776 de *Society of Photographic Instrumentation Engineers*, págs 605–612.
- Faugeras, O. D., Luong, Q.-T., y Maybank, S. J. (1992). Camera self-calibration: Theory and experiments. En G. Sandini, editor, *European Conference on Computer Vision*, volumen 588 de *Lecture Notes in Computer Science*, págs 321–334. Springer.
- Fischler, M. A. y Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the Association for Computing Machinery*, **24**(6), 381–395.
- Harris, C. y Stephens, M. (1988). A combined corner and edge detector. En *Proceedings of the 4th Alvey Vision Conference*, págs 147–151. The Plessey Company.
- Hartley, R. I. (1992). Estimation of relative camera positions for uncalibrated cameras. En G. Sandini, editor, *European Conference on Computer Vision*, volumen 588 de *Lecture Notes in Computer Science*, págs 579–587. Springer.
- Hartley, R. I. (1997). In defense of the eight-point algorithm. *IEEE Computer Society Transactions on Pattern Analysis and Machine Intelligence*, **19**(6), 580–593.
- Hartley, R. I. y Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, segunda edición.

- Heikkila, J. y Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. En *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pags 1106–1112. IEEE Computer Society.
- Helferty, J. P., Zhang, C., McLennan, G., y Higgins, W. E. (2001). Videoendoscopic distortion correction and its application to virtual guidance of endoscopy. *IEEE Transactions on Medical Imaging*, **20**(7), 605–617.
- Huang, T. S. y Faugeras, O. D. (1989). Some properties of the e matrix in two-view motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(12), 1310–1312.
- Kanatani, K. (1993). *Geometric computation for machine vision*. Oxford University Press, Inc., New York, NY, USA.
- Kingslake, R. (2010). *Lens design fundamentals*. Academic Press, segunda edición.
- Little, J. y Moler, C. (1984-2009). Matlab r2007a, the mathworks company. Información disponible en www.mathworks.com/products/matlab/.
- Liu, D., Wood, N. B., Xu, X. Y., Witt, N., Hughes, A. D., y SAMcG, T. (2008). *3D Reconstruction of the Retinal Arterial Tree Using Subject-Specific Fundus Images*, volumen 13 de *Computational Methods in Applied Sciences*, chapter 11, pags 187–201. Springer Netherlands, London, UK, first edición.
- Lourakis, M. I. A., Lourakis, M. I. A., Deriche, R., Deriche, R., y Robotvis, P. (1999). Camera self-calibration using the singular value decomposition of the fundamental matrix: From point correspondences to 3d measurements. En *Asian Conference on Computer Vision*, pags 403–408.
- Lowe, D. (1999). Object recognition from local scale-invariant features. En *International Conference on Computer Vision*, pags 1150–1157.
- Ma, Y., Soatto, S., Kosecká, J., y Sastry, S. S. (2005). *An invitation to 3D vision, from images to geometric models*, volumen 26 de *Interdisciplinary applied mathematics*. Springer.
- Marcos, S. (2005). Calidad óptica del ojo. *Investigación y ciencia*, (345), 66–74.
- Martinez-Perez, M., Hughes, A. D., Stanton, A. V., Thom, S. A., Chapman, N., Bharath, A. A., y Parker, K. H. (2002). Retinal vascular tree morphology: a semi-automatic quantification. *IEEE Transactions on Biomedical Engineering*, **49**, 912–917.

- Martinez-Perez, M. E. y Espinosa-Romero, A. (2004). 3d reconstruction of retinal blood vessels from two views. En *Indian Conference on Computer Vision, Graphics and Image Processing*, pages 258–263.
- Martinez-Perez, M. E., Hughes, A. D., Thom, S. A., Bharath, A. A., y Parker, K. H. (2007). Segmentation of blood vessels from red-free and fluorescein retinal images. *Medical Image Analysis*, **11**(1), 47–61.
- Mendonça, P. R. S. y Cipolla, R. (1999). A simple technique for self-calibration. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **1**, 1500.
- Moore, E. (1920). On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society*, **26**, 394–395.
- Moré, J. J. (1977). The Levenberg-Marquardt algorithm: Implementation and theory. En G. A. Watson, editor, *Numerical Analysis*, pages 105–116. Springer, Berlin.
- Otsu, N. (1979). A threshold selection method from gray level histograms. *IEEE Transactions Systems, Man and Cybernetics*, **9**, 62–66.
- Patton, N., MacGillivray, T., Aslam, T., Pattie, A., Deary, I., y Dhillon, B. (2005). Retinal vascular image analysis as a potential screening tool for cerebrovascular disease: a rationale based on homology between cerebral and retinal microvasculatures. *Journal of Anatomy*, **206**, 319–348.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., y Flannery, B. P. (1995). *Numerical Recipes in C*. Cambridge University Press, segunda edición.
- Rajendra Acharya, U., Ng, E. Y., y Suri, J. S. (2008). *Image Modeling of the Human Eye*. Artech House, ISBN: 9781596932081.
- Rousseeuw, P. J. y Leroy, A. M. (1987). *Robust regression and outliers detection*. John Wiley and Sons, New York.
- Tsai, R. Y. (1987). A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, **3**(4), 323–344.
- Zhang, Z. (1999). Flexible camera calibration by viewing a plane from unknown orientations. *IEEE International Conference on Computer Vision*, **1**, 666–673.