



BIBLIOTECA
FAC. DE INFORMÁTICA
U.N.L.P.



Tesina de Grado:
Sistema de procesamiento de imágenes aplicado al entrenamiento deportivo
Facultad de Informática – UNLP

Federico Cristina

Sebastián Dapoto

Director: Ms. Claudia Russo

Agosto de 2007

<p>TES 07/15 DIF-03103 SALA</p>	<p>UNIVERSIDAD NACIONAL DE LA PLATA FACULTAD DE INFORMÁTICA Biblioteca 50 y 120 La Plata catalogo.info.unlp.edu.ar biblioteca@info.unlp.edu.ar</p> <p> DIF-03103</p>
---	--



BIBLIOTECA
FAC. DE INFORMÁTICA
U.N.L.P.

DONACION... FACULTAD
\$
Fecha... 12-3-08
Inv. E.....Inv. B... 003103

TES
02/15
003103

Indice

Introducción	1
La tecnología y el deporte	1
Prototipo: objetivo y complejidad	2
Estructura de la tesina	3
Capítulo I. Dispositivos de captura	5
1. Introducción	5
2. Breve reseña histórica	5
3. Conceptos y fundamentos	7
3.1. Distancia focal	9
3.2. Tiempo de exposición	11
3.3. Apertura del diafragma	12
3.4. Sensibilidad (ISO)	14
3.5. Determinando el balance adecuado	15
3.6. Resolución y tecnología de captura	16
4. Interfaces de comunicación	19
5. Conclusiones	20
Capítulo II. Técnicas de procesamiento	21
1. Introducción	21
2. Filtrado por umbralización	21
3. Etiquetado	24
3.1. Vecindad y conectividad	24
3.2. Etiquetado de componentes	25
4. Asociación por bloques (block matching)	27
4.1. Búsqueda Completa	28
4.2. Algoritmos Basados en Firma	28
4.3. Algoritmos de Ventana de Búsqueda Dinámica	29
4.4. Algoritmo de Búsqueda de Tres Etapas	29
4.5. Algoritmo de Búsqueda Binaria	30
5. Flujo Óptico	31
5.1. Teoría del Flujo Óptico	32
5.2. Problemas del Flujo Óptico	34
5.2.1. Problemas de iluminación de la escena	34
5.2.2. Problema de la apertura	35
5.2.3. Problema de la oclusión	36
Capítulo III: Técnicas estereoscópicas	37
1. Introducción	37
2. Fotografía digital	37
2.1. Imágenes de intensidad	37
2.1.1. Óptica	39
2.1.1.1. Foco de la imagen	40
2.1.1.2. Lentes delgados	40

2.1.1.3. Campo de visión	42
2.1.2. Radiometría	43
2.1.3. Geometría	43
2.1.3.1. La Cámara de Perspectiva	43
2.1.3.2. La Cámara de Perspectiva Débil	44
2.2. Adquisición de imágenes digitales	45
2.3. Parámetros de Cámara	46
2.3.1 Parámetros Extrínsecos	46
2.3.2 Parámetros Intrínsecos	47
2.4. Modelo de Cámara	48
3. Calibración de cámara	51
4. Estereoscopía	53
4.1. Introducción	53
4.2 Geometría epipolar	55
5. Conclusiones	59
Capítulo IV. La interfaz gráfica de usuario	60
1. Introducción	60
2. Definición	60
3. La Interacción Hombre-Computadora	61
4. Aspectos de las GUI	63
4.1. Ejemplo 1	63
4.2. Ejemplo 2	64
4.3. Ejemplo 3	65
4.4. Ejemplo 4	65
5. Usabilidad	66
5.1. Principios generales de la usabilidad	67
5.1.1. Facilidad de aprendizaje	67
5.1.2. Consistencia	67
5.1.3. Flexibilidad	68
5.1.3.1. Control del usuario	68
5.1.3.2. Migración de tareas	68
5.1.3.3. Capacidad de sustitución	68
5.1.3.4. Adaptabilidad	68
5.1.3.5. Técnicas para lograr flexibilidad	69
5.2. Otros aspectos a tener en cuenta	69
6. Metáforas	69
7. Recomendaciones	71
8. Conclusiones	73
Capítulo V: Análisis preliminar	75
1. Introducción	75
2. Ubicación de las cámaras	76
3. Distancia focal de las lentes	80
4. Utilizando cámaras	82
4.1. Resolución	83
4.2. Distancia focal	84
4.3. Velocidad de captura	85
5. Requerimientos de las cámaras	86
6. Interfaz de comunicación y formato de captura	87

7. Arquitectura de captura	88
8. Librería de procesamiento de imágenes	90
9. Conclusiones	92
Capítulo VI: Módulo de calibración de cámaras	93
1. Introducción	93
2. Toolbox de calibración para matlab	93
3. Primera aproximación: 'StereoChecker'	96
4. Módulo de calibración de cámaras	99
5. Conclusiones	103
Capítulo VII: Determinación de la trayectoria de la pelota	104
1. Introducción	104
2. Estimación de la trayectoria	104
3. Módulo de filtrado de la pelota	106
4. Procesamiento de un tiro	108
5. Aspectos técnicos	112
6. Conclusiones	113
Capítulo VIII: Determinación de la velocidad de rotación de la pelota	116
1. Introducción	116
2. Procesamiento de un tiro	116
3. Análisis preliminar	117
4. Marcación de la pelota	119
5. Filtrado	120
6. Seguimiento de la marcas	122
7. Modelo computacional	124
7.1. Marca central definida	124
7.2. Marca central no definida	127
7.3. Realizando el cálculo	128
8. Conclusiones	129
Capítulo IX: Entorno de visualización de tiros	130
1. Introducción	130
2. La GUI principal	130
3. Componentes funcionales principales	133
4. Selección de vista	134
5. Reproducción de videos	135
5.1. Información gráfica a renderizar	135
5.2. La información numérica	136
6. Adiciones funcionales que facilitan el uso de la aplicación	137
7. Aspectos técnicos	139
7.1. Almacenamiento y carga de tiros	139
7.2. Renderización sobre vistas y videos	143
7.3. Generación del video de seguimiento	144
8. Conclusiones	144
Capítulo X: Pruebas realizadas	146
1. Introducción	146
2. Simulaciones	147

2.1. Disposición de las cámaras	147
2.2. Grilla de calibración	148
2.3. Iluminación	150
3. Datos obtenidos	150
4. Tiempos de procesamiento	152
5. Conclusiones	154
Conclusiones	155
1. Introducción	155
2. El prototipo	155
3. OpenCV	156
4. Ampliaciones	157
5. A futuro	159
Bibliografía	161

Introducción

1. La tecnología y el deporte

Actualmente, en toda disciplina deportiva, las sesiones de entrenamiento son necesarias para perfeccionar las técnicas de los jugadores, las cuales requieren de más y mejor tecnología.

El uso de sistemas automatizados de aprendizaje, más específicamente sistemas informáticos, permite a los deportistas incrementar considerablemente la curva de aprendizaje, y a los entrenadores aumentar las capacidades didácticas en las sesiones de entrenamiento.

La tecnología permite obtener información adicional más allá de los datos que pueden ser recavados mediante observación visual. En el caso del entrenamiento futbolístico por ejemplo, gracias a la filmación y reproducción en cámara lenta de los tiros libres, jugador y entrenador pueden observar y debatir la calidad y efectividad de los mismos, logrando así un aprendizaje a lo largo del tiempo.

Pero más y mejor información - y lo que es más importante, información precisa - puede obtenerse mediante la captura digital y procesamiento de los tiros libres. Realizando ciertas adaptaciones al entorno y a la pelota, es posible triangular la posición de ésta a lo largo de la trayectoria y calcular además la velocidad de rotación que la misma desarrolla en cada tiro; determinando también la dirección preponderante sobre la que ésta realiza el movimiento. A su vez, el procesamiento digital permite obtener información numérica importante para su posterior análisis.

Entre los datos de interés para su posterior análisis relacionados con los tiros libres, cabe mencionar: la aceleración y el ángulo de elevación iniciales del tiro y su velocidad promedio, la elevación máxima y las revoluciones por segundo que adopta la pelota; además de la posición de la misma a lo largo del recorrido que efectúa hacia la línea de fondo.

Por otro lado, es particularmente interesante poder reproducir gráficamente el tiro desde vistas especiales sin perspectiva (superior, lateral y frontal), en las cuales la interpretación visual de la curva lograda en cada tiro es considerablemente mayor con respecto a las capturas originales.

2. Prototipo¹: objetivo y complejidad

El prototipo desarrollado tiene como meta mejorar la técnica de tiros libres de los jugadores mediante la exposición inmediata de los resultados anteriormente detallados luego de cada intento. Esto permite además calificar cada tiro de acuerdo a la precisión y velocidad lograda; brindando comparativas entre varios intentos. Gracias a esto se podrá contar con un historial de cada jugador con el que será posible notar la evolución de los mismos en cuanto a destreza deportiva.

Resulta particularmente importante la reconstrucción 3D del objeto en movimiento (en este caso, la pelota), ya que no sólo se puede variar el punto de observación sino que se pueden estudiar otros efectos de importancia deportiva (por ejemplo la rotación de la pelota sobre su eje dentro de un trayecto lineal). Esto requiere identificar puntos propios del objeto móvil y reconstruir la trayectoria de giro, simultáneamente con la trayectoria de movimiento.

Este aspecto posee una complejidad particular; ya que debe aislarse adecuadamente el movimiento de traslación que sufre la pelota del de rotación. Se debe tener en cuenta

¹ Los términos *prototipo*, *sistema* y *aplicación* serán usados indistintamente a lo largo del presente trabajo

además el criterio de etiquetado sobre la superficie de la misma con el fin de lograr resultados correctos.

3. Estructura de la tesina

Para llevar a cabo el desarrollo del prototipo; fue necesario seguir una serie de pasos que abarcó la calibración de cámaras, el filtrado de imágenes, la triangulación de puntos en el espacio, la obtención de la trayectoria y rotación de la pelota, y la renderización visual de la información entre otros.

El presente trabajo se divide en dos partes principales. La primera abarca los fundamentos y conceptos que han sido analizados y estudiados, los cuales fueron utilizados como base para el desarrollo de la aplicación. En la segunda parte se realiza un recorrido detallado de la aplicación, partiendo del análisis previo, pasando por la implementación, y llegando hasta los resultados obtenidos en las pruebas realizadas; justificando en cada caso las decisiones de diseño adoptadas.

Federico Cristina
Sebastián H. Dapoto

Parte I

FUNDAMENTOS TEORICOS

Capítulo I. Dispositivos de captura

1. Introducción

De manera introductoria se detallarán a continuación los fundamentos físicos de la fotografía como dispositivo de captura, la cual es base para la captura y posterior procesamiento de video. Los conceptos volcados en este apartado - aunque básicos - tienen por objetivo ubicar al lector en el contexto adecuado, ya que son necesarios para las discusiones que se presentarán en capítulos posteriores.

2. Breve reseña histórica

Los dispositivos de captura tienen por finalidad impactar una imagen del mundo real sobre algún dispositivo fotosensible, tal como película filmica o CCD (Charge-Couple Device). La evolución que ha derivado en los dispositivos actuales requirió años de investigación y desarrollo. A modo de introducción, se realizará una breve reseña de la misma.

Inicialmente, la fotografía se basó en una caja oscura con un orificio en una de las paredes. La luz que penetraba en ella por éste proyectaba una imagen del exterior en la pared opuesta. Aunque la imagen formada resultaba invertida y borrosa, se utilizó esta técnica mucho antes de la invención de la película.

Posteriormente al orificio se le adicionó una lente óptica para conseguir una imagen más clara y definida, reduciendo dramáticamente los tiempos de exposición.

La sensibilidad a la luz de ciertos compuestos de plata, particularmente el nitrato y el cloruro de plata, era ya conocida a finales del siglo XVIII para obtener imágenes fotográficas. Las imágenes de cuadros, siluetas de hojas y perfiles humanos utilizando papel recubierto de cloruro de plata no eran permanentes, ya que después de exponerlas a la luz, toda la superficie del papel se ennegrecía.

Las primeras fotografías, conocidas como heliografías, fueron hechas en 1827 por el físico francés Nicéphore Niépce. Alrededor de 1831 el pintor francés Louis Jacques Mandé Daguerre realizó fotografías en planchas recubiertas con una capa sensible a la luz de yoduro de plata. Con el método de Daguerre se obtenía una imagen única en la plancha de plata por cada exposición.

Alrededor de 1884 el inventor estadounidense George Eastman patentó una película que consistía en una larga tira de papel recubierta con una emulsión sensible. En 1889 realizó la primera película flexible y transparente en forma de tiras de nitrato de celulosa. El invento de la película en rollo marcó el final de la era fotográfica primitiva y el principio de un periodo durante el cual miles de fotógrafos aficionados se interesarían por el nuevo sistema.

En 1907 se pusieron a disposición los primeros materiales comerciales de película en color, unas placas de cristal llamadas Autochromes Lumière en honor a sus creadores, los franceses Auguste y Louis Lumière. En esta época las fotografías en color se tomaban con cámaras de tres exposiciones.

La cámara de 35 mm, que requería película pequeña y que estaba, en un principio, diseñada para el cine, se introdujo en Alemania en 1925. Gracias a su pequeño tamaño y a su bajo coste se hizo popular entre los fotógrafos profesionales y los aficionados.

Con la aparición de la película de color Kodachrome en 1935 y la de Agfacolor en 1936, con las que se conseguían transparencias o diapositivas en color, se generalizó el

uso de la película en color. La película Kodacolor, introducida en 1941, contribuyó a dar impulso a su popularización.

Desde ese entonces hasta la actualidad, la fotografía “tradicional” no ha pegado grandes saltos tecnológicos. Sin embargo, a principios de los años '70, una nueva forma de captura nació: los dispositivos CCD y CMOS.

Los CCD (del inglés Charge Coupled Device) y CMOS (del inglés Complementary Metal Oxide Semiconductor) son sensores fotosensibles basados en unos circuitos integrados que contienen un número determinado de condensadores acoplados. Bajo el control de un circuito interno, cada condensador puede transferir su carga eléctrica a uno o a varios de los condensadores que estén a su lado en el circuito impreso. A lo largo de este trabajo se hará mención a CCD o CMOS sin distinción a menos que especifique lo contrario.

Los primeros dispositivos CCD, pilares fundamentales de la fotografía digital actual, fueron inventados por Willard Boyle y George Smith en 1969 en los Laboratorios Bell. La primer cámara digital fue desarrollada por Kodak en Diciembre de 1975. Ésta tenía el tamaño de una tostadora y una calidad equivalente a 0.01 Megapíxeles. Necesitaba 23 segundos para guardar una fotografía en blanco y negro en una cinta de cassette.

La capacidad de resolución o detalle de la imagen depende del número de células fotoeléctricas del CCD, el cual es expresado en píxeles. Actualmente las cámaras fotográficas digitales incorporan CCDs con capacidades de hasta 160 megapíxeles.

3. Conceptos y fundamentos

Las componentes básicas de las cámaras fotográficas y de video actuales pueden observarse en la figura 1.1:

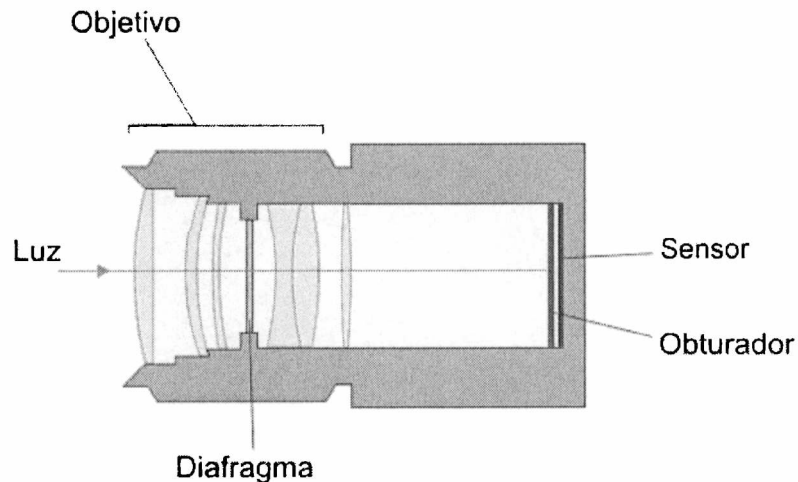


Figura 1.1 – Componentes principales de los dispositivos de captura

El objetivo es el conjunto de dispositivos ópticos o lentes de la cámara que permiten trasladar la luz de manera ordenada hacia el interior de la cámara hasta el dispositivo CCD.

El diafragma de la cámara se encarga de permitir el paso de una cantidad de luz determinada según sean los requerimientos de captura. Funciona como el iris del ojo humano, abriéndose o cerrándose para permitir que entre más o menos luz.

El propósito del obturador es regular la extensión de tiempo que la luz penetra en la cámara durante una exposición. Al realizar una captura, el obturador se abre, permitiendo el impacto de la luz en el dispositivo fotosensible.

Por último, el sensor CCD recibe la información lumínica, la cual es traducida a información eléctrica.

Existen ciertas nociones relacionadas con los componentes recién mencionados, que son comunes tanto para cámaras fotográficas digitales como para cámaras de video. De acuerdo a las características de éstos, se determina la capacidad de captura del dispositivo en cuestión. Principalmente se pueden nombrar:

- Distancia focal de la lente
- Tiempo de exposición
- Apertura del diafragma
- Sensibilidad (ISO)
- Resolución y tecnología de captura

La variación de estos parámetros incide directamente en los resultados obtenidos y cada uno de éstos incide además en el resto de los parámetros. Por ejemplo: a mayor apertura de diafragma, menor requerimiento de tiempo de exposición, pero menor profundidad de campo. A mayor sensibilidad, menor requerimiento de apertura y tiempo de exposición, pero mayor ruido resultante. A continuación, se explicará los resultados que presenta el cambio de estos parámetros en la captura de imágenes y video.

3.1. Distancia focal

La distancia focal de la lente es la que separa a ésta de la película o CCD. Un objetivo normal para una cámara de 35 mm tiene una longitud focal cercana a los 50 mm. Un objetivo de distancia focal larga podría tener alrededor de 100 mm (teleobjetivo).

Cuando se emplea un objetivo de distancia focal larga no solamente aumenta la distancia de éste a la película sino también el tamaño de la imagen. El tamaño de la imagen es proporcional a la longitud focal. Es decir, una imagen formada por un objetivo de 100 mm es el doble que la de uno de 50 mm. Entonces un objetivo de distancia focal larga ejerce un efecto de acercamiento. Como la imagen formada por el objetivo de 100 mm es más grande entrará una parte menor de la misma en la película. Por el contrario un objetivo de 28 mm (gran angular) de longitud focal formará a partir del mismo objeto una imagen de aproximadamente la mitad de tamaño que uno de 50 mm.

En definitiva la longitud focal determina, entre otras cosas, el campo de visión y el nivel de detalle posibles a observar en la captura. La figura 1.2 muestra la misma escena capturada con lentes de variadas distancias focales, desde gran angular hasta teleobjetivo:

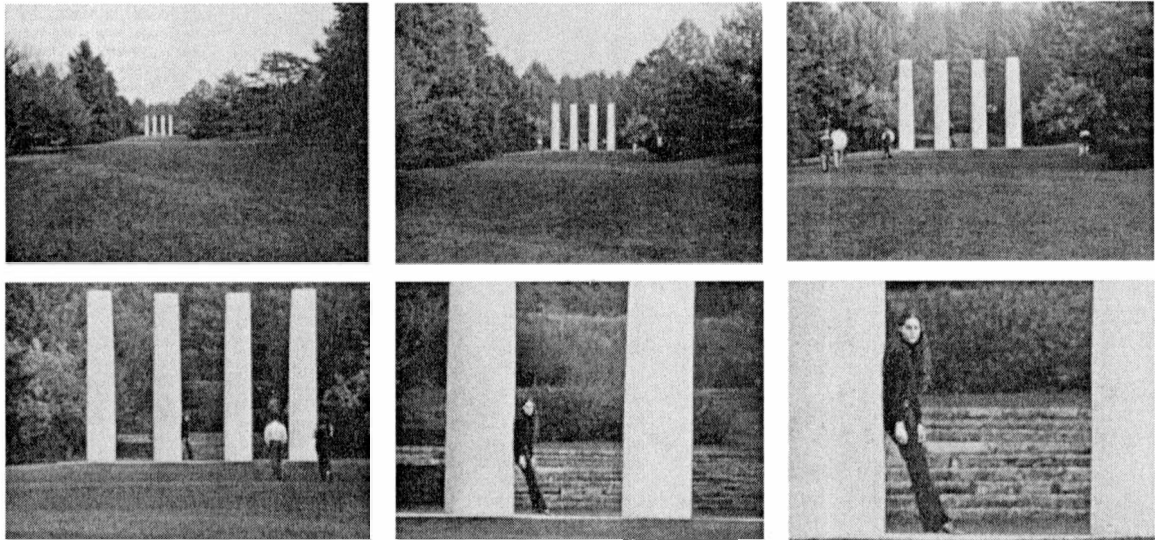


Figura 1.2 – Resultado de utilizar diferentes distancias focales

En las imágenes en movimiento, la longitud focal juega un papel importante en la imagen final. Una lente con una longitud focal larga magnificará la imagen y su movimiento aparente. El movimiento aparente es más grande cuando el objeto mueve en ángulos rectos a través de la línea de visión, y es menor cuando el objeto mueve directamente hacia la cámara o se aleja de ella.

Contar con un amplio campo de visión es necesario a fin de poder capturar la escena completa, pero esto es útil sólo hasta cierto punto; ya que dependerá principalmente del tamaño de los objetos que deben ser posteriormente procesados. Por ejemplo, en la figura 1.2, sería imposible realizar cualquier tipo de tratamiento digital sobre el rostro de la modelo en las primeras tomas de la serie.

3.2. Tiempo de exposición

El tiempo de exposición es el que determina la duración en la que el obturador permite el paso de luz. La luminosidad de una captura se encuentra altamente condicionada por el tiempo de exposición. La captura de una fotografía en condiciones normales de luz diurna se realiza generalmente a una velocidad de 1/60 de segundo.

Con tiempos de exposición mayores se pueden lograr capturas en condiciones de luz reducidas o nocturnas, en las cuales la duración de la exposición puede llegar a ser de varios segundos, en cuyo caso se requiere de un soporte fijo para la cámara.

Los tiempos de exposición rápidos (1/500 de segundo o más) son necesarios en los casos en que el objetivo a capturar se encuentre en movimiento, a fin de evitar un efecto de borronado en la imagen, conocido como PSF (Point Spread Function) de movimiento. Cabe destacar que a mayor velocidad de exposición, mayor es el requerimiento de lumínico sobre el objetivo, lo cual es un condicionante en ciertos casos.

La figura 1.3 muestra la misma escena en dos tomas diferentes. En la primer toma se puede observar el efecto borronado ya que fue obtenida con una velocidad de captura demasiado lenta para el movimiento que presenta la escena. En la segunda toma fue correctamente configurada la velocidad de exposición.

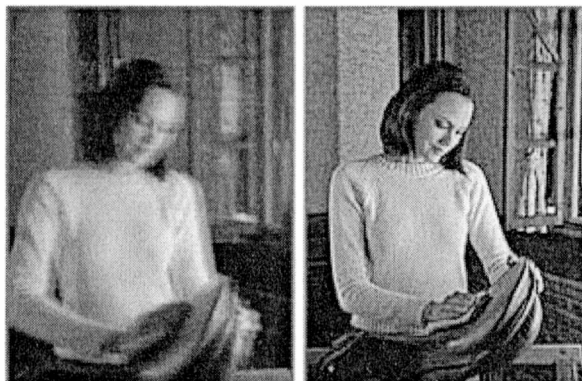


Figura 1.3 – Resultados al variar el tiempo de exposición.

Para el caso específico de capturas deportivas, es común utilizar tiempos de exposición muy breves, del orden de los 1/1000 de segundo o más. Por supuesto que estos valores deben ir acompañados de una fuente lumínica considerable o alta sensibilidad de la película.

Formalmente, para contrarrestar el movimiento y evitar el efecto borroso, se debe usar la exposición más corta posible. Una fórmula para determinar el máximo permitido es:

$$T = L / (500 \times V) \text{ segundos}$$

Donde L es el área del objetivo más grande a ser capturada y V es la velocidad de éste, en la misma unidad de medida que L por segundo.

Por ejemplo, para un auto desplazándose a 102 km/h (300 metros por segundo), la mayor longitud de cuadro de captura (L) puede ser de alrededor de 150 metros. Por lo tanto, en este caso $T = 150 / (500 \times 300) = 1 / 1000$ segundos de obturación.

Cabe destacar que tanto V como L determinan el tiempo de exposición: tanto incrementando V como decrementando L se logran menores tiempo de exposición.

3.3. Apertura del diafragma

Para conseguir el equilibrio de luz en cada exposición se combina el tiempo de exposición con la apertura de diafragma. Es decir, si se aumenta la apertura de diafragma (más luminosidad) hay que reducir el tiempo de exposición (tiempo que la película está expuesta a la luz), en caso contrario la fotografía sería muy clara (sobreexpuesta). Igualmente, para captar la luz que hay en la escena, si se reduce la apertura de diafragma (menos luminosidad) hay que alargar la exposición, o por lo contrario quedaría una imagen oscura (subexpuesta).

La apertura de diafragma se mide en números f . El menor de los números indica la apertura máxima. La relación entre los números f , es que cada vez que se pasa de un número a uno menor, la luminosidad se reduce a la mitad, de modo que $f/11$, por ejemplo, tiene el doble de luminosidad que $f/16$ y la mitad que $f/8$. La figura 1.4 muestra diferentes tomas de la misma escena capturadas con distintas aperturas del diafragma, manteniendo constante el tiempo de exposición en cada toma:

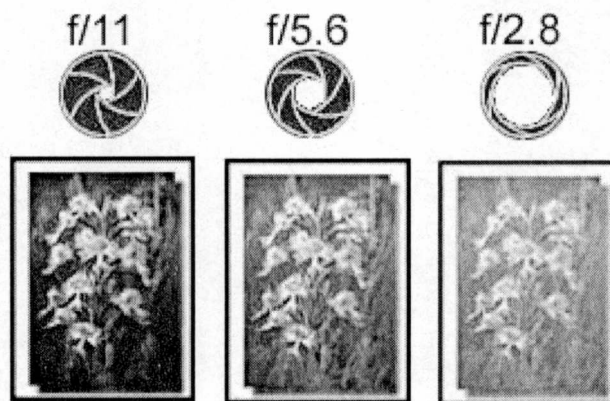


Figura 1.4 – Resultados al variar la apertura del diafragma

La apertura del diafragma, en conjunto con la distancia focal, determinan además la profundidad de campo, que es la zona en la cual la imagen captada por el objetivo es nítida. Como regla general, a mayor distancia focal y apertura, menor profundidad de campo. Puede observarse una corta profundidad de campo (tres renglones del texto) en la figura 1.5.

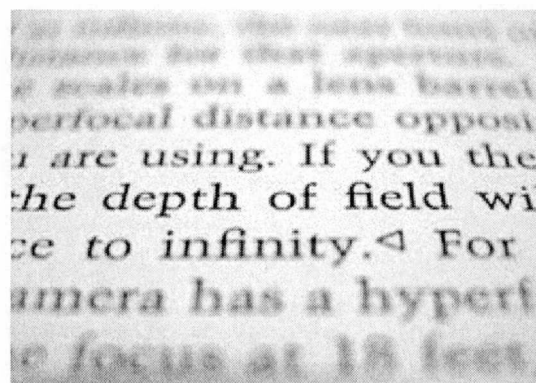


Figura 1.5 – Profundidad de campo

En distancias focales cortas este fenómeno no se presenta, inclusive trabajando con grandes aperturas de diagrama.

3.4. Sensibilidad (ISO)

La sensibilidad de una película fotográfica es la velocidad con la que su emulsión fotosensible reacciona a la luz. A mayor velocidad, menor cantidad de luz será necesaria para realizar la fotografía. Como contra, mayor ruido lumínico será captado en la toma.

El valor ISO es aquel en que a mayor numeración en su escala, mayor velocidad tendrá, y viceversa, a menor numeración menor velocidad. Por ejemplo, si tenemos una película ISO 800 será dos veces más rápida que una película ISO 400 y por lo tanto requerirá la mitad de luz que en el último caso.

Las fotografías que se realizan con un valor ISO alto permiten capturar imágenes en escenas de baja luminosidad con tiempos cortos de exposición, pero pueden presentar el problema de ruido de imagen.

La figura 1.6 muestra la misma escena capturada con película ISO100 e ISO800, en la cual la primera requiere de un tiempo de exposición y apertura de diafragma mayores que en la segunda. La desventaja de esta última es notoria: el ruido aditivo obtenido.

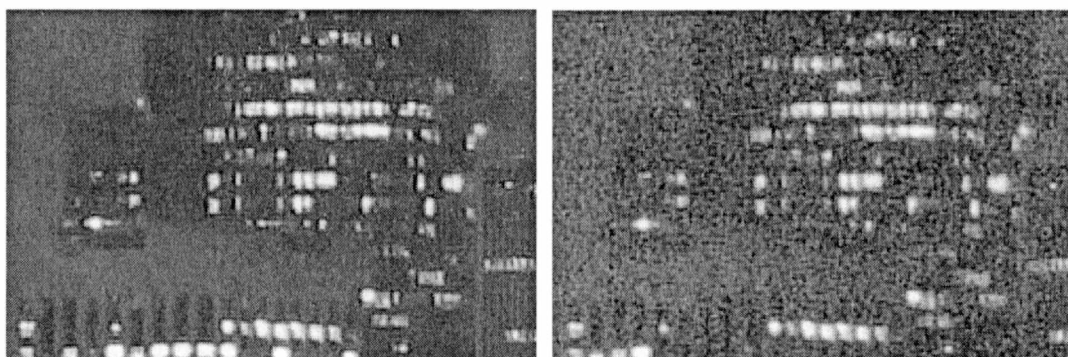


Figura 1.6 – Resultados al variar la sensibilidad

Las cámaras digitales permiten configurar el nivel de sensibilidad del CCD, simulando la utilización de películas de distinto ISO.

3.5. Determinando el balance adecuado

La figura 1.7 presenta el “triángulo de exposición”, el cual combina los tres parámetros que determinan la exposición final de la fotografía: el tiempo de exposición y la apertura afectan la cantidad de luz que llega hasta el sensor óptico de grabación; mientras que la velocidad ISO afecta la sensibilidad del mismo.

Estos tres parámetros deben encontrarse correctamente configurados entre sí con el fin de crear la exposición correcta. En caso de no conseguir una correcta configuración de éstos, la captura lograda puede encontrarse subexpuesta o sobreexpuesta, perdiéndose en ambos casos calidad e información vital para el procesamiento.

La figura 1.8 muestra estos casos: en la primera, la incorrecta combinación de parámetros logró una imagen subexpuesta; mientras que en la segunda una imagen sobreexpuesta. Debajo de cada ejemplo, se muestra el histograma correspondiente.

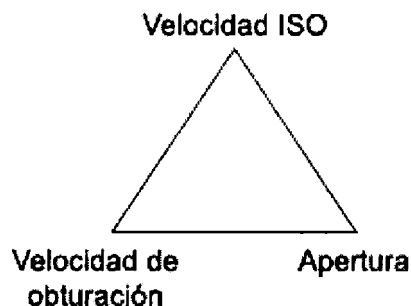


Figura 1.7 – El triángulo de exposición

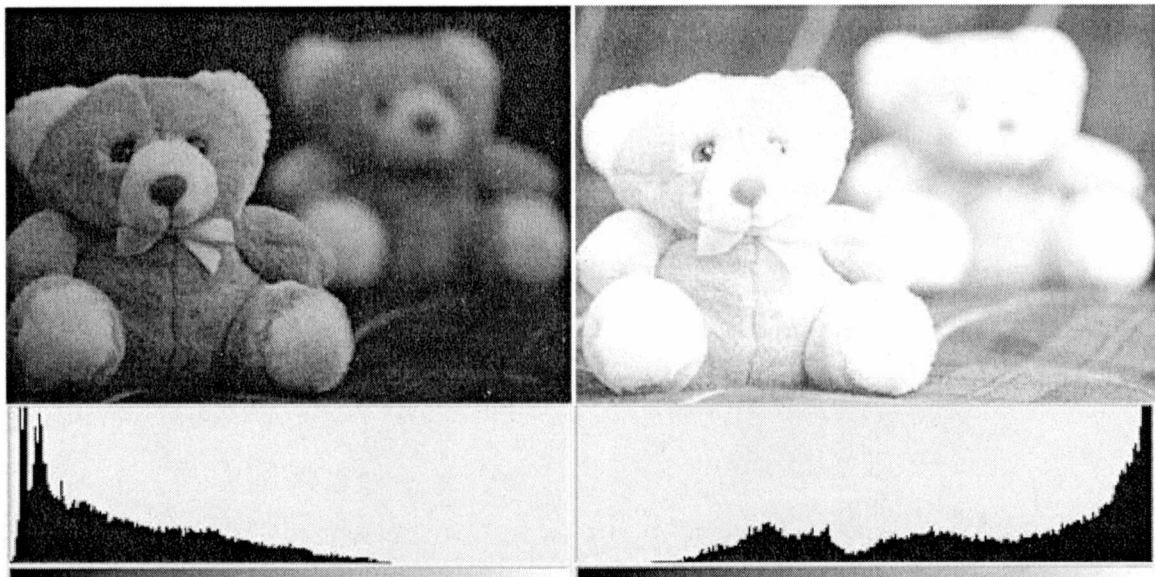


Figura 1.8 – Configuraciones incorrectas del triángulo de exposición

Afortunadamente, los dispositivos de captura actuales permiten automatizar la configuración de manera tal que pueda lograrse en la mayoría de los casos balances correctos de exposición. En condiciones ideales de luminosidad, la mejor opción es configurar un valor ISO lo más bajo posible, minimizando el nivel de ruido; asignar una velocidad de obturación acorde a las necesidades de la toma; y dejando que el dispositivo determine de manera automática el nivel de apertura.

3.6. Resolución y tecnología de captura

Al capturar una imagen mediante una cámara digital, o al digitalizar una secuencia de video, se nos presenta el siguiente problema: la información que fotografiamos tiene que ser discretizada y plasmada en píxeles para que pueda ser representada en una computadora.

Se pierde entonces tanto información espacial (llevamos una imagen de la vida real a una grilla discreta y finita de píxeles) como información sobre la intensidad de cada píxel (llevamos los niveles de brillo y color a, por ejemplo, una escala de 256 niveles de intensidad en una escala de grises). La figura 1.9 muestra un ejemplo de la

discretización espacial y la cuantización lumínica en una parte de la captura (indicado en el recuadro rojo):

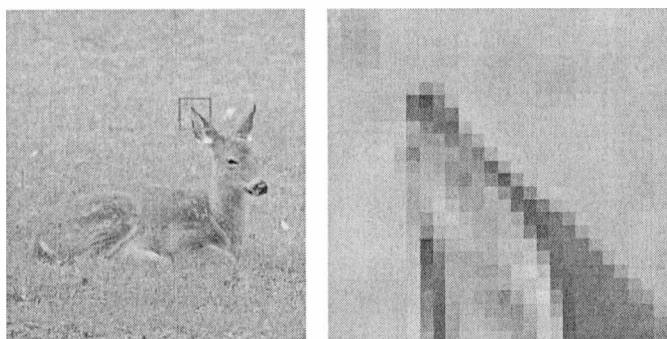


Figura 1.9 – La imagen derecha muestra en detalle una zona de la imagen izquierda.

Aunque actualmente incluso en cámaras fotográficas hogareñas se cuenta con resoluciones y profundidades de color elevadas (matrices mayores a 4000 x 3000 píxeles, con 24 bits por píxel), la tecnología de captura digital es la misma: dispositivos de captura con tres CCDs o con sólo un CCD.

Las cámaras con tres CCDs poseen un prisma que divide la luz en diferentes rangos de longitud de onda: rojo, verde y azul. Cada CCD se encarga de capturar una de estas señales para luego combinar la información, como se muestra en la figura 1.10.

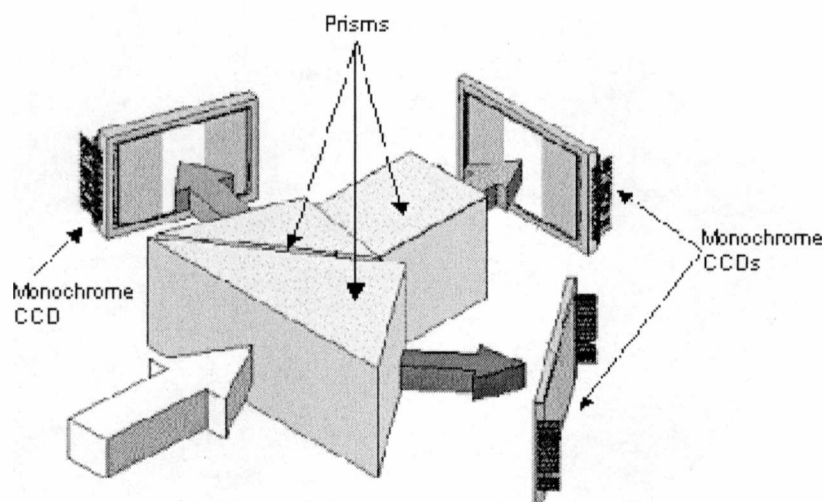


Figura 1.10 – Dispositivos con 3 CCDs.

En las cámaras con un solo CCD, los píxeles que componen el sensor poseen diferentes sensibilidades lumínicas según la longitud de onda, generando lo que se conoce como mosaico Bayer, como se muestra en la figura 1.11.

p_{00}	p_{10}	p_{20}	p_{30}	p_{40}
p_{01}	p_{11}	p_{21}	p_{31}	p_{41}
p_{02}	p_{12}	p_{22}	p_{32}	p_{42}
p_{03}	p_{13}	p_{23}	p_{33}	p_{43}
p_{04}	p_{14}	p_{24}	p_{34}	p_{44}

Figura 1.11 – Mosaico Bayer.

La razón de que se use mayor cantidad de puntos verdes es que el ojo humano es más sensible a ese color. Para obtener la información RGB de cada píxel, se interpolan las muestras verdes, roja, y azul de los píxeles vecinos. Según los requerimientos de tiempo y calidad se utilizan diferentes métodos de interpolación. Uno de los más adoptados es el que se muestra en la figura 1.12 para los píxeles p_{11} , p_{21} , p_{12} y p_{22} ; utilizando como base el mosaico Bayer de la figura 1.11.

$$\begin{aligned}
 p_{11} (R) &= (p_{01}+p_{21}) / 2 \\
 p_{11} (V) &= p_{11} \\
 p_{11} (A) &= (p_{10}+p_{12}) / 2 \\
 \\
 p_{21} (R) &= p_{21} \\
 p_{21} (V) &= (p_{20}+p_{31}+p_{22}+p_{11}) / 4 \\
 p_{21} (A) &= (p_{10}+p_{30}+p_{12}+p_{32}) / 4 \\
 \\
 p_{12} (R) &= (p_{01}+p_{21}+p_{03}+p_{23}) / 4 \\
 p_{12} (V) &= (p_{11}+p_{22}+p_{13}+p_{02}) / 4 \\
 p_{12} (A) &= p_{12} \\
 \\
 p_{22} (R) &= (p_{21}+p_{23}) / 2 \\
 p_{22} (V) &= p_{22} \\
 p_{22} (A) &= (p_{12}+p_{32}) / 2
 \end{aligned}$$

Figura 1.12 – Método de interpolación del mosaico Bayer para la obtención de los canales Rojo (R), Verde (V) y Azul (A).

Por lo tanto, las cámaras con tres CCDs poseen una mayor definición que las de un solo CCD, ya que estas últimas generan imágenes a color mediante técnicas de interpolación. De todas formas, el costo de las primeras es más elevado que el de las últimas.

4. Interfaces de comunicación

Las principales formas de capturar video mediante una computadora son dos, dependiendo del tipo de señal que recibe desde la cámara.

Si la información recibida es analógica (por ejemplo mediante conector RCA o SVHS) es necesario realizar una digitalización mediante alguna placa específica para el esto. Este proceso de captura digital, envío analógico y posterior redigitalización de la señal conlleva una degradación en el video final, ya que con cada conversión realizada se pierde calidad en las imágenes resultantes.

Si la información es digital, se evita la conversión analógico digital. La secuencia capturada es directamente enviada como datos digitales. Actualmente dos interfaces de comunicación digital muy utilizadas son la USB (Universal Serial Bus) y la Firewire (IEEE 1394). Cada una de éstas posee diferentes características, tales como la velocidad de transferencia, longitud de cable, etc.

La velocidad máxima de transferencia en la versión de USB 1.1 es de aproximadamente 12Mbps, mientras que en la versión 2.0 este valor ascendió hasta 480Mbps con una longitud máxima de cable entre dispositivos de 5 metros.

En cuanto a la interfaz Firewire, la especificación IEEE 1394a permite velocidades de hasta 400Mbps y la IEEE 1394b de hasta 800Mbps, con una longitud de cable de hasta 100 metros.

La velocidad con la que la cámara puede enviar información a la computadora es determinante, ya que limita la cantidad de cuadros por segundo (FPS) o la resolución del video a capturar.

5. Conclusiones

Como se detallará en capítulos posteriores, la selección del hardware para el sistema requirió un minucioso estudio de las cámaras existentes, sus capacidades de captura e interfaz con las computadoras.

El conocimiento y correcta aplicación de los parámetros vistos en este capítulo son los que permitieron determinar las condiciones óptimas de captura a fin de evitar tomas de pobre calidad cuyo contenido presenten imágenes con borronado, fuera de foco, oscuras o escasas de definición entre otras.

Al tratarse además de un sistema de captura en tiempo real, se debió contemplar en la determinación del hardware, la interfaz de comunicación óptima con la computadora.

De esta forma, la secuencia de imágenes obtenidas en cada captura estarán idealmente listas para su procesamiento evitando preprocesamientos costosos o que degraden la señal original.

Capítulo II. Técnicas de procesamiento

1. Introducción

Las características de la aplicación implementada para el presente trabajo hicieron necesario la investigación de diferentes técnicas de procesamiento. En particular, en este capítulo se analizan técnicas de filtrado, etiquetado y estimación de movimiento.

Cabe destacar que algunas de éstas finalmente no fueron utilizadas, debido a que no resultaron lo suficientemente precisas como se esperaba. En estos casos, se debió realizar un método ad-hoc para realizar dichas tareas.

2. Filtrado por umbralización

El filtrado por umbralización es especialmente útil cuando se trabaja con imágenes que contienen objetos de distintos colores o niveles de grises.

En las imágenes en escala de grises, la intensidad de luminosidad de cada píxel puede variar entre 0 y 255. Por lo tanto, se puede definir un valor o *umbral* que servirá como límite para determinar si se filtra o no un píxel de la imagen.

El modelo de color RGB define tres canales: el canal rojo (R), el canal verde (G) y el canal azul (B). RGB asigna un valor de intensidad a cada píxel que oscila entre 0 y 255 para cada uno de los canales de una imagen en color. Por ejemplo, un color rojo puro sería un valor R de 255, un valor G de 0 y un valor B de 0. Un color amarillo puro sería

un valor R de 255, un valor G de 255 y un valor B de 0, tal como se puede observar en la figura 2.1. Cuando los valores de los tres componentes son idénticos, se obtiene un matiz de gris. Si el valor de todos los componentes es de 255, el resultado será blanco puro y será negro puro si todos los componentes tienen un valor 0.

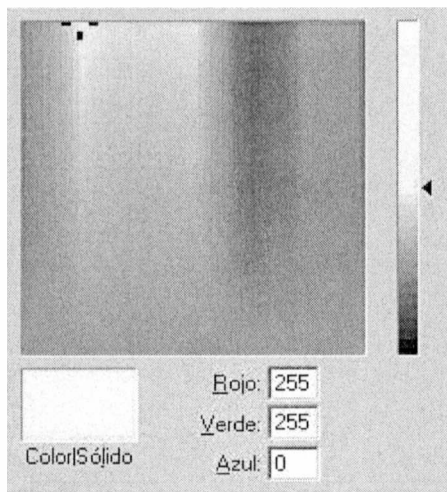


Figura 2.1. Modelo de color RGB

Cada canal de una imagen RGB puede ser procesado de forma independiente. En la figura 2.2 se observa una imagen en color RGB y sus respectivos canales de forma separada. Las intensidades de luminosidad de cada canal están representadas en una escala de grises entre 0 y 255. De esta forma, el valor de intensidad 0 (negro puro) representa la carencia completa de luminosidad, y el valor de intensidad 255 (blanco puro) representa la mayor luminosidad posible.

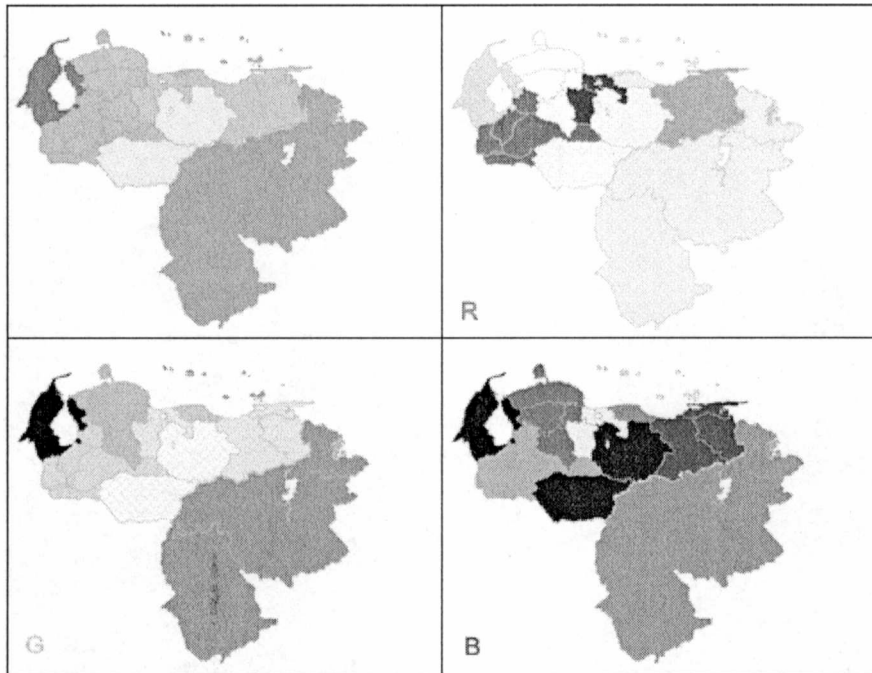


Figura 2.2. Una imagen color RGB y sus correspondientes canales

Tomando en cuenta el modelo de color RGB, realizar un filtrado por umbralización consiste en definir tres umbrales, uno por cada canal.

Si se necesita, por ejemplo, filtrar los colores rojizos de una imagen, se puede realizar un filtro con los siguientes valores:

- $180 \leq R \leq 255$
- $0 \leq G \leq 70$
- $0 \leq B \leq 70$

De esta forma, cada píxel en la imagen será analizado y si sus valores RGB cumplen las tres condiciones traspasará el filtro, de lo contrario no lo hará. Como se puede observar en el filtro generado, los valores que permite pasar no son exclusivamente los de rojo puro, ya que raramente un píxel de una imagen correspondiente a una escena real cumple con dicha condición.

Si se aumentaran los valores de G y B, o se disminuyera el valor de R, se permitiría traspasar el filtro a una gama de colores rojos mas cercanos a los grisáceos.

3. Etiquetado

El objetivo del etiquetado de una imagen es el de determinar los objetos de una escena. Para poder realizar esto se necesita inicialmente determinar cómo se individualiza y separa cada uno de los objetos en una escena.

A continuación se abordarán los conceptos de vecindad y conectividad, para luego analizar las distintas formas de realizar un correcto etiquetado. Por simplicidad se trabajará con imágenes binarias, pero todos los conceptos aquí tratados pueden aplicarse a imágenes en escala de grises o color.

3.1. Vecindad y conectividad

Un píxel en una imagen digital se encuentra esta adyacente a un grupo de píxeles vecinos. Se dice que dos píxeles son 4-vecinos u 8-vecinos si comparten un vecindario. Si se toma en cuenta un píxel central $[i,j]$, la diferencia entre ambos vecindarios es la siguiente:

- los 4-vecinos son $[i+1, j]$, $[i-1, j]$, $[i, j+1]$ y $[i, j-1]$; es decir, se toman en cuenta solo los vecinos horizontales y verticales.
- los 8-vecinos son todos los 4-vecinos más $[i+1, j+1]$, $[i+1, j-1]$, $[i-1, j+1]$ y $[i-1, j-1]$; es decir, se toman en cuenta los vecinos horizontales, verticales y diagonales.

Luego, se dice que un píxel está *4-conectado* a sus 4-vecinos, y está *8-conectado* a sus 8-vecinos. Un *camino* del píxel $[i_0, j_0]$ al píxel $[i_n, j_n]$ es una secuencia de píxeles tal

que $[i_k, j_k]$ es vecino de $[i_{k+1}, j_{k+1}]$ para todo $k=0..n-1$. Se pueden definir caminos 4-conectados y 8-conectados, tal como se observa en la figura 2.3

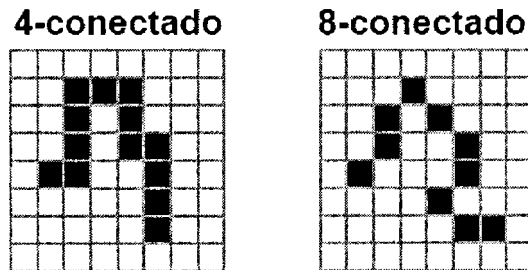


Figura 2.3. Ejemplos de 4-caminos y de 8-caminos

En la figura 2.3 se puede observar dos tipos de píxeles. Los píxeles negros, que pertenecen al *frente* de la imagen, y los blancos que pertenecen al *fondo* de la imagen. Dados dos píxeles p, q pertenecientes al frente de la imagen, los mismos están *conectados* si existe un camino que lleve de p a q cuyos píxeles son todos del frente.

La *conectividad* es una relación de equivalencia entre dos píxeles, ya que cumple con las propiedades de reflexividad, conmutatividad y transitividad.

Un conjunto de píxeles en el que cada píxel está conectado a todos los demás píxeles se llama *componente conexas*.

3.2. Etiquetado de componentes

El etiquetado es una operación que consiste en encontrar las componentes conexas de una imagen. Los puntos en una componente conexas forman una región candidata a representar un objeto.

Un algoritmo de etiquetado encuentra todas las componentes conexas en una imagen y les asigna una etiqueta única a cada punto de una misma componente. Cada componente conexas tendrá entonces una etiqueta que lo identifica y diferencia del resto.

En la figura 2.4 se puede observar un ejemplo de etiquetado de una imagen.

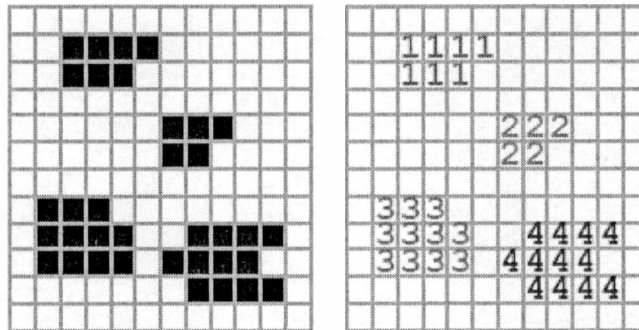


Figura 2.4. Una imagen y sus correspondientes componentes conexas

Existe mas de una forma de realizar un algoritmo de etiquetado de las componentes conexas de una imagen. Una forma simple de hacerlo es con un algoritmo recursivo que realice lo siguiente:

1. Examinar la imagen buscando un píxel que aun no tenga una etiqueta asignada, y asignarle una nueva etiqueta L.
2. Recursivamente, asignar la etiqueta L a todos sus vecinos adyacentes.
3. Detenerse si no hay más píxeles sin etiqueta. En caso contrario volver al primer paso.

Este tipo de algoritmos recursivo es muy simple y fácil de comprender, pero tiene una desventaja: en un procesador secuencial es muy ineficiente. Sin embargo, el mismo puede ser paralelizado para superar esta pérdida.

Una algoritmo mas eficiente para realizar el etiquetado es el secuencial. El mismo consiste en producir una tabla auxiliar que contenga las equivalencias. A continuación se da un ejemplo de una implementación de este algoritmo usando 4-conectividad:

1. Examinar la imagen de izquierda a derecha, de arriba hacia abajo.

2. Si el píxel es del frente de la imagen entonces:
 - a. Si solo uno de sus vecinos de arriba o de la izquierda tiene una etiqueta, o ambos tienen la misma, entonces se copia la etiqueta.
 - b. Si ambos tienen etiquetas diferentes, entonces se copia la etiqueta del vecino de arriba y se anota la equivalencia de etiqueta en la *tabla de equivalencias*.
 - c. En caso contrario, asignar una nueva etiqueta a este píxel y anotar la etiqueta en la *tabla de equivalencias*.
3. Si no hay mas píxeles que considerar, ir al paso 2.
4. Encontrar la menor etiqueta para cada conjunto de equivalencia en la *tabla de equivalencias*.
5. Examinar la imagen reemplazando cada etiqueta por la menor en su conjunto de equivalencia.

Como se observa, el algoritmo es sencillo y usualmente necesita dos pasadas sobre la imagen.

4. Asociación por bloques (block matching)

Los algoritmos de asociación por bloques o *block matching* son utilizados en la estimación de movimiento y se han convertido en una técnica fundamental para la mayoría de los estándares de compresión y codificación de video basados en la estimación de movimiento. Estos algoritmos consisten en la eliminación de redundancia espacial y temporal que existe en imágenes consecutivas de secuencias de video.

Los algoritmos de asociación por bloques estiman la cantidad de movimiento en base a bloques de acuerdo a un criterio. Existen variados criterios para evaluar la

correspondencia de bloques entre dos imágenes de una secuencia de video, entre los que se encuentran:

- Función de Correlación en Cruz (CCF)
- Clasificación de Diferencia de Píxeles (PDC)
- Diferencia Absoluta Media (MAD)
- Diferencia Cuadrada Media (MSD)
- Proyección Integral (IP)

Algunas de estas técnicas son simples de evaluar y otras son más complejas. A continuación se revisarán algoritmos que utilizan alguno de estos criterios:

4.1. Búsqueda completa

Uno de los primeros algoritmos utilizados para la compensación de movimiento basado en bloques fue el que se denomina *búsqueda completa* o *búsqueda exhaustiva*. En este algoritmo, cada bloque dentro de una ventana dada de la búsqueda es comparado al bloque actual, obteniéndose la mejor correspondencia basándose en alguno de los criterios anteriores, como por ejemplo la diferencia absoluta media.

Este algoritmo es el mejor en términos de calidad de predicción de la imagen y posee una gran sencillez algorítmica; sin embargo, es muy costoso computacionalmente. Es por esto que se ha buscado otro tipo de algoritmos que sean mas eficientes. De todas formas, aunque mas eficientes, estos algoritmos no llegan a obtener la calidad del algoritmo de *búsqueda completa*.

4.2. Algoritmos basados en firma

Existen muchas aproximaciones para reducir la complejidad computacional. Una de ellas son los algoritmos basados en firma, que reducen los cálculos mediante la utilización de varias etapas. En cada una de las etapas usa un criterio de comparación

diferente. En la primera etapa todos los bloques se evalúan utilizando un criterio computacionalmente sencillo. Luego, basados en los resultados de esta etapa, se escoge un subconjunto de los candidatos para la próxima etapa, donde se utiliza un criterio más complejo.

4.3. Algoritmos de ventana de búsqueda dinámica

Algunos algoritmos están basados en la naturaleza de la imagen y se clasifican como algoritmos basados en el *principio de localidad*. Uno de los problemas con estos algoritmos es que pueden convergir a un mínimo local antes que al mínimo global. Estos algoritmos pueden ser modificados cambiando la manera en que el algoritmo reduce el área de búsqueda.

Por ejemplo, la reducción del área de búsqueda se puede hacer en función de las dos deformaciones más pequeñas en la etapa anterior, en lugar de solo la deformación más pequeña. Estos algoritmos se denominan de *ventana de búsqueda dinámica*.

4.4. Algoritmo de búsqueda de tres etapas

Este algoritmo se hizo muy popular debido a su simplicidad, robustez y cercanía a una óptima performance. Busca los mejores vectores de movimiento en un patrón de búsqueda que va de grueso a fino. El algoritmo puede describirse de la siguiente forma:

- Paso 1: se escoge un tamaño inicial de paso. Para la comparación, se escogen ocho bloques a una distancia del tamaño de paso elegido, alrededor del bloque central.
- Paso 2: el tamaño del paso se reduce. El centro se mueve al punto con la distorsión mínima.

Los pasos 1 y 2 se repiten hasta que el tamaño del paso se vuelve menor a 1. Un camino particular para la convergencia de este algoritmo se muestra en la figura 2.5.

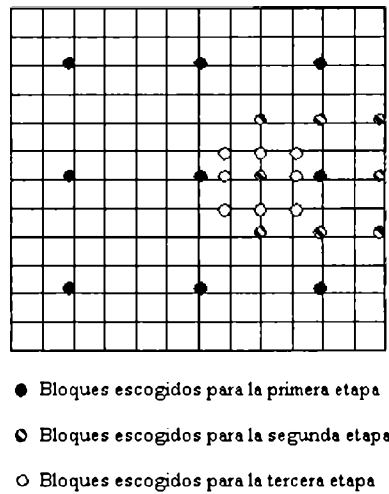


Figura 2.5. Ejemplo de un camino para la convergencia de la Búsqueda de Tres Etapas

Un problema que ocurre con este algoritmo es que utiliza un patrón de chequeo uniforme en el primer paso, que se vuelve ineficiente para la estimación de movimiento pequeño.

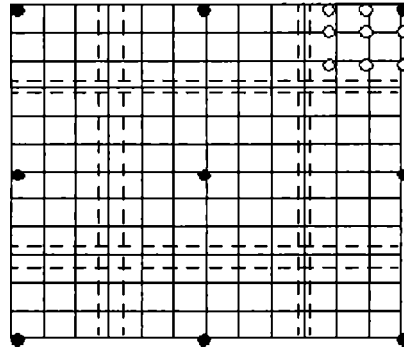
4.5. Algoritmo de Búsqueda Binaria

Al igual que el anterior, este algoritmo también se ha vuelto muy popular para la estimación de movimiento. De hecho, es el utilizado por MPEG. La idea básica detrás de este algoritmo es dividir la ventana de búsqueda en un número de regiones y realizar una búsqueda completa en sólo una de estas regiones. Se puede describir de la siguiente manera:

- Paso 1: se evalúa la MAD en una grilla de nueve píxeles, que incluye el centro, las cuatro esquinas de la ventana de búsqueda y cuatro píxeles en el vecindario. La ventana de búsqueda es dividida en regiones basadas en estos puntos.

- Paso 2: se realiza una búsqueda completa en la región correspondiente al punto con la menor MAD.

En esta técnica existen píxeles que nunca son considerados. Por esto, a pesar de requerir una menor cantidad de comparaciones, su performance no es muy buena. La convergencia de este algoritmo se puede observar en la figura 2.6.



- Puntos iniciales basados en qué regiones están formadas
- Puntos de la región donde se realiza una búsqueda completa

Figura 2.6. Ejemplo de convergencia de la Búsqueda Binaria

Además de los métodos de búsqueda revisados en este capítulo, existen muchos otros que no serán analizados, tales como la búsqueda logarítmica, la búsqueda de las cuatro etapas, el algoritmo de búsqueda en cruz, la búsqueda en espiral, etc.

5. Flujo Óptico

Cuando se estudia una secuencia de imágenes, se pretende estudiar la evolución temporal de los objetos que son capturados por la cámara, para tratar de extraer la mayor cantidad de características de los mismos. Una de las principales características que puede ser extraída de una secuencia de imágenes es la velocidad de un objeto. Las

técnicas de flujo óptico tratan de dar solución a la determinación de la velocidad de los objetos dentro de una secuencia de imágenes.

5.1. Teoría del Flujo Óptico

Una de las principales características que puede ser extraída de una secuencia de imágenes es la velocidad de un objeto. Mediante esta característica, es posible tener una noción diferencial de la posición que tendrá el objeto en el siguiente instante.

Las técnicas de flujo óptico tratan de dar solución a la determinación de la velocidad de los objetos dentro de una secuencia de imágenes.

En la figura 2.7 se puede observar dos imágenes consecutivas en las que un objeto se encuentra en movimiento. El desplazamiento aparente de los puntos de una imagen se produce cuando:

- la cámara se mueve en relación con los objetos que se enfocan
- uno o mas objetos se mueven dentro de la escena

En el caso ideal, en el que la iluminación se mantiene a lo largo de la secuencia de imágenes, el Flujo Óptico se corresponde con el campo de velocidades.

Suponiendo que la iluminación es constante, un punto de la imagen conserva su intensidad al moverse. Extendiendo este hecho a todos los puntos de la imagen, se deduce una ecuación que indique la componente de la velocidad, en dirección normal a los contornos. Para hallar la segunda componente de la velocidad deben hacerse suposiciones adicionales.

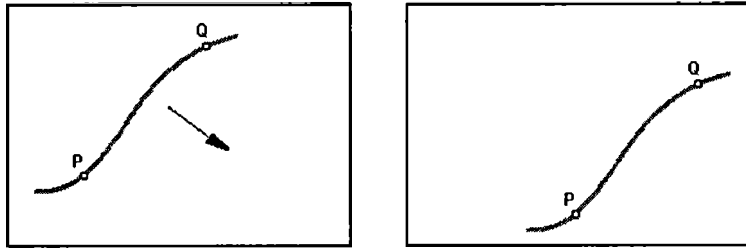


Figura 2.7 - Desplazamiento de dos puntos de un objeto

Sea $I(x, y, t)$ la intensidad del punto (x, y) de la imagen en el instante de tiempo t . Sean $u(x, y)$ y $v(x, y)$ las componentes en x e y del flujo óptico en ese punto. Después de un pequeño intervalo de tiempo δt , este punto se encontrará en $(x + u\delta t, y + v\delta t)$, conservando su intensidad. Por lo tanto: $I(x + u\delta t, y + v\delta t, t + \delta t) = I(x, y, t)$.

Considerando que la intensidad varía suavemente en casi toda la imagen, es posible desarrollar el lado izquierdo de la ecuación en la serie de Taylor. Si se realizan operaciones matemáticas y se utilizan las abreviaciones:

$$u = \frac{\delta x}{\delta t}, \quad v = \frac{\delta y}{\delta t}, \quad I_x = \frac{\delta I}{\delta x}, \quad I_y = \frac{\delta I}{\delta y}, \quad I_t = \frac{\delta I}{\delta t},$$

se obtiene finalmente la *Ecuación de Contención del Optical Flow*: $I_x u + I_y v + I_t = 0$

Esta ecuación es la base de las técnicas diferenciales del cálculo del Flujo Óptico. Identifica una línea, perpendicular al gradiente de la intensidad del punto dado, sobre la cual estará su velocidad. Esta línea se denomina *línea de condición*. Por tanto, con esta ecuación, se conoce solamente una componente de la velocidad.

Hay que tener en cuenta que en el desarrollo de esta ecuación se asume que los contornos de intensidad de la imagen se mueven sin discontinuidades en el campo de velocidades.

5.2. Problemas del Flujo Óptico

Las técnicas de flujo óptico buscan correspondencias entre puntos de dos imágenes basándose en las intensidades de luminosidad. Existen diversos problemas que hacen que esta búsqueda no se pueda llevar a cabo de una forma correcta. A continuación se analizarán diferentes situaciones en las que es altamente complejo o imposible determinar el flujo óptico.

5.2.1. Problemas de iluminación de la escena

Dado que se está estimando el movimiento a partir de las intensidades de luminosidad de los puntos en las imágenes, surgen dos problemas básicos que están relacionados con la iluminación de la escena.

El primer problema se refiere a la *falta de detección de movimiento*. Tiene que haber suficiente variación en el nivel de gris de la región que se mueve para que su movimiento se detecte. En la figura 2.8a se puede observar una esfera con nivel de gris constante que gira sobre su eje. Aunque existe el movimiento, el mismo no se detecta por lo que el flujo óptico es nulo.

El segundo problema se refiere a la *detección incorrecta de movimiento*. En la figura 2.8b se puede observar una esfera estática. Cuando se produce un cambio en la iluminación de la escena, es posible detectar flujo óptico, aún cuando no exista el movimiento.

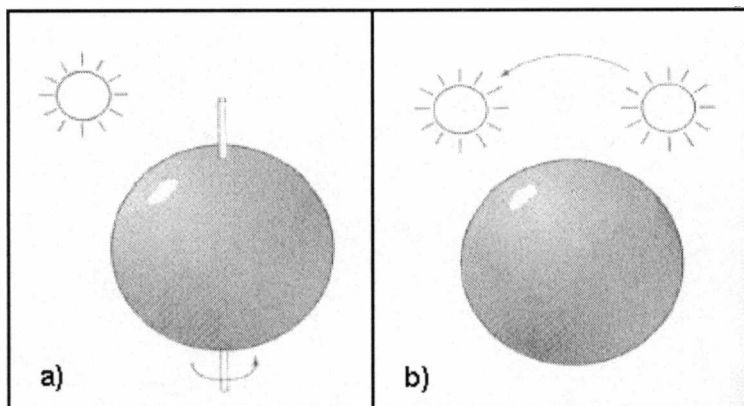


Figura 2.8 - Problemas de iluminación de la escena

5.2.2. Problema de la apertura

El problema de la apertura surge de la imposibilidad de asignar un valor unívoco al desplazamiento de un punto debido a que la región analizada es demasiado pequeña. Es un caso particular del problema de la correspondencia.

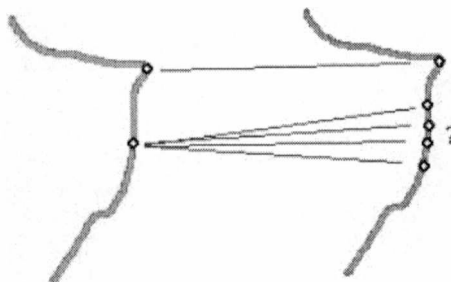


Figura 2.9 - Problema de la apertura

Como se ha visto, utilizando la *ecuación de contención del flujo óptico* sólo se obtiene información sobre la componente de la velocidad en la dirección del gradiente de intensidad, esto es, perpendicularmente a los contornos. No se podrá determinar la componente del flujo óptico en la dirección del contorno. Esta es la causa del problema de la apertura.

Una resolución al problema de la apertura es la aplicación de criterios de vecindad. Con ello se pretende mejorar el cálculo del flujo óptico en regiones ambiguas mediante la observación en regiones próximas que presenten mejores características.

5.2.3. Problema de la oclusión

El problema de la oclusión ocurre cuando una superficie es cubierta o descubierta como consecuencia de una rotación o traslación tridimensional de un objeto que ocupa sólo parte del campo de visión.

Si se cuenta con una secuencia de imágenes, puede suceder que un punto de un objeto en una imagen no sea encontrado en la siguiente imagen de la secuencia y viceversa. Si esto sucede, no es posible calcular el flujo óptico.

La figura 2.10 muestra dos imágenes consecutivas en las que un objeto se traslada lateralmente hacia la derecha. Como se puede observar, existen regiones de las imágenes que se cubren o descubren, originando un problema de correspondencia de puntos entre las dos imágenes.

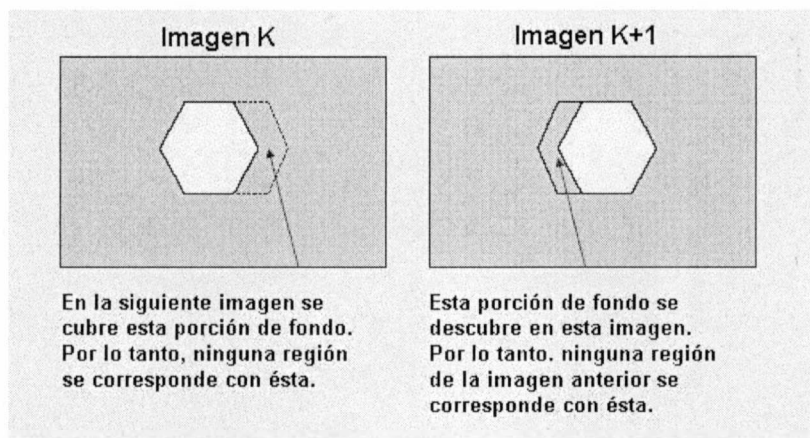


Figura 2.10. Problema de la oclusión

Capítulo III: Técnicas estereoscópicas

1. Introducción

Los ojos humanos, debido a su separación, obtienen dos imágenes de una misma escena con pequeñas diferencias entre ellas. Nuestro cerebro procesa las diferencias entre ambas imágenes y las interpreta de forma que percibimos la sensación de profundidad, lejanía o cercanía de los objetos que nos rodean. Este proceso se denomina estereopsis.

Las técnicas estereoscópicas intentan reproducir la sensación de profundidad de la visión estereoscópica natural. Si obtenemos dos fotografías con una separación adecuada, correspondientes a la visión que se obtendría con cada ojo es posible reconstruir la profundidad.

2. Fotografía digital

2.1. Imágenes de intensidad

Las imágenes de intensidad miden la cantidad de luz que impacta contra el dispositivo fotosensible de la cámara. En una imagen digital, se ha sustituido el soporte fotosensible (la película de sales de plata) por otro soporte, el sensor digital. De este modo, la luz penetra a través de la lente y llega al sensor.

Una imagen digital es una matriz de dos dimensiones $M(x,y)$, en donde las coordenadas x e y indican la posición del punto, y el valor numérico guardado en dicha posición se corresponde con la intensidad de luz.

En un sistema visual como el del hombre, el proceso de formación de la imagen comienza con los rayos de luz del mundo exterior impactando contra los receptores de la retina. De esta forma, existen una serie de parámetros físicos que participan en la formación de la imagen.

Los parámetros ópticos de la lente caracterizan la óptica del sensor:

- tipo de lente
- distancia focal
- campo de visión
- apertura angular

Los parámetros fotométricos aparecen en los modelos de energía de luz que alcanza el sensor después de ser reflectada desde los objetos en la escena:

- tipo, intensidad y dirección de la iluminación
- propiedades de reflectancia de las superficies
- efectos de la estructura del sensor en la acumulación de luz de los fotorreceptores

Los parámetros geométricos determinan la posición de la imagen en la que los puntos en tres dimensiones son proyectados:

- tipo de la proyección
- posición y orientación de la cámara en el espacio
- distorsiones de la perspectiva introducidas en el proceso de formación de la imagen

Además de estos parámetros, son necesarios otros para caracterizar la imagen digital y sus sistemas de adquisición:

- propiedades físicas de la matriz fotosensible de la cámara
- la naturaleza discreta de los fotorreceptores
- la cuantización de la escala de intensidad

A continuación se revisarán los aspectos de óptica, radiometría y geometría que son relevantes para la formación de la imagen. Cabe destacar que estos temas se analizarán de manera sintética, teniendo en cuenta sólo los aspectos de relevantes relacionados con el presente trabajo.

2.1.1. Óptica

Como se ha visto, el proceso de formación de la imagen comienza con los rayos de luz que entran en la cámara a través de la apertura angular e impactan el plano de imagen, que es el dispositivo fotosensible que registra las intensidades de luz. La mayoría de estos rayos son el resultado de las reflexiones de los rayos emitidos por los orígenes de luz que impactan sobre las superficies de los objetos.

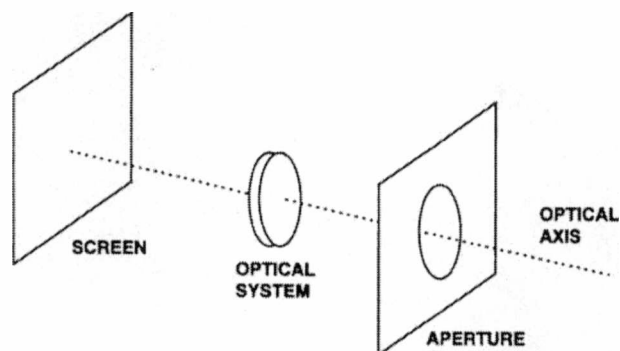


Figura 3.1 – Elementos básicos de un dispositivo de imágenes

2.1.1.1. Foco de la imagen

Cualquier punto en la escena refleja luz que llega desde muchas posibles direcciones. Entonces, varios rayos reflejados por el mismo punto pueden entrar en la cámara. Para obtener imágenes definidas, todos los rayos que vienen desde un solo punto en la escena, P , debe converger en un solo punto del plano de la imagen p (la imagen de P). Cuando esto sucede se dice que la imagen de P está en foco.

Existen dos posibilidades para lograr focalizar todos los rayos de luz de un punto de la escena en un único punto de la imagen:

- Reducir la apertura de la cámara a un punto, al que llamaremos pinhole. De esta forma solo un rayo desde cualquier punto dado podrá entrar en la cámara, creando una correspondencia “uno a uno” entre los puntos visibles, los rayos, y los puntos de la imagen. Una desventaja de esta alternativa es que se necesita un tiempo de exposición bastante largo (varios segundos), ya que el pinhole permite penetrar muy poca luz por unidad de tiempo.
- Introducir un sistema óptico compuesto por lentes, aperturas y otros elementos, diseñados explícitamente para hacer que todos los rayos que vienen de un mismo punto tridimensional converja en un único punto en la imagen.

2.1.1.2. Lentes delgadas

Para continuar explorando los conceptos básicos de la formación de una imagen analizaremos un sistema óptico simple denominado lente delgada. El mismo está caracterizado por dos elementos:

- el eje óptico
- dos puntos especiales F_l y F_r , llamados foco izquierdo y derecho respectivamente

El eje óptico pasa a través del centro del lente O y perpendicularmente al plano. F_l y F_r están ubicados en el eje óptico en lados opuestos del lente a la misma distancia de O . La distancia focal se indica con f . En la figura 3.2 se puede observar un lente delgado.

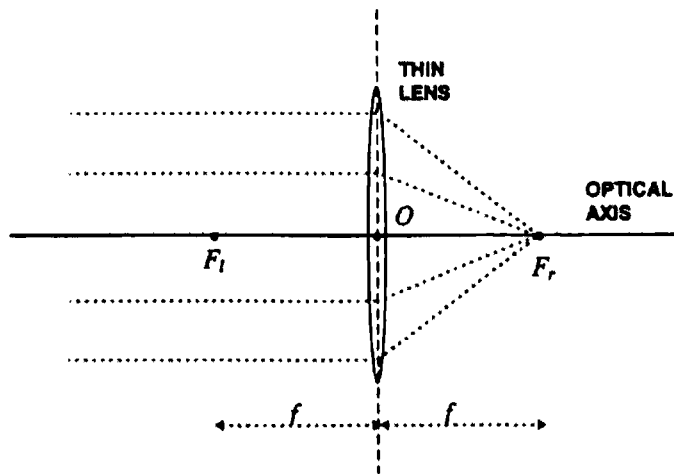


Figura 3.2 – Lente delgado

Por construcción, un lente delgado desvía todos los rayos paralelos al eje óptico que llegan de un lado, al foco del otro lado. Esto queda expuesto en las siguientes dos propiedades:

- Propiedad 1: cualquier rayo que entra paralelo al eje en un lado, se dirige a través del foco hacia el otro lado.
- Propiedad 2: cualquier rayo que entra paralelo al foco de un lado, emerge paralelo al eje en el otro lado.

En base a estas dos propiedades se deriva la *Ecuación Fundamental de Lentes Delgadas*. Se considera un punto P , no muy lejano al eje óptico; $Z + f$ es la distancia de P desde el lente a lo largo del eje óptico, tal como se puede observar en la figura 3.3.

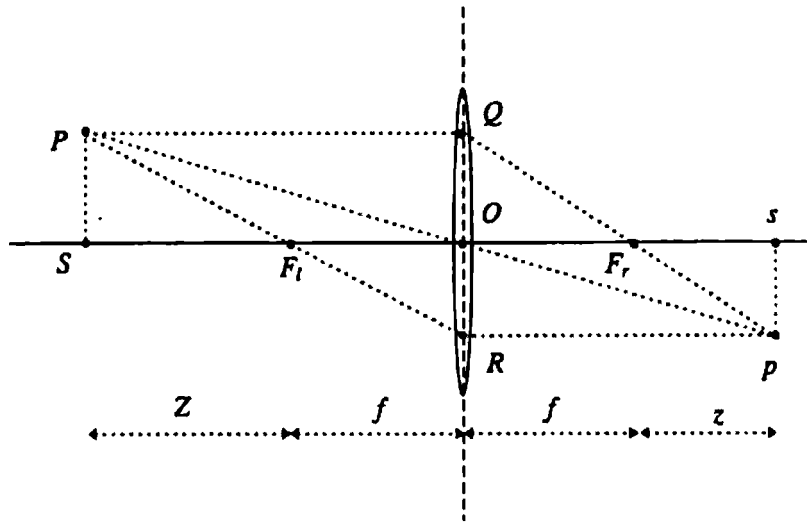


Figura 3.3 – Formación de la imagen en un lente delgado

Un lente delgado focaliza todos los rayos de P en el mismo punto, el punto de la imagen p . Por lo tanto, se puede localizar p intersectando solo dos rayos conocidos, sin tomar en cuenta los caminos de los demás.

Al aplicar la propiedad 1 al rayo PQ y la propiedad 2 al rayo PR , PQ y PR son desviados para intersectar a un cierto punto en el otro lado del lente delgado. Pero dado que el lente focaliza todos los rayos que vienen de P en el mismo punto, PQ y PR deben intersectar a p . De la anterior figura y utilizando dos pares de triángulos similares $\langle PFI \rangle$, $\langle ROFI \rangle$ y $\langle psFr \rangle$, $\langle QOFr \rangle$ se obtiene inmediatamente: $Zz = f^2$.

Valuando $\hat{Z} = Z + f$ y $\hat{z} = z + f$, la Ecuación Fundamental de Lentes Delgadas se reduce a:

$$\frac{1}{\hat{Z}} + \frac{1}{\hat{z}} = \frac{1}{f}$$

El rayo que pasa a través del centro del lente O se denomina *rayo principal*.

2.1.1.3. Campo de visión

La distancia focal junto con el diámetro efectivo del lente d (no el diámetro físico del lente) determinan el campo de visión del lente, que no es más que una medida angular de la porción de espacio 3-D que será visto por la cámara. Se acostumbra definir el campo de visión w como la mitad del ángulo sostenido por el diámetro del lente visto desde el foco: $\tan w = \frac{d}{2f}$.

2.1.2. Radiometría

La radiometría es la parte del proceso de formación de la imagen que se refiere a la relación entre la cantidad de energía emitida desde los orígenes de iluminación, reflectada desde las superficies y registrada por los sensores.

La *Ecuación Fundamental de la Formación Radiométrica de la Imagen* indica cuánta luz es recolectada por la cámara dependiendo de la dirección de la luz y la distancia focal. Para los fines de este trabajo no es necesario analizar en profundidad esta ecuación.

2.1.3. Geometría

Para lograr relacionar la posición de los puntos de la escena con sus puntos correspondientes en la imagen es necesario modelar la proyección geométrica realizada por el sensor.

2.1.3.1. La Cámara de Perspectiva

El modelo geométrico de una cámara de intensidad más común es el modelo de perspectiva o pinhole. El mismo consiste del plano de imagen π , y un punto tridimensional O que denota el centro o foco de proyección. La distancia entre π y O es la distancia focal. La línea que pasa a través de O y es perpendicular a π es el eje óptico. El centro de la imagen o punto principal es la intersección entre π y el eje

óptico, y se notará o . La imagen de P es el punto en el que la línea recta a través de P y O intersectan el plano de imagen π , y se notará p . En la figura 3.4 se puede observar un modelo de cámara de perspectiva.

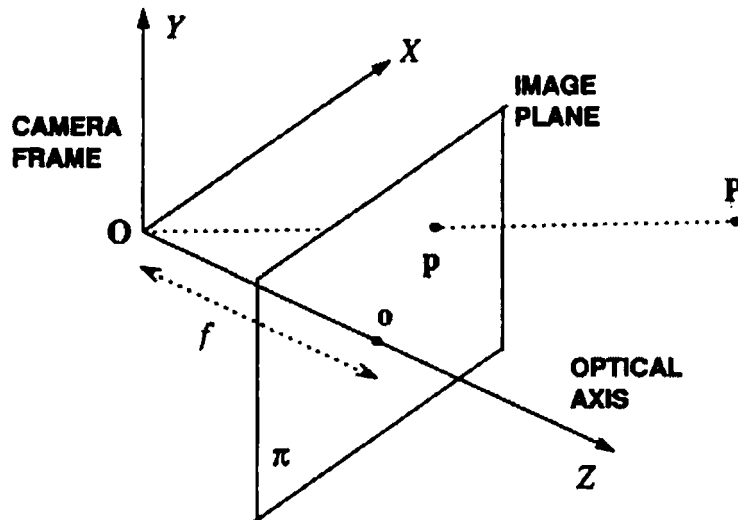


Figura.3.4 – Modelo de cámara de perspectiva

Se denominará *marco de cámara* al marco de referencia tridimensional en el que O es el origen y el plano π es ortogonal al eje Z , y siendo $P = [X, Y, Z]^T$ y $p = [x, y, z]^T$. En este marco se tienen las llamadas *Ecuaciones Fundamentales de la Cámara de Perspectiva*:

$$x = f \frac{X}{Z}, \quad y = f \frac{Y}{Z}.$$

2.1.3.2. La Cámara de Perspectiva Débil

Las ecuaciones fundamentales de la cámara de perspectiva no son lineales, por lo que no preservan las distancias entre puntos. Una aproximación que torna estas ecuaciones lineales es el modelo de cámara de perspectiva débil. El mismo requiere que la distancia relativa a lo largo del eje óptico, δz , de dos puntos cualesquiera, sea mucho menor que la distancia promedio, \bar{Z} (desde el punto de vista de la cámara). De esta forma, para cada punto de la escena P se puede escribir:

$$x = f \frac{X}{Z} \approx \frac{f}{Z} X$$

$$y = f \frac{Y}{Z} \approx \frac{f}{Z} Y$$

2.2. Adquisición de imágenes digitales

Un sistema de adquisición de imágenes digitales consiste típicamente de tres componentes de hardware: una cámara (generalmente una de tipo CCD), un capturador de cuadros y una computadora.

La entrada de la cámara es la luz que ingresa en el lente e impacta el plano de imagen. En una cámara CCD, el plano de imagen físico es el vector CCD, que no es más que una grilla rectangular de $n \times m$ fotosensores. Cada uno de estos fotosensores es sensible a la intensidad de la luz, y puede ser visto como una pequeña caja negra rectangular que convierte la energía de la luz en voltaje.

La salida del vector CCD es usualmente una señal eléctrica continua llamada *señal de video*, la cuál se genera examinando los fotosensores en un orden determinado y leyendo sus voltajes. Esta señal es enviada a un dispositivo eléctrico llamado capturador de cuadros, en donde es digitalizada en un vector rectangular de dos dimensiones $N \times M$ de valores enteros y almacenado en un buffer de memoria.

Finalmente contamos con la imagen en píxeles, la cual es transferida a una computadora para poder procesarla. Sin embargo, el número de elementos a lo largo y a lo ancho del vector CCD es usualmente diferente a la dimensión en píxeles del buffer. Dado esto, la relación entre la posición (x_{im}, y_{im}) <en píxeles> en la imagen del buffer y la posición (x_{CCD}, y_{CCD}) <en elementos de CCD> en el vector CCD está dada por:

$$x_{im} = \frac{n}{N} x_{CCD} \quad , \quad y_{im} = \frac{m}{M} y_{CCD}$$

Tomando en cuenta esta relación, se puede concluir que la diferencia de escala en la imagen con respecto al CCD se debe tanto a las relaciones $\frac{n}{N}$ y $\frac{m}{M}$ como a la diferente relación de aspecto horizontal-vertical del CCD. Es conveniente asumir que los elementos del CCD tienen una correspondencia de uno a uno con los píxeles de la imagen e introducir tamaños efectivos horizontales y verticales para obtener ambas escalas.

2.3. Parámetros de Cámara

Los algoritmos que reconstruyen la estructura tridimensional de una escena o que calculan la posición de objetos en el espacio necesitan ecuaciones que relacionen las coordenadas de los puntos en el espacio tridimensional con las coordenadas de sus correspondientes puntos en la imagen. Estas ecuaciones están escritas en el marco de referencia de la cámara, pero a menudo se asumen dos cosas:

- el marco de referencia de la cámara puede ser localizado con respecto a un marco de referencia conocido, como el marco de referencia del mundo real
- las coordenadas de los puntos de la imagen en el marco de referencia de la cámara pueden ser obtenidos desde las coordenadas de píxel, las únicas directamente disponibles desde la imagen

Al decir esto, se está asumiendo que se conocen ciertas características de la cámara. Estas características son los parámetros extrínsecos e intrínsecos de la cámara.

2.3.1 Parámetros Extrínsecos

El marco de referencia de la cámara es generalmente desconocido, por lo que se plantea el problema de determinar la localidad y orientación de la cámara con respecto a un marco de referencia conocido, que es el del mundo real. Para resolver este problema solo se cuenta con la información de la imagen.

Los parámetros extrínsecos están definidos como cualquier conjunto de parámetros geométricos que identifique unívocamente la transformación entre el marco de referencia de cámara desconocido y uno de referencia conocido llamado marco de referencia del mundo. Típicamente, se usan los siguientes parámetros:

- un vector de traslación tridimensional T describiendo las posiciones relativas de los orígenes de los dos marcos de referencia.
- una matriz de rotación R de tamaño 3×3 ortogonal ($R^T R = R R^T = I$) que nos brinda los ejes correspondientes de los dos marcos entre si. Las relaciones ortogonales reducen el número de grados de libertad de R a tres.

La relación entre las coordenadas de un punto P en los marcos del mundo y de la cámara, P_w y P_c respectivamente, es entonces $P_c = R(P_w - T)$ donde:

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

2.3.2 Parámetros Intrínsecos

Los parámetros son necesarios para caracterizar los aspectos ópticos, geométricos y digitales de la cámara. Para una cámara pinhole se necesitan:

- la proyección perspectiva, o sea, la distancia focal.

- la transformación entre las coordenadas del marco de la cámara y las coordenadas en píxeles
- la distorsión geométrica introducida por la óptica

Dejando de lado las distorsiones geométricas posiblemente introducidas por la óptica, la transformación de las coordenadas sería:

$$x = -(x_{im} - o_x)s_x, \quad y = -(y_{im} - o_y)s_y \text{ donde:}$$

(o_x, o_y) son las coordenadas en píxel del centro de la imagen (el punto principal)

(s_x, s_y) es el tamaño efectivo del píxel en milímetros en dirección horizontal y vertical respectivamente

Puede existir el caso de que la óptica introduzca distorsiones originadas en la periferia de la imagen o por el uso de ópticas con campos de visión muy grandes. Estas distorsiones son modeladas como distorsiones radiales. Sin embargo, la distorsión radial puede ser ignorada cuando no se necesita una alta precisión en todas las regiones de la imagen, o cuando los píxeles periféricos pueden ser descartados.

Los parámetros intrínsecos son entonces:

- la distancia focal f
- el punto principal (o_x, o_y)
- el tamaño efectivo del píxel (s_x, s_y)

2.4. Modelo de Cámara

Tomando en cuenta los conceptos revisados anteriormente, se pueden escribir relaciones que unan directamente a las coordenadas de píxel de un punto de la imagen con las

coordenadas del mundo del punto correspondiente tridimensional sin hacer referencia explícita del marco de referencia de la cámara.

La versión lineal de la Ecuación de Proyección Perspectiva relaciona las coordenadas en 3-D de un punto en el marco del mundo y las coordenadas de su correspondiente punto de la imagen, a través de los parámetros extrínsecos e intrínsecos de la cámara. La misma se enuncia como:

$$-(x_{im} - o_x)s_x = f \frac{R_1^T(P_w - T)}{R_3^T(P_w - T)}, \quad -(y_{im} - o_y)s_y = f \frac{R_2^T(P_w - T)}{R_3^T(P_w - T)} \quad \text{donde:}$$

$R_i, i = 1,2,3$ es un vector formado por la fila i de la matriz R .

Esta ecuación se puede reescribir como un producto simple de matrices de la siguiente forma:

$$M_{int} = \begin{pmatrix} -f/s_x & 0 & o_x \\ 0 & -f/s_y & o_y \\ 0 & 0 & 1 \end{pmatrix} \quad \text{y} \quad M_{ext} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & -R_1^T T \\ r_{21} & r_{22} & r_{23} & -R_2^T T \\ r_{31} & r_{32} & r_{33} & -R_3^T T \end{pmatrix}$$

de forma tal que M_{int} depende solo de los parámetros intrínsecos y M_{ext} solo de los extrínsecos. Si se adiciona un "1" como cuarta coordenada de P_w (o sea, se lo expresa en coordenadas homogéneas), se obtiene una ecuación de matriz lineal realizando el siguiente producto, describiendo las proyecciones perspectivas:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = M_{int} M_{ext} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

El *Modelo Cámara de Perspectiva* puede derivarse del anterior definiendo las restricciones apropiadas sobre la *matriz de proyección* $M = M_{\text{int}} M_{\text{ext}}$. Asumiendo que $o_x = o_y = 0$ y que $s_x = s_y = 1$ M puede reescribirse como:

$$M = \begin{pmatrix} -fr_{11} & -fr_{12} & -fr_{13} & fR_1^T T \\ -fr_{21} & -fr_{22} & -fr_{23} & fR_2^T T \\ r_{31} & r_{32} & r_{33} & -R_3^T T \end{pmatrix}$$

Para el *Modelo Cámara de Perspectiva Débil* la matriz de proyección M quedaría escrita de la siguiente forma:

$$p = M \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = \begin{pmatrix} fR_1^T (T - P) \\ fR_2^T (T - P) \\ R_3^T (T - P) \end{pmatrix}$$

Pero $\|R_3^T (T - P)\|$ es simplemente la distancia de P desde el centro de proyección a lo largo del eje óptico. Por lo tanto, la restricción básica para la aproximación de perspectiva débil puede ser escrita como:

$$\frac{|R_3^T (P_i - \bar{P})|}{|R_3^T (\bar{P} - T)|} \ll 1$$

donde P_1 y P_2 son los dos puntos en el espacio 3-D y \bar{P} es el centroide de P_1 y P_2 . Usando esto en la anterior ecuación quedaría:

$$p_i \approx \begin{pmatrix} fR_1^T (T - P_i) \\ fR_2^T (T - P_i) \\ R_3^T (\bar{P} - T) \end{pmatrix}$$

Finalmente. La matriz M para la aproximación de perspectiva débil resulta de la siguiente manera:

$$M = \begin{pmatrix} -fr_{11} & -fr_{12} & -fr_{13} & fR_1^T T \\ -fr_{21} & -fr_{22} & -fr_{23} & fR_2^T T \\ 0 & 0 & 0 & -R_3^T (\bar{P} - T) \end{pmatrix}$$

3. Calibración de cámara

Con el fin de lograr una correcta triangulación del objeto de interés en el espacio, se requiere primeramente conocer las características de las cámaras a utilizar. Existen dos conjuntos de parámetros que deben ser obtenidos mediante el proceso de calibración de cámaras: los parámetros intrínsecos y extrínsecos.

El proceso de calibración tiene como objetivo especificar las ecuaciones de proyección que relacionan las coordenadas de un punto 3-D y sus proyecciones. Se utiliza un conjunto de imágenes de un patrón o grilla de calibración, con la intención de obtener una serie de puntos 3-D. La grilla de calibración debe ser un objeto tridimensional de geometría y posición conocida. La figura 3.5 muestra una grilla de calibración típica.

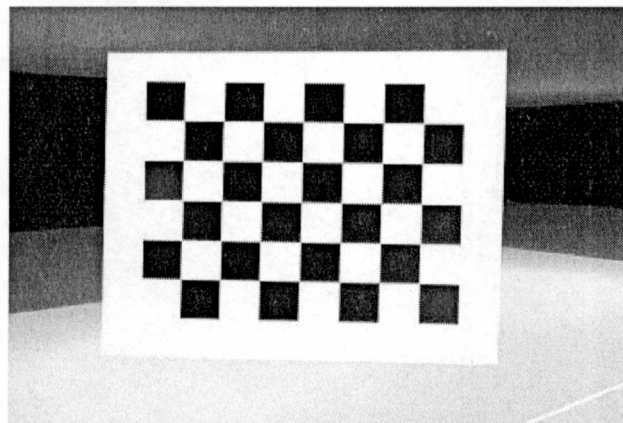


Figura 3.5 – Grilla de calibración

Como se ha visto anteriormente, los parámetros extrínsecos son:

- R , la matriz de rotación de tamaño 3×3
- T , el vector 3-D de traslación

Entre los parámetros intrínsecos se ha nombrado a la distancia focal f y los tamaños efectivos (s_x, s_y) . Desde ahora, y para que cada uno sea independiente de los demás, los parámetros intrínsecos quedarán definidos de la siguiente forma:

- $f_x = f / s_x$, la distancia expresada en unidades de tamaño efectivo de píxel (horizontal)
- $\alpha = s_y / s_x$, la relación de aspecto
- (o_x, o_y) , las coordenadas del centro de la imagen

La calibración extrínseca de una cámara permite obtener los parámetros extrínsecos de la misma. Éstos definen la orientación y la posición de la cámara, respecto a un sistema de coordenadas determinado, al que se conoce como sistema de coordenadas del mundo.

Considerando un punto 3-D $P = [X^w, Y^w, Z^w]^T$ en el marco de referencia del mundo conocido, las coordenadas de P en el marco de referencia de la cámara serán $[X^c, Y^c, Z^c]^T$, con $Z^c > 0$ si P es visible. El origen del marco de referencia de la cámara es el centro de proyección y Z es el eje óptico. A diferencia del marco de referencia del mundo real, la posición y orientación del marco de referencia de la cámara es desconocido, es decir, no se conocen R y T tal que:

$$[X^c, Y^c, Z^c]^T = R[X^w, Y^w, Z^w]^T + T$$

o, escrito en forma de componentes:

$$X^c = r_{11}X^w + r_{12}Y^w + r_{13}Z^w + T_x$$

$$Y^c = r_{21}X^w + r_{22}Y^w + r_{23}Z^w + T_y$$

$$Z^c = r_{31}X^w + r_{32}Y^w + r_{33}Z^w + T_z$$

Por otro lado, sin tomar en cuenta las distorsiones radiales se puede notar $[X_c, Y_c, Z_c]$ en el marco de referencia de la imagen como:

$$x_{im} = -\frac{f}{s_x} \frac{X^c}{Z^c} + o_x, \quad y_{im} = -\frac{f}{s_y} \frac{Y^c}{Z^c} + o_y$$

Reemplazando X_c por su definición en función de X_w logramos escribir las coordenadas de la imagen directamente en función de las coordenadas del mundo:

$$x_{im} - o_x = -f_x \frac{r_{11}X^w + r_{12}Y^w + r_{13}Z^w + T_x}{r_{31}X^w + r_{32}Y^w + r_{33}Z^w + T_z}, \quad y_{im} - o_y = -f_y \frac{r_{21}X^w + r_{22}Y^w + r_{23}Z^w + T_y}{r_{31}X^w + r_{32}Y^w + r_{33}Z^w + T_z}$$

Asumiendo que las coordenadas del centro de la imagen son conocidas y utilizando estas definiciones junto a una cantidad suficiente de puntos de una grilla de calibración de tamaño y orientación conocidos, se pueden resolver los parámetros desconocidos.

4. Estereoscopía

4.1. Introducción

La visión estéreo se refiere a la habilidad de inferir información de la distancia y

estructura 3-D de una escena desde dos o más imágenes tomadas desde diferentes puntos de vista.

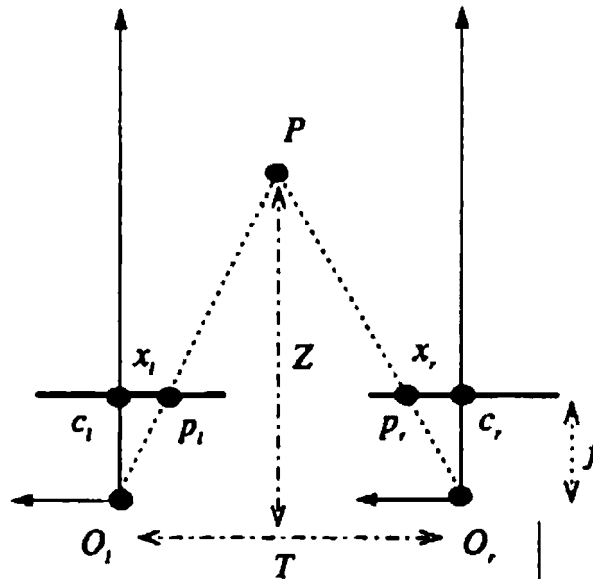


Figura 3.6 – Un sistema estéreo simple

En la figura 3.6 se puede observar un sistema estéreo básico construido a partir de dos cámaras pinhole. La forma en la que se determina la posición en el espacio de P es la triangulación, es decir, la intersección de los rayos definidos por los centros de proyección con p_l y p_r . En la figura se puede observar lo siguiente:

- P es el punto 3-D
- I_l y I_r son el plano de imagen izquierdo y derecho respectivamente (y son coplanares).
- O_l y O_r son el centro de proyección izquierdo y derecho respectivamente.
- p_l y p_r son las proyecciones de P en el plano de imagen izquierdo y derecho respectivamente.
- Los ejes ópticos son paralelos por lo que el *punto de fijación*, definido como el punto de intersección de los ejes ópticos, se unen infinitamente lejos de las cámaras.

- T es la distancia entre los centros de proyección O_l y O_r , y se le llama *línea base*.
- x_l y x_r son las coordenadas de p_l y p_r , con respecto a los puntos principales c_l y c_r .
- f es la distancia focal.
- Z es la distancia entre P y la *línea base*.

Los valores de los parámetros tales como f , T , c_l y c_r , se deben encontrar realizando una calibración estéreo. La misma estará constituida tanto por los parámetros extrínsecos como por los intrínsecos, y se realiza de forma análoga a lo visto anteriormente en el presente capítulo. La única diferencia es que el marco de referencia de una de las cámaras actuará como marco de referencia del mundo.

A continuación se explicará en detalle la denominada *geometría epipolar*, la cual es esencial para lograr una reconstrucción apropiada de la escena.

4.2 Geometría epipolar

La geometría epipolar corresponde a la geometría que relaciona una escena y dos proyecciones de ella. Es por esto que tiene una gran importancia en la reconstrucción tridimensional de la escena. A continuación se explica el concepto de geometría epipolar junto con la importancia y dificultad de establecer la correspondencia de puntos entre dos imágenes de una misma escena.

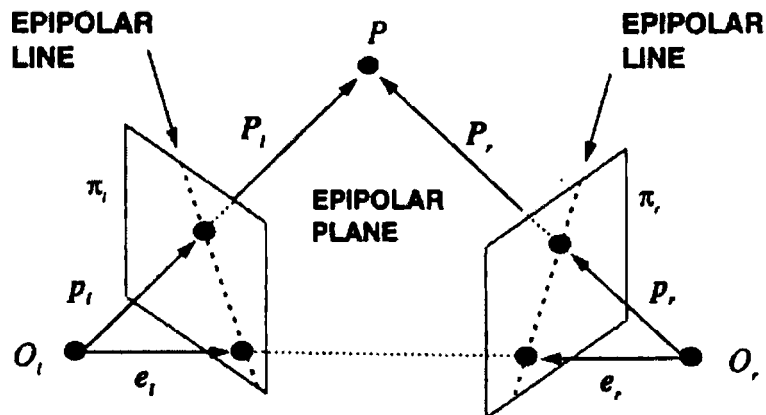


Figura 3.7 – Geometría epipolar

En la figura 3.7 se puede observar:

- Dos cámaras de tipo “pinhole”, sus centros de proyección, O_l y O_r , y sus planos imagen π_l y π_r . Las longitudes focales se notan como f_l y f_r .
- Cada cámara identifica un sistema de referencia tridimensional, el origen del cual se sitúa en el centro de proyección de la cámara; y el eje Z en el eje óptico.
- Los vectores $P_l[X_l, Y_l, Z_l]^T$ y $P_r[X_r, Y_r, Z_r]^T$ se refieren al mismo punto $P[X, Y, Z]$, como vectores en los sistemas de referencia de la cámara izquierda y derecha respectivamente.
- Los vectores $p_l[x_l, y_l, z_l]^T$ y $p_r[x_r, y_r, z_r]^T$ definen las proyecciones del punto P en la imagen izquierda y derecha respectivamente y están expresados en su correspondiente sistema de referencia. De esta forma, para todos los puntos de las imágenes se tiene $z_l = f_l$ o $z_r = f_r$, de acuerdo a la imagen que sea.
- El epipolo izquierdo es e_l y el epipolo derecho es e_r .

Los sistemas de referencia de las cámaras izquierda y derecha están relacionados a través de los parámetros extrínsecos. Estos definen una transformación rígida en el espacio tridimensional definida por:

- Un vector de traslación $T = O_r - O_l$.
- Una matriz de rotación R .

De esta forma, dado un punto en el espacio P , la relación entre P_l y P_r es:

- $P_r = R(P_l - T)$.

El nombre de geometría epipolar es debido a que los puntos en los cuales la recta que une los centros de proyección de las cámaras corta a los planos de proyección se llaman epipolos.

El epipolo e_l es la imagen del centro de proyección de la cámara derecha y el epipolo e_r es la imagen del centro de proyección de la cámara izquierda. Si se diera el caso de que uno de los planos imagen fuera paralelo a la recta que une los centros de proyección, el epipolo de ese plano estaría situado en el infinito.

La relación entre un punto del espacio tridimensional y su proyección se describe por las ecuaciones usuales de proyección de perspectiva, que en forma vectorial se escriben:

- $P_l = (f_l/Z_l)P_l$
- $P_r = (f_r/Z_r)P_r$

La importancia práctica de la geometría epipolar se desprende del hecho que el plano identificado por P, O_l, O_r llamado plano epipolar, interseca cada imagen en una línea llamada línea epipolar.

Dado p_l, P puede ubicarse en cualquier punto del rayo definido por O_l y p_l . Pero dado que la imagen de este rayo en la imagen derecha es la línea epipolar a través del punto correspondiente p_r , dicho punto debe estar sobre la línea epipolar.

Esta correspondencia establece una aplicación entre puntos de la imagen izquierda y rectas de la imagen derecha y viceversa. Una consecuencia de esta correspondencia es que dado que todos los rayos pasan por construcción por el centro de proyección, todas las rectas epipolares deben pasar por el epipolo.

Por lo tanto si determinamos la aplicación entre puntos de la imagen izquierda y las rectas epipolares de la imagen derecha, se puede restringir la búsqueda para el emparejamiento de p_l a lo largo de la línea epipolar correspondiente.

De la misma forma, si determinamos la aplicación entre puntos de la imagen derecha y las rectas epipolares de la imagen izquierda, se puede restringir la búsqueda para el emparejamiento de p_r a lo largo de la línea epipolar correspondiente.

Finalmente la búsqueda de las correspondencias se reduce a un problema de una sola dimensión.

Este conocimiento también se puede usar para verificar si una pareja de puntos son correspondientes o no. Por esto, es una técnica muy efectiva para detectar las posibles falsas correspondencias debidas a oclusión.

La restricción epipolar expresa entonces que los puntos correspondientes deben estar sobre líneas epipolares conjugadas.

5. Conclusiones

Los conceptos teóricos detallados en este capítulo presentan las bases para la triangulación de la pelota a lo largo de su trayectoria. Utilizando los conceptos de formación de la imagen, calibración y geometría epipolar es posible reconstruir la posición de la misma en el espacio tridimensional en cualquier instante de tiempo.

En el presente trabajo, la posición de la pelota es conocida en cada uno de los planos imagen. De esta forma, no es necesario calcular la correspondencia de puntos entre un plano y otro, sino que se puede pasar directamente al paso de reconstrucción del punto 3D.

Capítulo IV. La interfaz gráfica de usuario

1. Introducción

Dadas las características no convencionales de la aplicación implementada para el presente trabajo, los patrones clásicos de diseño de interfaces no aplican completamente a este caso en particular.

Este breve capítulo se centra en discutir la forma adecuada de desarrollo de una interfaz de usuario, más allá del tipo de aplicación a la que ésta pertenezca; los cuales pueden ser utilizados como lineamientos generales de diseño.

2. Definición

Las Interfaces Gráficas de Usuario, más conocidas como GUI por sus iniciales en inglés de Graphical User Interface han sido objeto de estudio desde hace más de cincuenta años.

Se trata de un tipo de interfaz que posibilita la comunicación entre humanos y computadoras. Es un sistema de convenciones gráficas y cinéticas mediante el cual podemos operar la computadora.

El concepto de "convención gráfica" es particularmente importante, ya que ha dominado los modelos sobre los cuales se desarrollaron las GUI durante años. Mediante el uso de diferentes metáforas (escritorio, cesto de basura, etc.) se establecieron las convenciones

que permiten a un usuario reconocer rápidamente muchos de los elementos y funciones de un sistema operativo. Hoy en día, las GUI poseen su propio metalenguaje y éste forma parte del imaginario simbólico de quienes usan computadoras.

Cuando se nos pregunta acerca de un programa determinado, inmediatamente asociamos su nombre al aspecto de su interfaz gráfica. Para el usuario, la interfaz es la aplicación.

El desarrollo de GUI efectivas y amigables fue el factor determinante para la popularización de las computadoras y de la informática en general. Gracias a ellas estas máquinas dejaron de ser un instrumento científico y se convirtieron en algo similar a un electrodoméstico.

En muchos casos la interfaz se convierte en una barrera debido a un pobre diseño y una escasa atención a los detalles de la tarea a realizar. Es por esto que, actualmente, cerca del 48% del código de una aplicación está dedicada a la interfaz; inclusive el tiempo de desarrollo dedicado a éstas se triplicó a lo largo de los años.

3. La Interacción Hombre-Computadora

El diseño de las GUI es parte de los estudios que se realizan en el campo de la Interacción Hombre Computadora (de ahora en más HCI, proveniente del inglés Human-Computer Interaction).

La ACM (Association for Computer Machinery) posee un grupo especial de trabajo en temas relacionados con la HCI; la cual propuso la siguiente definición para ésta:

La Interacción Hombre-Máquina es la disciplina relacionada con el diseño, evaluación e implementación de sistemas informáticos interactivos para el uso de seres humanos y con el estudio de los fenómenos más importantes con los que está relacionado.

El tema principal de esta disciplina está en la interacción entre uno o más seres humanos y una o más computadoras. Aunque la situación clásica sería una persona usando un programa sentado en su puesto de trabajo, alrededor de las palabras interacción, persona y computadora hay un amplio conjunto de opciones posibles.

Por ejemplo, las computadoras pueden formar parte de cabinas de avión, monitores de ruta en coches, teléfonos móviles, block de notas, libros electrónicos, sistemas de realidad virtual, etc. Por otro lado, los usuarios pueden formar parte de grupos u organizaciones y por tanto se necesitarán interfaces para sistemas distribuidos o de trabajo cooperativo.

La HCI surgió como un campo entrecruzado de disciplinas: gráficos por computadora, sistemas operativos, factores humanos, factores ergonómicos, ingeniería industrial, psicología cognitiva, y la ingeniería informática.

Los objetivos de la HCI son desarrollar o mejorar la seguridad, utilidad, efectividad, eficiencia y usabilidad de sistemas que incluyan computadoras.

Para hacer sistemas interactivos correctos hace falta:

- 1.- Comprender los factores psicológicos, ergonómicos, organizativos y sociales que determinan cómo la gente trabaja y hace uso de las computadoras.
- 2.- Desarrollar herramientas y técnicas para ayudar a los diseñadores a conseguir que los sistemas informáticos sean los idóneos para las actividades a las cuales se quieran aplicar.
- 3.- Conseguir una interacción eficiente, efectiva y segura, tanto a nivel individual como de grupo.

Para finalizar, es importante comprender que los usuarios no cambian radicalmente su manera de ser, sino que los sistemas deben estar diseñados para satisfacer los requisitos del usuario.

4. Aspectos de las GUI

En la vida cotidiana hay muchos ejemplos de interfaz. En el caso de la puerta, la manija de la puerta es la interfaz entre la puerta y la persona. El volante, el acelerador y otros instrumentos y herramientas son la interfaz entre un auto y el conductor. Es muy importante darse cuenta en un primer nivel que la interfaz refleja las cualidades físicas de las dos partes de la interacción. La manija está hecha de un material sólido y esta bien pegada a la puerta, la cual por otra parte, como tiene que interactuar con la mano, está puesta a la altura de ésta, y tiene la forma que se le adapta. Esta es una idea muy importante en el diseño que puede concretarse en dos conceptos:

- Correcta visibilidad del objeto al cual se realizará una acción.
- Comprensión intuitiva del objeto sobre el que se realizará la acción y cómo hacerlo.

Se detallarán a continuación algunos ejemplos de diseños que han fallado en cumplir dichos conceptos:

4.1. Ejemplo 1

La figura 4.1 muestra un extracto de una botonera de un ascensor. En la parte superior, los botones para dirigirse a un piso y su respectiva indicación se encuentran correctamente colocados; ayudando así a comprender el uso de dicho botón.



Figura 4.1 – Distribución de la botonera de un ascensor

En la parte inferior, esto no ocurre. Los botones y sus respectivas indicaciones son muy similares (igual tamaño y color); por lo tanto es fácil presionar una indicación en vez de un botón por error. Este caso es un ejemplo en el cual se rompe con el concepto de comprensión intuitiva.

4.2. Ejemplo 2

La figura 4.2 muestra el panel de un semáforo interactivo, el cual cambia de rojo a verde cuando un peatón lo requiere.



Figura 4.2 - ¿Cuál es el botón?

El cartel indica que hay que presionar un botón, pero en ningún lugar se indica la ubicación del mismo. La interfaz no es comprensible. Algún usuario podría preguntarse por qué no puede ser el botón el tornillo superior que sujeta la tapa de plástico. Nos encontramos en este caso con una carencia del concepto de visibilidad.

4.3. Ejemplo 3

Por último; un caso en el que los controles son visibles; y existe cierta comprensión intuitiva sobre qué acción ejecutará cada botón. La figura 4.3 muestra un equipo de audio con reproducción de CD y cassette.

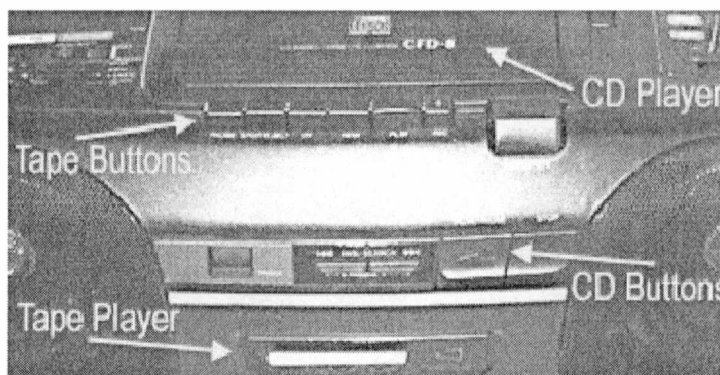


Figura 4.3 – Controles funcionales incorrectamente distribuidos

El problema aquí es que los botones se encuentran mal distribuidos. ¿Cuales son los botones para el reproductor de CD y cuales son los botones para el reproductor de cassettes? Los botones para el reproductor de CD se encuentran cerca de la cassette y los botones para el reproductor de cassettes se encuentran cerca de la unidad de CDs.

4.4. Ejemplo 4

No hay que olvidar además, que en el diseño de una GUI también se ha de tener en cuenta el entorno y el ámbito cultural.

Por ejemplo, el símbolo de reciclado de la figura 4.4, muy utilizado actualmente en los sistemas operativos gráficos para denotar la papelera de reciclaje, quizás no sea muy común en ciertas culturas. Para esta gente; la imagen no indica que en dicho lugar se encuentran los archivos y documentos eliminados.

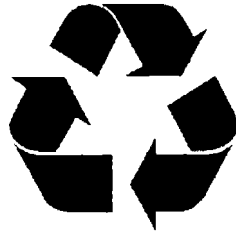


Figura 4.4 – Semántica no intuitiva según el contexto cultural

5. Usabilidad

La usabilidad es aquella característica que hace que el software sea fácil de utilizar y de aprender. Un software es fácil de utilizar si realiza la tarea para la que lo estamos usando. La facilidad de aprendizaje se puede medir por lo rápido que se realiza una tarea, cuantos errores se cometen y la satisfacción de la gente que lo utiliza.

También incluye aspectos como que sea seguro, útil y que tenga un costo adecuado. Una aplicación usable es la que permite que el usuario se concentre en su tarea y no en la aplicación.

La GUI es la puerta del usuario a la funcionalidad del sistema. Que las interfaces de usuario estén mal diseñadas es un factor que frena el uso de las funcionalidades. Por lo tanto es muy importante diseñar interfaces de usuario usables.

Desafortunadamente, se da en muchas ocasiones que los diseñadores evaden el tema de la usabilidad con excusas como “los usuarios no necesitan mejores interfaces, sino un entrenamiento mejor”, “la usabilidad es subjetiva”, “la usabilidad aumenta los costes de desarrollo y el tamaño del ciclo de desarrollo”.

5.1. Principios generales de la usabilidad

5.1.1. Facilidad de aprendizaje

Se requiere que sea mínimo el tiempo necesario que se requiere desde el no conocimiento de una aplicación a su uso productivo.

Para que el sistema sea fácil de aprender, éste debe ser:

Sintetizable: El usuario tiene que poder evaluar el efecto de operaciones anteriores en el estado actual. Es decir cuando una operación cambia algún aspecto del estado anterior, es importante que el cambio sea captado por el usuario.

Familiar: Es la correlación que existe entre los conocimientos que posee el usuario obtenidos a partir de experiencias interactivas con otros sistemas y los conocimientos requeridos para la interacción en un sistema nuevo.

5.1.2. Consistencia

Este es un concepto clave en la usabilidad de un sistema informático. Un sistema es consistente si todos los mecanismos que se utilicen en un sistema, sean siempre usados de la misma manera siempre que se utilicen y sea cuando sea el momento en que se haga.

A fin de lograr sistemas consistentes convendría:

- Seguir guías de estilo siempre que sea posible.
- Diseñar con una apariencia o “*look & feel*” común.
- No hacer modificaciones si no es necesario hacerlas.

- Añadir nuevas técnicas al conjunto preexistente, en vez de pedir que cambie las conocidas

5.1.3. Flexibilidad

La flexibilidad se refiere a la multiplicidad de maneras en que el usuario y el sistema intercambian información. Los parámetros que miden la flexibilidad son:

5.1.3.1. Control del usuario

Permite a los usuarios conducir la interacción y no estar forzados a trabajar para la aplicación. Para ésto se debe dar a los usuarios por ejemplo la posibilidad de deshacer y el control para poder comenzar y detener las operaciones siempre que sea posible. Cuando el proceso no se pueda interrumpir, advertir a los usuarios y visualizar mensajes apropiados durante el proceso.

5.1.3.2. Migración de tareas

La migración de tareas está relacionada con la transferencia del control, entre el usuario y el sistema. Tanto el usuario como el sistema han de poder pasar de una tarea a la otra o promocionarla, de manera que pueda ser completamente interna o compartida entre los dos.

5.1.3.3. Capacidad de sustitución

Permitir que valores equivalentes pueden ser substituidos los unos por los otros. Por ejemplo si queremos introducir el valor que determina el margen de una carta, se puede preferir entrarlo en centímetros o en pulgadas. Eliminando cálculos innecesarios al usuario, se pueden minimizar errores y esfuerzo.

5.1.3.4. Adaptabilidad

Es la adecuación automática de la interfaz al sistema. Las decisiones para llevarla a cabo pueden estar basadas en la experiencia del usuario o en la observación de la repetición de ciertas secuencias de tareas. Se puede preparar a un sistema para reconocer

el comportamiento de un experto o de un usuario inicial y de acuerdo con esto ajustar automáticamente el control del diálogo o el sistema de ayuda con tal de adaptarlo a las necesidades del usuario actual.

5.1.3.5. Técnicas para lograr flexibilidad

Permitir a los usuarios suspender una acción y comenzar otra para atender un trabajo inesperado. Tener atajos y soporte de navegación por teclado para no tener que utilizar siempre el mouse. Facilitar la realización de tareas repetitivas.

5.2. Otros aspectos a tener en cuenta

La *robustez* de una interacción usuario máquina, cubre las características para poder cumplir los objetivos de ésta.

La *recuperabilidad* es el grado de facilidad de una aplicación para permitir al usuario corregir una acción una vez reconocido un error.

El *tiempo de respuesta*, el cual se define como el tiempo que necesita el sistema para expresar los cambios de estado del usuario. Es importante que los tiempos de respuesta sean soportables para el mismo.

Reducir la *carga cognitiva*: los usuarios tienen que confiar más en los reconocimientos que en los recuerdos, y no tienen que recordar abreviaciones y códigos muy complicados.

6. Metáforas

En el diseño de las interfaces actuales, las metáforas tienen un papel dominante. La metáfora se utiliza cuando queremos comunicar un concepto abstracto de una manera familiar y accesible.

La metáfora del escritorio introducida por la computadora Macintosh y de uso generalizado actualmente, supuso un cambio en la usabilidad de las computadoras. En esta metáfora los objetos en la pantalla, los nombres que se dan a las órdenes de comando, conceptos como escritorio, icono, menús, ventanas están basados en temas familiares y lo mismo sucede con las acciones a realizar, arrastrar, soltar, pegar, etc.

Cuando nos encontramos delante de una nueva herramienta tecnológica como una computadora, se tiende a compararla con alguna cosa conocida. Un ejemplo clásico es comparar a ésta con una máquina de escribir de una manera metafórica. Cuando se utiliza un editor de texto pensamos que se comporta igual que una máquina de escribir. Estos enlaces metafóricos proporcionan los fundamentos por los cuales el usuario desarrolla su modelo de computadora propio. Los conocimientos de los elementos y sus relaciones en un dominio familiar se traspasan a los elementos y sus relaciones en un dominio no familiar.

La metáfora visual normalmente es una imagen que nos permite representar alguna cosa y que el usuario puede reconocer lo que representa y por extensión puede comprender el significado de la funcionalidad que recubre. Las personas entendemos las metáforas por *intuición*, lo que significa “conocimiento de una cosa obtenida sin recurrir a inferencia o razonamiento”.

Las metáforas se basan en asociaciones percibidas de manera similar por el diseñador y el usuario. Como hemos visto, si el usuario no tiene la misma base cultural que el desarrollador es fácil que la metáfora falle, incluso teniendo la misma base cultural puede haber faltas de comprensión importantes.

Como ejemplo de la limitación de las metáforas podemos nombrar al portapapeles. Éste es un componente del sistema operativo que nos permite pasar información entre aplicaciones. Es fácil de intuir como utilizar el portapapeles, pero si nos limitamos estrictamente a la metáfora es muy pobre. No se puede recoger más de una cosa, no se

sabe de donde vienen los documentos u objetos, no se sabe todo lo que ha ido recibiendo. Todas estas posibilidades no están cubiertas por la metáfora y por tanto tienen que ser aprendidas. Esta metáfora, presenta problemas de cobertura.

7. Recomendaciones

Como un lineamiento general, se considera importante para una correcta interacción entre el usuario y la aplicación:

- 1) Facilitar la visibilidad y limitar el cansancio visual diseñando desde parámetros ergonómicos que faciliten y permitan centrarse en el contenido y no en el formato.
- 2) Utilizar diálogos simples y naturales para facilitar que la interfaz sea lo más simple posible, con un lenguaje lo más cercano a los usuarios. La información verbal debe estar expresada en palabras y frases claras, utilizando conceptos familiares.
- 3) Reducir la memorización de modo que las instrucciones de uso del sistema sean visibles y fácilmente recuperables cuando sea necesario, evitando que los usuarios tengan que hacer esfuerzos de memoria para recordar la información que hubiera desaparecido al saltar a otra página de ayuda.
- 4) Reducir la complejidad de las acciones de manera que los usuarios no tengan que enfrentarse a situaciones o acciones diferentes para una misma función, así será más sencilla su tarea. Los productos interactivos, deben responder a las actuaciones de los usuarios de un modo predecible. Se debe mantener al usuario informado sobre lo que sucede en la aplicación en cada momento del proceso, para evitar que se sienta perdido.
- 5) Proporcionar atajos, a modo de funciones rápidas. Para limitar la sensación de desorientación del usuario, es aconsejable mantener sistemas de localización y salto

idénticos en toda la estructura del sistema, a fin de que el usuario sepa en todo momento donde se encuentra.

6) La aplicación interactiva debe permitir que el usuario controle la velocidad con que interactúa con la información, sin sentir el desplazamiento y sin tener que sufrir esperas excesivas o velocidades extremas que le provoquen ansiedad.

7) Incluir mensajes de error en un lenguaje lo más natural posible (sin utilización de códigos) haciendo referencia explícita al problema y sugiriendo, de forma constructiva, una solución. De todas formas, un adecuado diseño de interacción debería anticiparse a los posibles errores de los usuarios.

8) Mostrar opciones de ayuda de modo que aparezca cierta información fácilmente localizable y enfocada a ofrecer a los usuarios listas de actuaciones posibles para realizar una función o una localización.

9) Agrupar los datos lógicamente de un modo coherente y a lo largo de las diferentes pantallas. La información debe organizarse jerárquicamente en diferentes niveles que vayan de los más general a lo más particular. Es importante mostrar sólo la información necesaria en un momento determinado y excluir la innecesaria.

10) Flexibilizar la presentación de la interfaz, para que los elementos que aparecen en pantalla puedan modificarse y personalizarse a fin de satisfacer las preferencias de los usuarios.

8. Conclusiones

Un correcto diseño de la interfaz gráfica de usuario no solo simplifica la tarea al usuario, sino que facilita además la metodología de desarrollo; estructurando la implementación de manera adecuada.

Como se detallará posteriormente, los conceptos y lineamientos repasados en el presente capítulo serán volcados en el desarrollo del prototipo.

Parte II

IMPLEMENTACIÓN

Capítulo V: Análisis preliminar

1. Introducción

Una vez estudiados los conceptos y fundamentos necesarios para el desarrollo de la aplicación, se inició la etapa de análisis con el fin de determinar la mejor alternativa y lograr así una correcta selección de hardware; estableciendo además los lineamientos generales de la implementación del software.

El presente capítulo detalla la metodología utilizada para determinar la óptima ubicación de las cámaras para la captura de los tiros, así como las características que las mismas debían cumplir. Esto es, establecer los parámetros intrínsecos y extrínsecos de los dispositivos de captura.

Primeramente se justifican las decisiones tomadas con respecto al posicionamiento de las cámaras para lograr una captura óptima, indentando maximizar la calidad y campo de visión de la misma.

Posteriormente se detalla el proceso de análisis de requerimiento que derivó en las especificaciones de hardware de las cámaras, así como la metodología de captura para su posterior procesamiento.

2. Ubicación de las cámaras

Con el fin de agilizar las estimaciones iniciales, un modelo virtual de la zona de captura de los tiros fue desarrollado mediante el programa 3D Studio. Además de poder determinar los requerimientos técnicos que debían cumplir las cámaras y su posición, gracias a este modelo se lograron simular pruebas de calibración de las mismas y de seguimiento de la pelota.

Las figuras 5.1 y 5.2 muestran la primera versión del modelo desarrollado (mallado y renderizado respectivamente). A su vez, se dispusieron varias pelotas de fútbol de tamaño estándar con el fin de verificar que en distancias lejanas a la cámara, las mismas todavía puedan ser correctamente capturadas.

En primera instancia, se ubicaron dos cámaras en la parte posterior de la zona de captura para realizar el seguimiento de la trayectoria de la pelota. Las mismas conformarían un sistema estereo que se encargaría de capturar la totalidad del tiro. Luego de varias pruebas, se determinó que la mejor opción en cuanto a capacidad de captura y triangulación se daba ubicando las cámaras de manera ortogonal entre si. Sin embargo, el uso de un único sistema estereo presentaba dos inconvenientes.

El primero se refiere a la posibilidad de oclusión en los cuadros iniciales de la secuencia de video de un tiro, dado que quien realice el mismo puede llegar a ubicarse entre la pelota y la cámara. De esta manera podría estar perdiéndose información vital, como por ejemplo los datos necesarios para el cálculo de la estimación de la trayectoria basándose en la información obtenida al inicio del tiro.

El segundo radicaba en que las pruebas de calibración realizadas (que se detallarán posteriormente en el presente trabajo) demostraron que la precisión de triangulación disminuía notoriamente en la zona más alejada a las cámaras (cerca de la línea de fondo). Idealmente, la precisión en la triangulación debe ser lineal a lo largo de toda la trayectoria. Utilizando sólo un sistema estéreo, el error aumentaba exponencialmente con la distancia; lo cual era inaceptable a los fines prácticos de la aplicación.

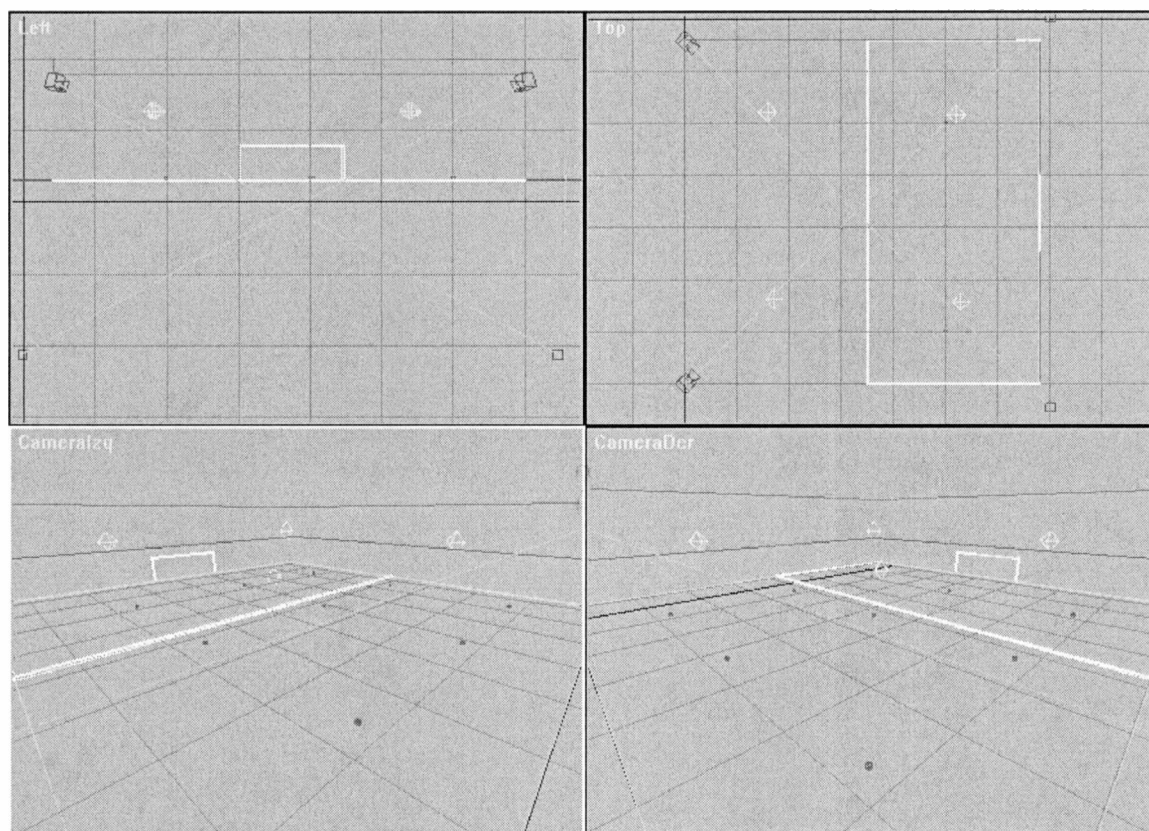


Figura 5.1 – Primera versión del modelo virtual 3D
(1) vista posterior del modelo. (2) vista superior del modelo
(3) vista desde la cámara izquierda. (4) vista desde la cámara derecha

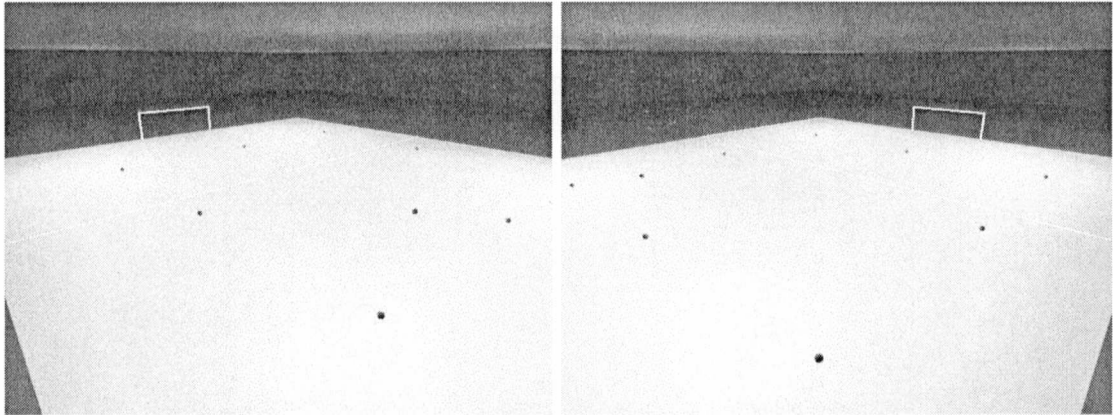


Figura 5.2 – Imágenes renderizadas a partir del modelo 3D
(1) vista de cámara izquierda (2) vista de cámara derecha

Se contemplaron dos alternativas con el fin de paliar el segundo problema: utilizar cámaras con mayor resolución o colocar otras cámaras que capturen la zona final del tiro.

La primer alternativa era utilizar cámaras con mayor resolución. Esta solución tenía por contra la necesidad de contar con una interfaz de comunicaciones hacia la computadora no convencional, dado el volumen de información a transmitir debido al tamaño de las imágenes que componen la secuencia de video. Esta solución además no resolvía el problema de la oclusión y aumentaba el costo del prototipo.

Es por esto que se decidió llevar a una nueva configuración con cuatro cámaras en lugar de dos. Éstas poseen una resolución de captura estándar por lo que no presentan complicaciones de uso e implementación. De esta manera, la precisión en la triangulación es mayormente lineal.

Por lo tanto, dos sistemas estéreo serán utilizados para el cálculo de trayectoria. El primero se encargará de capturar la parte inicial del tiro, mientras que el segundo la parte final.

La figura 5.3 muestra la configuración definitiva con ambos sistemas estéreo (cámaras S1CI y S1CD del sistema estereo 1 y cámaras S2CI y S2CD del sistema estereo 2) junto con la cámara encargada de realizar la captura de la pelota para el cálculo de rotación (CR), la cual es ubicada en la zona cercana al arco.

Como se explicó anteriormente, las cámaras de cada sistema estereo fueron ubicadas con un ángulo cercano a los 90 grados entre sí ya que de esta manera se obtiene el menor nivel de error posible sobre la trayectoria completa de los tiros. Esta configuración además es la que permite el mayor campo de visión de la zona de captura.

Las pruebas realizadas, que se detallan en el capítulo 10, confirman que esta disposición es la que mejor resultados presenta en ambos aspectos con respecto a las restantes alternativas probadas.

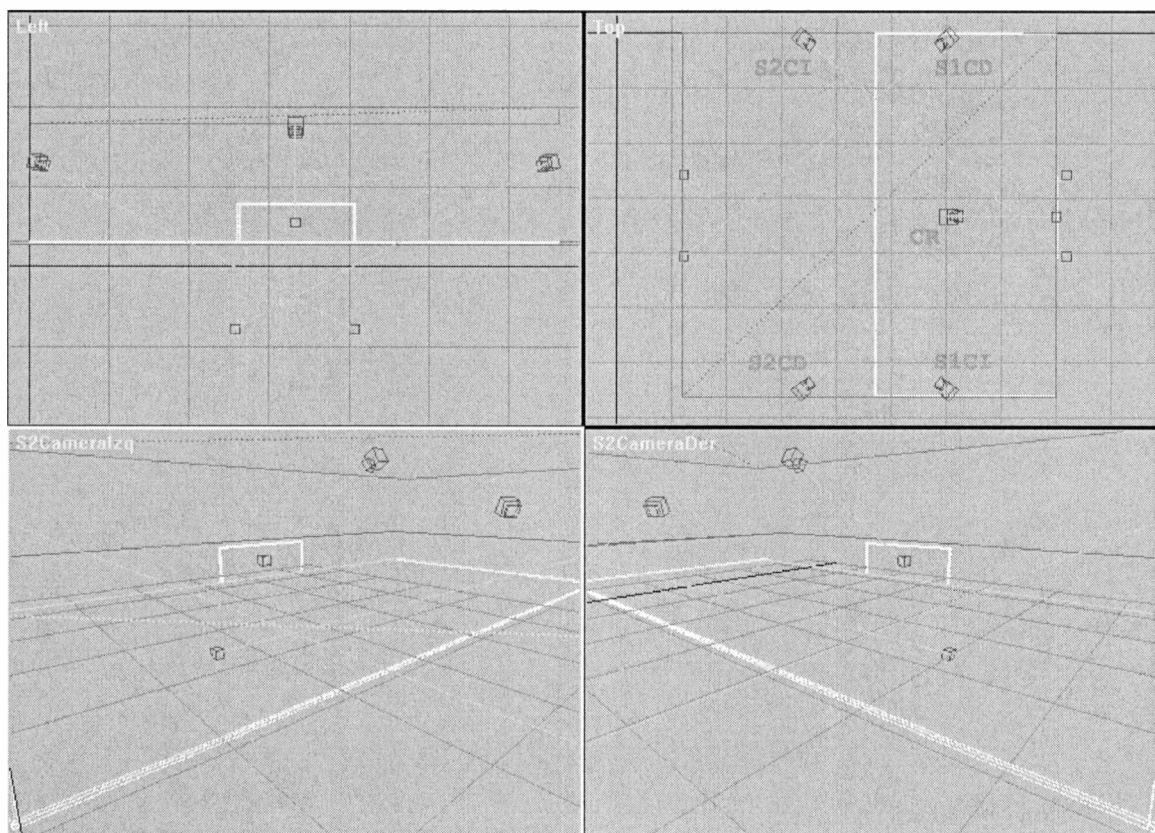


Figura 5.3. Configuración de cámaras utilizada
(1) vista posterior del modelo. (2) vista superior del modelo

(3)(4) vistas desde la cámaras izquierda y derecha respectivamente del sistema 1

Al contar con 2 sistemas estereo de captura, fue posible colocar las cámaras de manera tal que se logre minimizar el error de triangulación y abarcar además el mayor campo de visión posible, modificando la distancia focal de las lentes a ser utilizadas.

El problema de oclusión también fue resuelto debido a que quien realice el tiro siempre se encontrará por detrás del sistema 1, el cual realiza la captura del inicio del tiro. Para el momento en que la pelota se encuentre en la zona de captura del sistema 2, el ejecutor del tiro ya no podrá interponerse entre la pelota y la cámara.

Utilizar dos sistemas estereo en lugar de uno no incide mayormente en los costos ya que se corre con la ventaja de utilizar cámaras con una resolución relativamente baja.

3. Distancia focal de las lentes

Como se ha explicado en el capítulo 1, el campo de visión con la que cuenta una cámara se encuentra directamente relacionada con la longitud focal de la lente. A menor longitud, mayor campo de visión y viceversa.

La figura 5.4 muestra distintas tomas de la misma escena renderizadas con diferente longitud focal. Como puede apreciarse, una lente de 28mm no permite abarcar el campo de visión necesario, mientras que una lente de 20mm genera una distorsión geométrica en la imagen resultante demasiado notoria, llevando a posteriores errores de triangulación; e incluso impidiendo determinar la posición de la pelota.

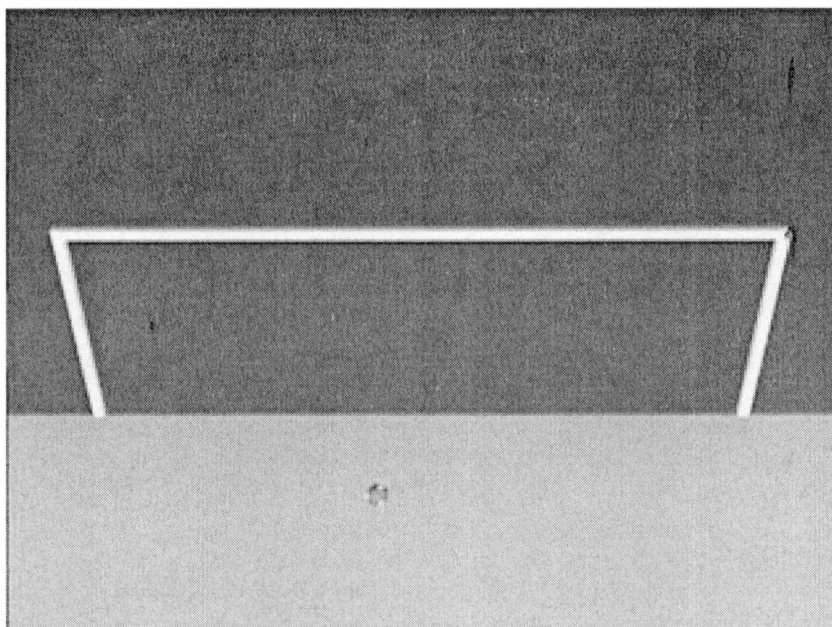


Figura 5.5 – Vista desde la cámara de rotación.

Dichas marcas son las que permitirán realizar el procesamiento para el cálculo de rotación de la pelota. De utilizar para esta cámara una lente con distancia focal muy corta, la perspectiva obtenida sería demasiado elevada, con lo cual la relación píxels/cm no sería mayormente homogénea entre las capturas en las cuales la pelota se encuentra cerca de la cámara y las que ésta se encuentra más alejada.

Por supuesto que utilizando una distancia focal excesivamente larga impediría capturar de manera completa la zona del arco, con lo cual esta alternativa debe ser eliminada, optando entonces por una lente de distancia focal intermedia; en la que se obtiene entonces el balance adecuado entre distorsión perspectiva y ángulo de visión.

4. Utilizando cámaras

A fin de corroborar las conclusiones obtenidas de la discusión previa, se realizaron pruebas con una cámara digital sobre una zona con un área similar a la estipulada para

la captura. Dichas pruebas confirmaron las estipulaciones iniciales referidas a resolución, distancia focal y velocidad de captura.

4.1. Resolución

En cuanto a la resolución, se verificó que en las capturas realizadas se observe la pelota incluso en la mayor distancia posible entre la cámara y la pelota. La figura 5.6 muestra esta situación. En el recuadro de la misma se encuentra ampliada la zona de interés, para facilitar la observación de este detalle.

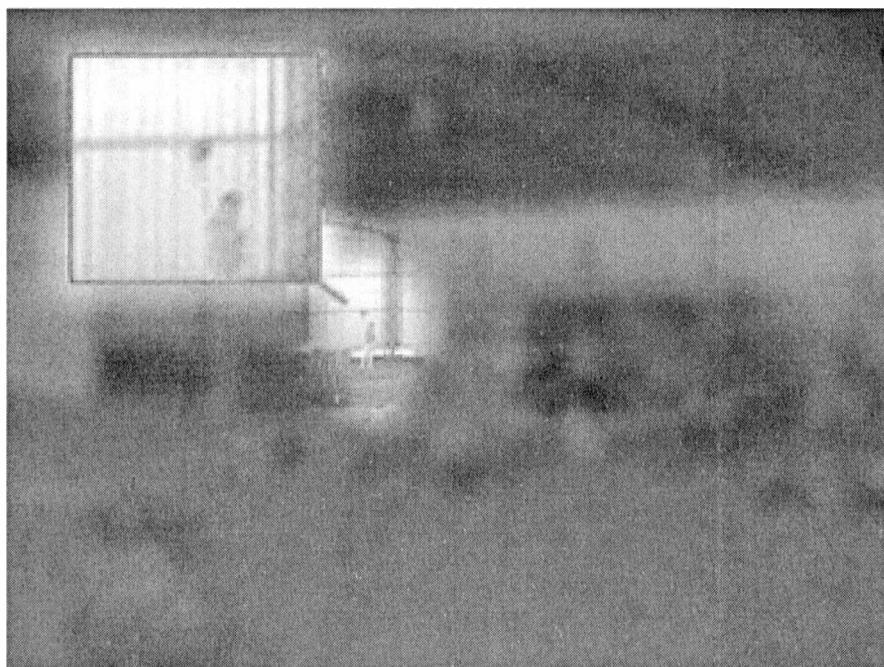


Figura 5.6 – Mayor distancia de captura.

La imagen de la figura 5.6 fue tomada a una resolución de 640x480 píxeles. Ya que con esta resolución la pelota es correctamente divisada incluso en la mayor distancia, y dado que la misma es un standard en los dispositivos de captura, este parámetro fue utilizado como base para la mayoría de las pruebas realizadas; una vez que se determinó el uso de dos sistemas estéreo.

4.2. Distancia focal

Con respecto a la distancia focal, se debía corroborar que la misma abarque completamente la zona de captura, tal como se muestra en la figura 5.7. Sobre esta zona, se consiguió capturar solo parte del área de captura y del arco. La distancia focal con la que se contó para esta toma fue de 35 mm, lo cual indicaba la necesidad de un mayor campo de visión, tal como se había determinado previamente.

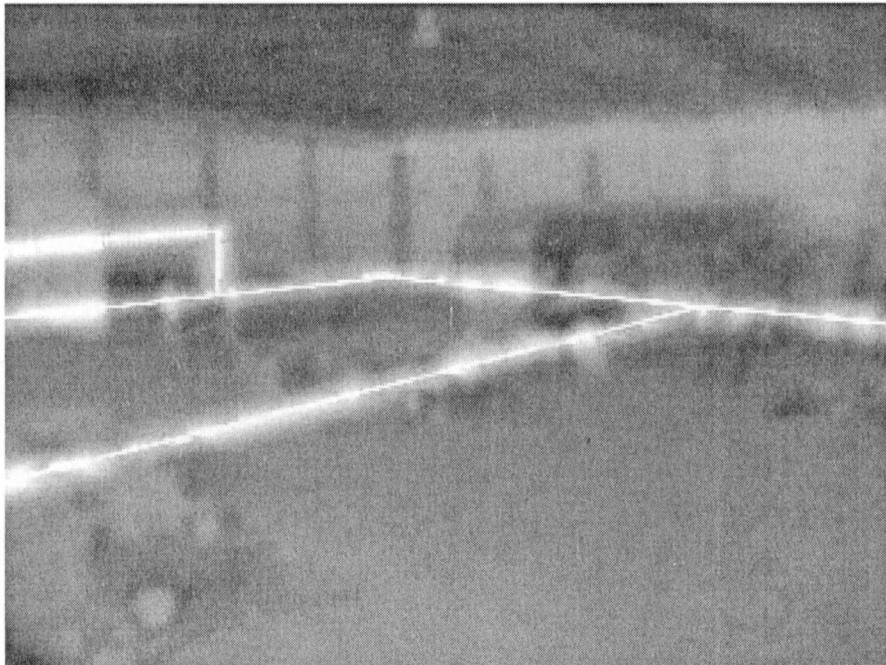


Figura 5.7 – Area de captura

Dado que en las capturas realizadas para el cálculo de la trayectoria únicamente se requiere determinar la ubicación del centro de la pelota en cada cuadro de las capturas de cada cámara a lo largo del tiro, una perspectiva relativamente alta no presenta un problema. Es por esto que utilizar lentes con distancia focal relativamente corta no será una contrariedad para el proceso de triangulación.

Debido a la utilización de dos sistemas estereo en lugar de uno, a lo largo de todo el trayecto desde el inicio hasta el final del tiro, la pelota nunca se encontrará a una

distancia tan amplia como para que la perspectiva generada por la distancia focal de las lentes incidan fuertemente en la precisión de la triangulación.

4.3. Velocidad de captura

Una última prueba concerniente con la velocidad de captura se llevó a cabo con el fin de determinar la velocidad de obturación necesaria a fin de evitar el efecto de borronado que se observa en la figura 5.8. Dentro de la zona indicada con un círculo se puede observar la *estela* que dejó la pelota en la captura. La velocidad de captura para este ejemplo fue de 1/15 segundos.



Figura 5.8 – La pelota no se observa completamente definida debido a la baja velocidad de obturación

Dado el tipo de aplicación, es obvia la necesidad de utilizar velocidades de obturación superiores a 1/500 de segundo. Sin embargo, se debe contemplar el factor de luminosidad en el área de captura; ya que éste puede llegar a limitar el tiempo de exposición. Incluso aumentando la sensibilidad del dispositivo de captura, y utilizando

una lente con gran apertura, esto puede no ser suficiente para evitar una subexposición en la captura.

Es deseable entonces, mantener el nivel ISO dentro de un nivel aceptable de ruido, maximizando la apertura del diafragma lo más posible y llevando el tiempo de exposición a la fracción de tiempo más corta para un nivel de luminosidad aceptable.

5. Requerimientos de las cámaras

Las pruebas en computadora y reales llevadas a cabo permitieron llegar a una serie de conclusiones en cuanto a los requerimientos de las cámaras. A modo de resumen, se detalla a continuación las características que deben poseer los dispositivos de captura para el cálculo de la trayectoria y rotación de la pelota.

- Las cámaras deben poseer un CCD con una resolución de 640x480 píxeles. Con esto se logra capturar la pelota incluso en la mayor de las distancias, minimizando el tamaño de las imágenes obtenidas y consiguiendo además un error aceptable en la triangulación del objetivo.

- La velocidad de captura debe ser de 100 cuadros por segundo, a fin de poder triangular la pelota cada 35 cm. aproximadamente. Con esta velocidad se logra además que la rotación de la pelota entre un cuadro y el siguiente no sea mayor a 1/10 de revolución¹. De esta forma, se está garantizando que las marcas sobre la pelota utilizadas para realizar el cálculo de rotación no se pierdan en los cuadros anterior y posterior al determinado para realizar esta tarea. Esto será explicado en detalle en el capítulo 8: “Determinación de la velocidad de rotación de la pelota”.

¹ La velocidad máxima en un tiro libre se encuentra cerca de los 35 metros por segundo, mientras que la velocidad máxima de rotación de la pelota no supera las 10 revoluciones por segundo

· Se requiere una velocidad de obturación de 1/1000 segundos a fin de obtener un efecto de borronado despreciable: el desplazamiento de la pelota en dicho intervalo de tiempo no supera los 2 cm, y la rotación de la misma no es mayor a 7 mm. A nivel pixel, este movimiento es minimamente capturado por la cámara, y la PSF de movimiento resultante es prácticamente nula.

· Las lentes deben poseer una distancia focal de 25 mm para las cámaras encargadas de la captura de la trayectoria y 50 mm para la cámara de rotación, logrando con esto un ángulo de visión de 73 y 40 grados respectivamente; suficientes para abarcar el área de captura buscada en cada caso. Idealmente, las mismas deben permitir una apertura del diafragma superior a $f/2$ para lograr un buen nivel de luminosidad.

6. Interfaz de comunicación y formato de captura

Teniendo en cuenta la velocidad de captura y resolución de video, se optó por utilizar la interfaz Firewire 1394a. Las ventajas de ésta son varias. Al ser una interfaz digital, la captura no resulta en pérdidas de señal dado que en ningún momento hay conversión analógico-digital. Por otro lado, la posibilidad de una posterior expansión a Firewire 1394b en caso de ser necesario hace de ésta una opción ideal.

Dado que capturar 100 cuadros de 640 x 480 píxeles a 24 bits por píxel por segundo requiere mínimamente una tasa de transferencia de 87 Megabytes por segundo² se decidió realizar la captura de los videos en formato RAW, a fin de minimizar en ancho de banda tanto en la comunicación como en el tiempo de almacenamiento del video.

Al capturar de esta forma, se está leyendo la información tal cual es capturada por el CDD según el Bayer Mosaic descrito en el capítulo 1. El proceso de interpolación

² $640 \times 480 \text{ pixels} * 24 \text{ bits por pixel} * 100 \text{ FPS} = 921600 \text{ bytes} * 100 \text{ FPS} = 87 \text{ Mbytes/seg}$

correspondiente para la conversión a RGB es entonces delegado a la capa de software. De esta forma, la tasa de transferencia se reduce a 29 Megabytes por segundo³.

A modo de ejemplo, la figura 5.9 muestra una captura en formato RAW y su correspondiente conversión a formato RGB.



Figura 5.9 – Reconstrucción RAW a RGB

7. Arquitectura de captura

Idealmente, la configuración de hardware mas completa en la que se pensó inicialmente es la que se muestra en la figura 5.10. La zona verde de la misma es el área de captura mientras que la zona gris se encuentra fuera de ésta. El arco se encuentra en la zona superior de la imagen.

³ $640 \times 480 \text{ pixels} * 8 \text{ bits por pixel} * 100 \text{ FPS} = 307200 \text{ bytes} * 100 \text{ FPS} = 29 \text{ Mbytes/seg.}$

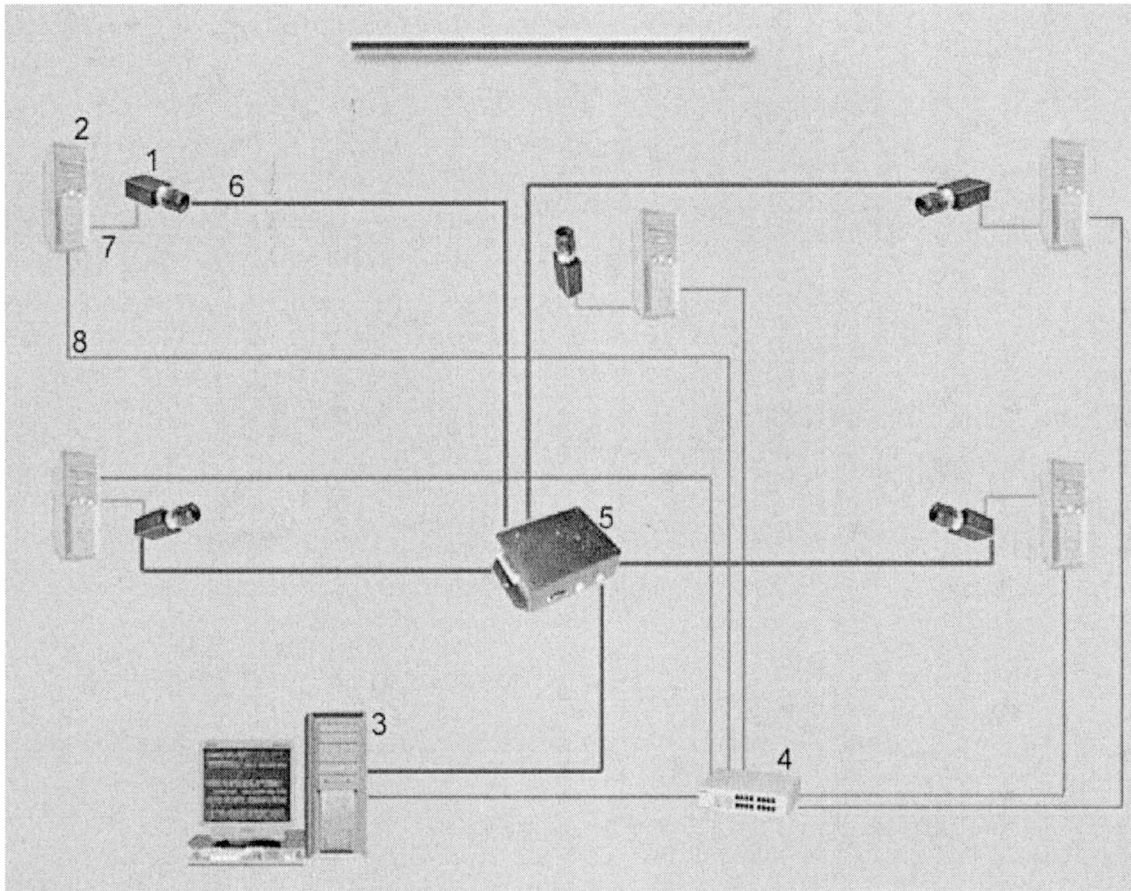


Figura 5.10 – Arquitectura de captura

Las configuración presenta la siguiente arquitectura de hardware, indicado con los correspondientes números en la figura 5.10. Cuatro cámaras (1) para el cálculo de la trayectoria, más una quinta para el cálculo de rotación. Éstas se encuentran conectadas a sendas computadoras (2) mediante interfaz Firewire (7); cada una de éstas almacenando el video capturado. Las mismas se encuentran conectadas mediante una red Ethernet (4)(8) a la computadora central (3), encargada de realizar el procesamiento de los videos. La sincronización de iniciación de captura es enviado mediante un circuito exclusivo para esta tarea (5)(6), disparado por la computadora central.

Una alternativa a esta configuración, a fin de reducir el costo del mismo, es utilizar una computadora para realizar la captura de dos cámaras simultáneamente. Dado que se esta recibiendo la información en formato RAW, esto podría llegar a ser posible implementar. La computadora central podría ser la encargada de realizar la captura del

video para el cálculo de rotación, para posteriormente realizar el procesamiento completo (trayectoria y rotación).

Obviamente, la configuración de hardware adoptada, no debe modificar en absoluto la solución de software implementada; permitiendo un nivel de abstracción intermedio entre la arquitectura de captura y el sistema de procesamiento.

8. Librería de procesamiento de imágenes

Entre las librerías de procesamiento de imágenes disponibles para el desarrollo de la aplicación, se determinó que la Open Source Computer Vision Library (conocida como Open CV) es la que más se adecua a los requerimientos especificados; tanto en funcionalidad como en performance.

Esta librería es una colección de algoritmos con la finalidad de facilitar la tarea de resolución de problemas de visión por computadora en general. La misma se encuentra optimizada para microprocesadores con tecnología Intel®.

La librería se divide en cuatro módulos principales:

- CV library: Brinda las funciones tales como: procesamiento de imágenes básicos, análisis estructural, análisis de movimiento y seguimiento de objetos, reconocimiento de patrones y calibración de cámara.
- CXCORE: Las operaciones sobre arrays y estructuras dinámicas, funciones de dibujo y persistencia de datos se presentan en este módulo.
- HighGUI: Permite la interacción de entrada/salida de datos (carga y almacenamiento de imágenes, entrada y salida de video, manejo de ventanas para la interfaz gráfica).

- **Experimental:** Este módulo presenta las funciones que se encuentran en etapa experimental, tales como: funciones de correspondencia estéreo, funciones de morfosis, funciones de seguimiento 3D, modelos de Markov, etc.

A modo de ejemplo se muestra la forma de interacción con la librería. Dado que OpenCV prioriza performance sobre simplicidad, la forma de acceso a la información contenida en las imágenes es la siguiente.

Ejemplo 1. Sea *img* una imagen en escala de grises(24bpp), el acceso al píxel situado en la posición (x,y) se logra mediante:

```
I(x, y) ~ ((uchar*)(img->imageData + img->widthStep*y))[x]
```

Ejemplo 2. Sea *img* una imagen color basado en RGB (24bpp), el acceso a cada canal del píxel situado en la posición (x,y) se logra mediante:

```
((uchar*)(img->imageData + img->widthStep*y))[x*3]    (canal B)
((uchar*)(img->imageData + img->widthStep*y))[x*3+1]  (canal G)
((uchar*)(img->imageData + img->widthStep*y))[x*3+2]  (canal R)
```

Ejemplo 3. Incrementar el brillo del píxel (100,100) en 30 unidades de *img* puede realizarse de la siguientes manera:

```
CvPoint pt = {100,100};
((uchar*)(img->imageData + img->widthStep*pt.y))[pt.x*3] +=
30;
((uchar*)(img->imageData + img->widthStep*pt.y))[pt.x*3+1] +=
30;
((uchar*)(img->imageData + img->widthStep*pt.y))[pt.x*3+2] +=
30;
```

o más eficientemente de la siguiente manera:

```
CvPoint pt = {100,100};
uchar* temp = &((uchar*)(img->imageData + img->widthStep*pt.y))[x*3];
temp[0] += 30;
temp[1] += 30;
temp[2] += 30;
```

9. Conclusiones

El análisis preliminar permitió no solo estudiar las alternativas tecnológicas para llevar a cabo el desarrollo, se logró además conocer las posibles limitaciones y complicaciones a enfrentar a lo largo de la implementación.

El conocimiento previo de las características necesarias de hardware a utilizar y las técnicas de software a implementar facilitan el correcto planeamiento y puesta en marcha del desarrollo.

Capítulo VI: Módulo de calibración de cámaras

1. Introducción

El proceso de calibración de cámaras es uno de los más importantes dentro de la aplicación desarrollada. Esto se debe a que una incorrecta calibración de cámaras puede llevar a errores de precisión al momento de la triangulación de puntos en el espacio.

El presente capítulo detalla primeramente el análisis previo realizado mediante una herramienta de calibración desarrollada para Matlab. Posteriormente se explica la aproximación a la solución implementada; y por último, el módulo de calibración de cámaras finalmente desarrollado.

2. Toolbox de calibración para Matlab

En primera instancia se estudió en detalle un toolbox de calibración de cámaras desarrollado para Matlab. El mismo fue de gran utilidad para el entendimiento de los pasos prácticos que se deben llevar a cabo para realizar una calibración de cámaras completa.

En dicha herramienta, con el fin de lograr una correcta calibración, se deben seguir una serie de pasos que se encuentran agrupados en diferentes funciones. Los pasos a seguir son los siguientes:

- Indicación del nombre de las imágenes a leer. Cada una de estas imágenes debe ser una vista de la grilla o patrón de calibración.

- Lectura de las imágenes de la grilla.
- Extracción de las esquinas de la grilla.
- Calibración.

Además de éstas, la herramienta cuenta con otras funciones que sirven para visualizar y analizar la precisión de la calibración efectuada. En la figura 6.1 se puede observar la interfaz de usuario con la que cuenta el toolbox.

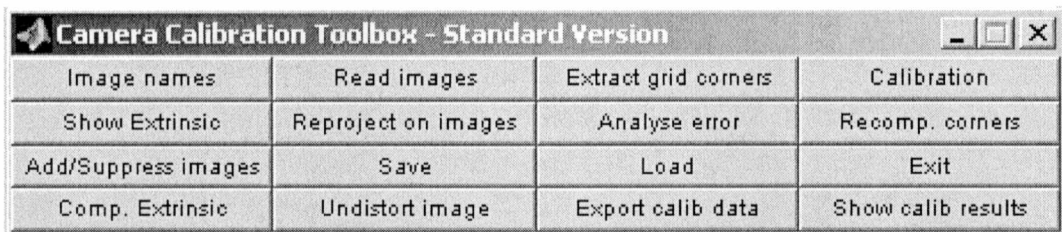


Figura 6.1 - Toolbox de calibración de cámaras para Matlab

Cabe destacar que ciertos pasos de la calibración no son automáticos. En el proceso de extracción de esquinas, el usuario debe clickear sobre las cuatro esquinas de cada una de las imágenes de la grilla, comenzando por la esquina que será el origen de las coordenadas. Este proceso puede observarse en la figura 6.2.

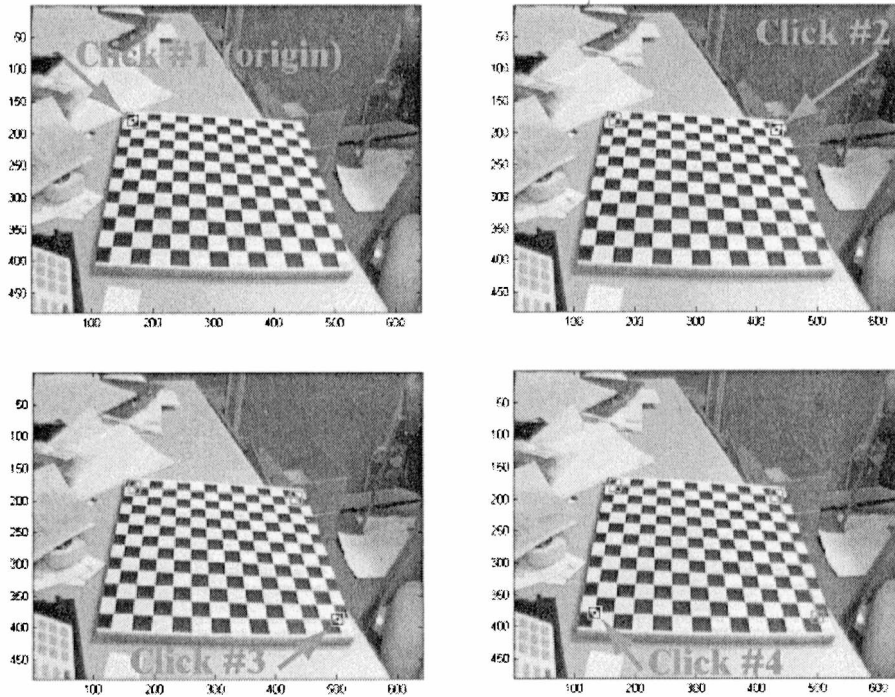


Figura 6.2 - Proceso de extracción de esquinas

Una vez indicadas las cuatro esquinas de la grilla, el toolbox estará listo para detectar las intersecciones internas. En la figura 6.3 se pueden observar todas las intersecciones encontradas de la grilla de calibración de la figura anterior.

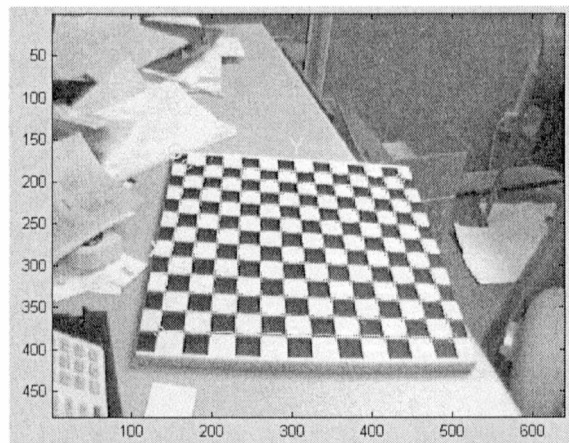


Figura 6.3 - Detección de las esquinas internas

Al terminar el proceso completo de calibración, el toolbox permite visualizar la posición y orientación de las grillas de cada una de las imágenes que se utilizaron, tomando en cuenta el origen que el usuario designó. En la figura 6.4 se puede observar el conjunto de grillas utilizados en la calibración junto con la ubicación espacial de la cámara.

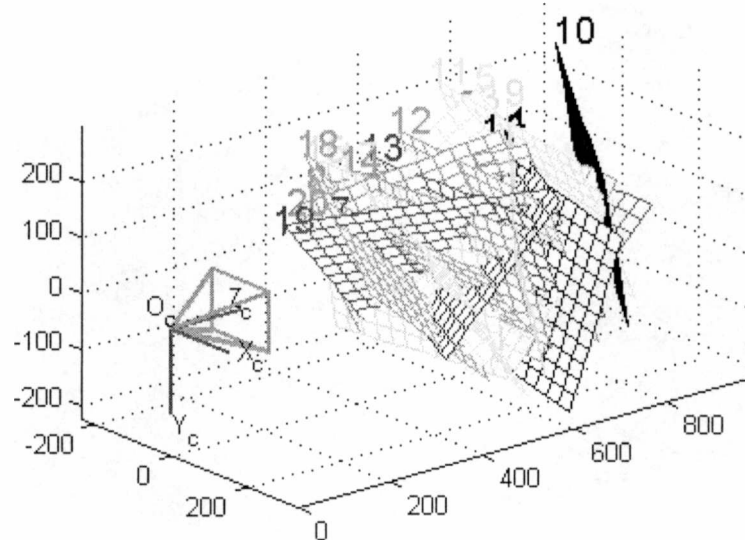


Figura 6.4 - Visualización de la posición y orientación de las grillas de cada imagen

3. Primera aproximación: ‘StereoChecker’

Tomando en cuenta los pasos vistos en el toolbox de Matlab, se desarrolló un prototipo de calibración de cámaras en lenguaje Borland Builder C++. Para el mismo se utilizó la librería de visión por computador OpenCV.

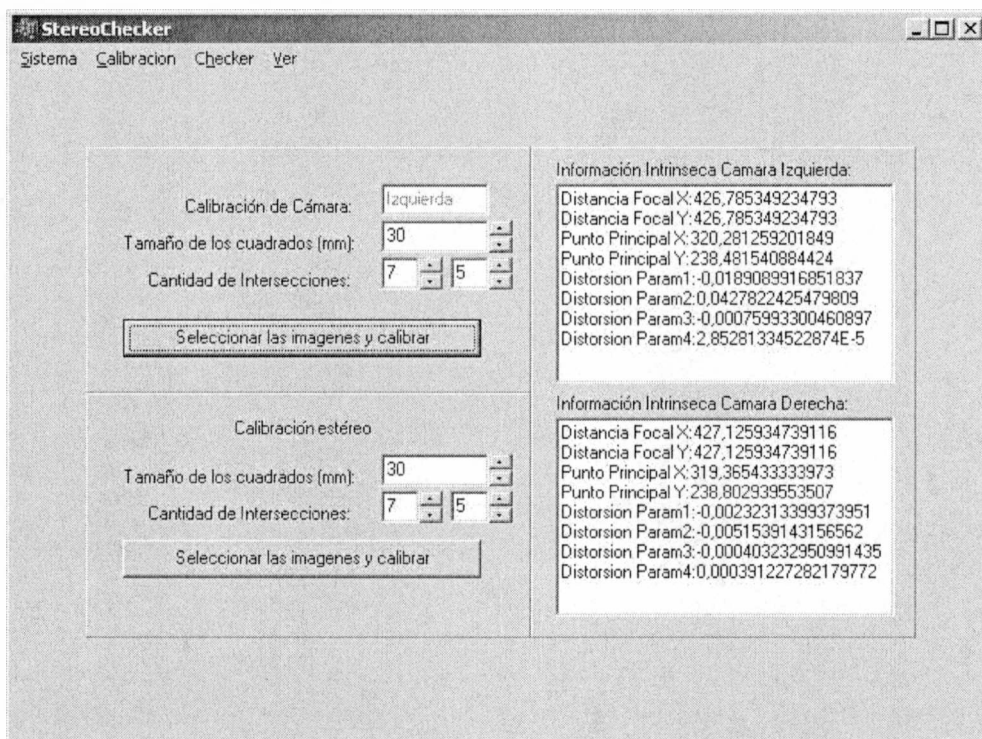


Figura 6.5 - Prototipo de calibración de cámaras StereoChecker

En la figura 6.5 se puede observar la pantalla principal del prototipo, en donde se ha realizado el proceso completo de calibración. En el sector derecho se pueden observar parte de los datos obtenidos. Los pasos a seguir para una correcta calibración de cámaras son los siguientes:

1. Indicar en el menú de Calibración que se realizará la calibración de la cámara izquierda.
2. En el sector de Calibración de Cámaras, configurar el tamaño de los cuadrados de la grilla de calibración en milímetros y la cantidad de intersecciones internas que posee la misma, tanto horizontal como verticalmente.
3. Seleccionar la serie de imágenes de la grilla tomadas con la cámara izquierda. Una vez seleccionadas, el sistema detectará automáticamente las intersecciones de la grilla de cada imagen y se efectuará y visualizará la calibración intrínseca de dicha cámara (figura 6.6).
4. Repetir los pasos 1, 2 y 3 con la cámara derecha. Una vez realizadas ambas

calibraciones intrínsecas el sistema está listo para realizar la calibración extrínseca estéreo.

5. En el sector de Calibración Estéreo, configurar el tamaño de los cuadrados de la grilla de calibración en milímetros y la cantidad de intersecciones internas que posee la misma, tanto horizontal como verticalmente.
6. Seleccionar una imagen de la grilla tomada con la cámara izquierda y una imagen de la grilla tomada con la cámara derecha. Estas imágenes deben ser proyecciones de la misma escena, y las cámaras deben estar en la posición final donde se situarán dentro del sistema estéreo. Una vez seleccionadas, se efectuará la calibración extrínseca estéreo y se visualizará el origen de coordenadas.

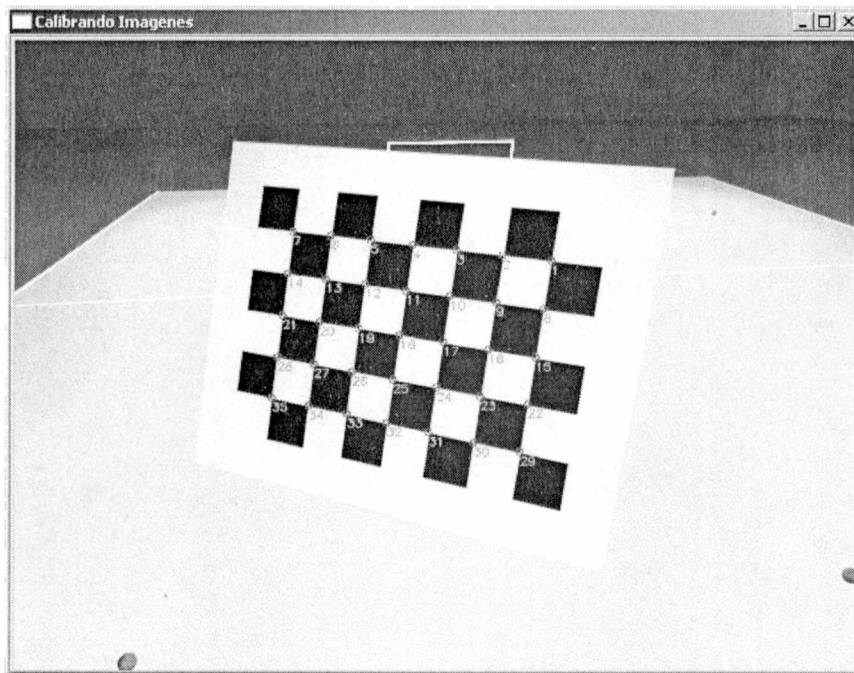


Figura 6.6 - Detección automática de intersecciones en el proceso de calibración intrínseca

Una vez realizados los pasos anteriormente enumerados, es posible realizar un chequeo de la calibración obtenida. En la figura 6.7 se puede observar el módulo de chequeo de calibración. El mismo permite seleccionar un punto en la imagen izquierda y su punto

correspondiente en la imagen derecha, verificando si los mismos se encuentran a una distancia correcta sobre el eje Z.

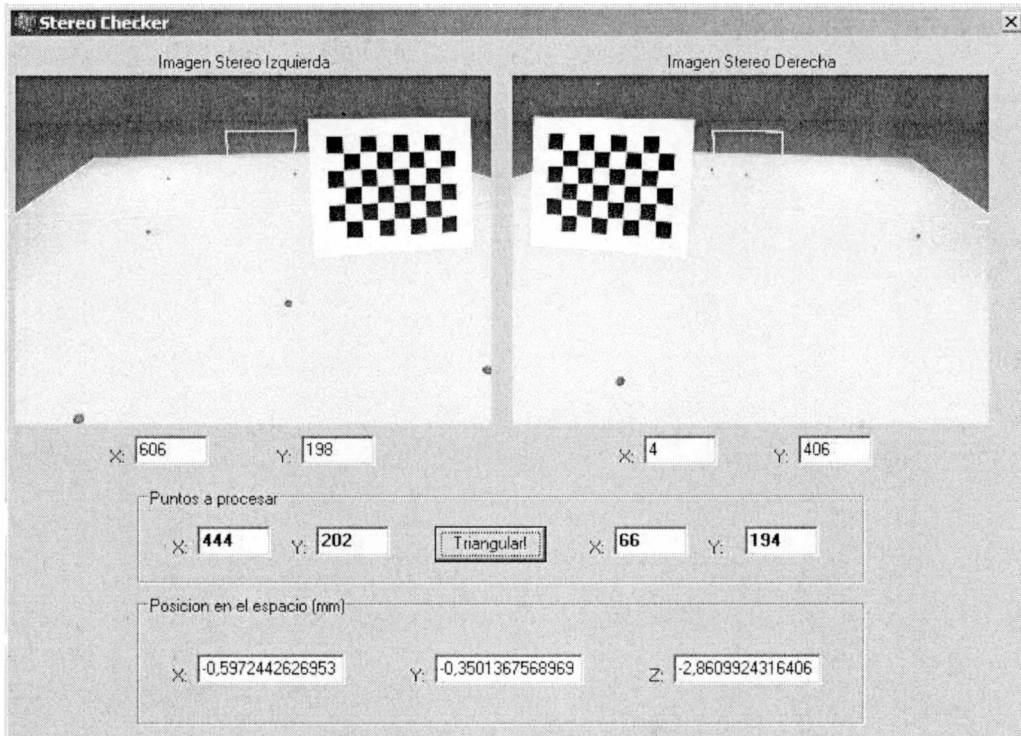


Figura 6.7 - Chequeo de calibración

El prototipo de calibración de cámaras fue exhaustivamente probado antes de ser acoplado a la aplicación principal. A continuación se examinará en detalle el módulo final de calibración de cámaras de la solución implementada.

4. Módulo de calibración de cámaras

En el capítulo 5 se determinó la utilización de dos sistemas estéreo por cuestiones de performance. El módulo de calibración fue desarrollado para realizar dichas calibraciones por separado.

Para calibrar el primer sistema estéreo, se deberá entrar al menú “Configuración → Calibración → Calibrar Sistema de Cámaras 1”. Una vez realizado ésto, se podrá observar el módulo que se muestra en la figura 6.8.

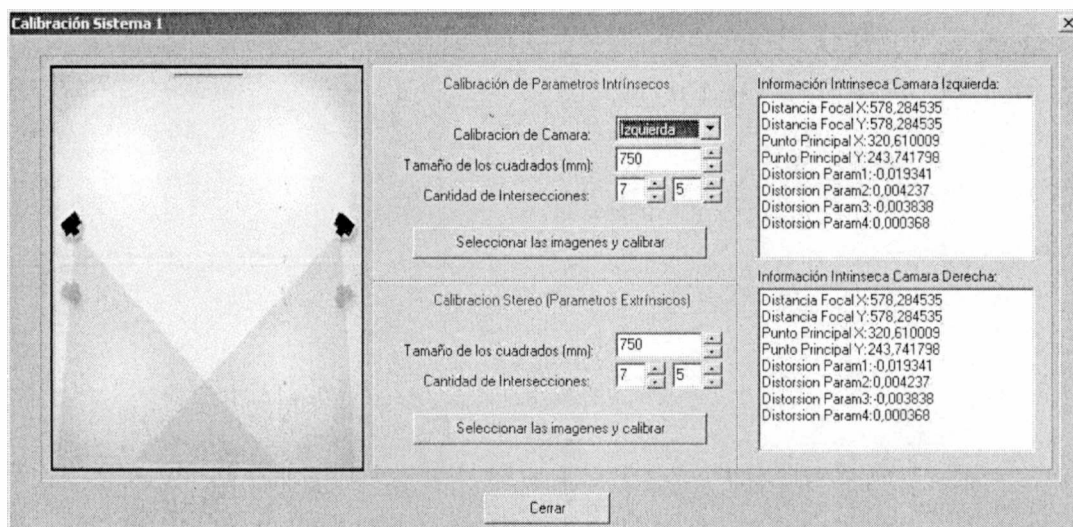


Figura 6.8 - Módulo de calibración del sistema de cámaras 1

En el sector izquierdo se puede observar gráficamente qué sistema estéreo se está calibrando. En este caso se trata del sistema de cámaras que captura la primer parte del tiro, desde que el jugador patea la pelota hasta el *punto de cambio*.

El *punto de cambio* es el punto en el cual se deja de utilizar el sistema de cámaras 1 y se pasa a utilizar el sistema de cámaras 2. El sistema de cámaras 2 captura desde *el punto de cambio* hasta que la pelota detiene su curso o cruza la línea de meta.

La interacción con el usuario es similar a la vista en el prototipo *Stereochecker*. Se escoge el conjunto de imágenes de la grilla de calibración para cada una de las cámaras y se realiza la calibración intrínseca automáticamente. Al finalizar ambas calibraciones intrínsecas se puede continuar con la calibración extrínseca estéreo.

El proceso de calibración extrínseca difiere escasamente con el visto en el

StereoChecker. Para éste, se escoge también una imagen izquierda y derecha correspondientes a la misma escena. La diferencia que presenta es que al finalizar la calibración, el sistema informa cuál es el origen de coordenadas definido sobre la grilla, tal como se puede observar en la figura 6.9. Luego, el usuario debe indicar la distancia desde ese punto a un origen de coordenadas definido en el centro del arco sobre el nivel del suelo. Esto último se puede observar en la figura 6.10.

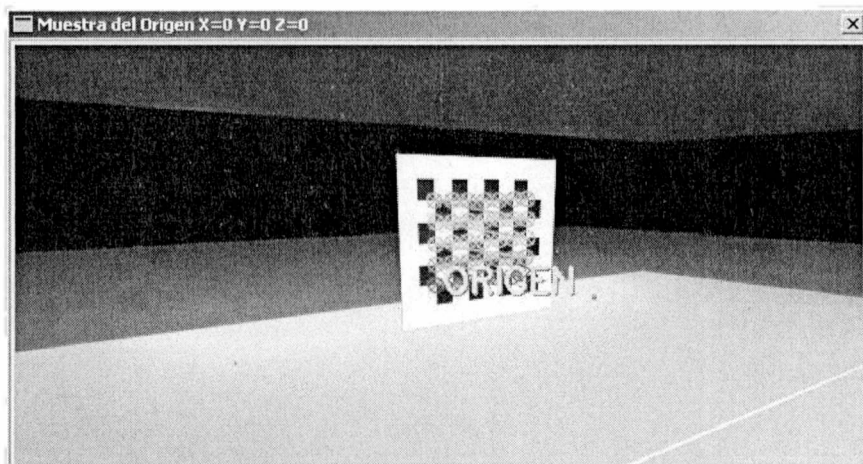


Figura 6.9 - Origen de coordenadas sobre la grilla de calibración

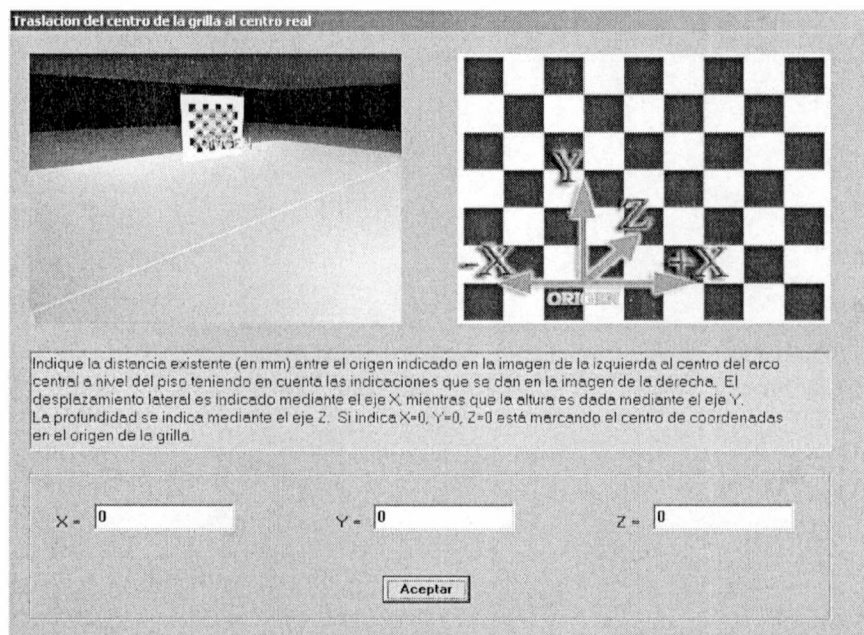


Figura 6.10 - Traslación del origen de coordenadas al centro del arco sobre el nivel del suelo

Una vez realizada la traslación, el origen queda definido en el centro del área de entrenamiento y sobre la línea de meta. En la figura 6.11 se puede observar como quedan establecidos los ejes de coordenadas.



Figura 6.11 - Ejes de coordenadas definidos sobre el área de entrenamiento

Una vez finalizado el proceso de calibración, se puede comprobar que el mismo es correcto mediante la utilización del módulo de “Chequeo de Calibración del Sistema 1”. El mismo funciona de forma similar al visto en el prototipo *Stereochecker* y se puede observar en la figura 6.12.

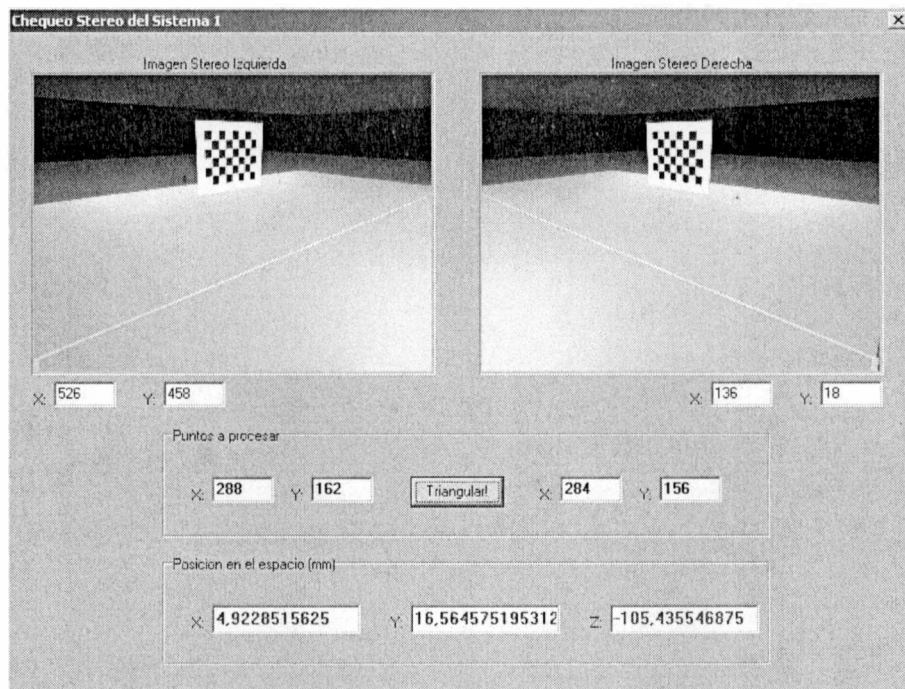


Figura 6.12 - Chequeo de calibraciones

Análogamente, es posible realizar la calibración y comprobación del segundo sistema estéreo.

5. Conclusiones

Se ha logrado el desarrollo de un módulo de calibración de cámaras conforme a los requerimientos del sistema planteado. Para lograr esto, se estudiaron diferentes aplicaciones de calibración de cámaras entre las que se destaca el toolbox de calibración en lenguaje Matlab. Este análisis fue de gran utilidad para el desarrollo de la metodología de calibración implementada.

El módulo desarrollado permite la calibración independiente de cada uno de los sistemas estéreos y la posterior traslación a un punto de origen conocido, el cual permitirá la posterior triangulación de la pelota a lo largo del recorrido que ésta efectúa en cada tiro.

Capítulo VII: Determinación de la trayectoria de la pelota

1. Introducción

La posición de un móvil en el espacio en un instante de tiempo puede ser establecida mediante los conceptos de visión estereoscópica revisados en el capítulo 2.

El presente capítulo detalla el procesamiento de los videos necesario para lograr determinar la posición de la pelota en cada instante de tiempo, y de esa forma, establecer la trayectoria descrita por la misma.

2. Estimación de la trayectoria

Un proyectil o bala de cañón es un objeto en movimiento que posee sólo una fuerza actuando sobre él: la *fuerza de gravedad*. El camino que sigue el proyectil se denomina trayectoria. Los proyectiles viajan con una trayectoria parabólica, dado que la fuerza de la gravedad los acelera hacia abajo.

Dejando a un lado la resistencia del aire, un proyectil mantendría una velocidad horizontal constante, ya que no existen otras fuerzas horizontales actuando sobre él. Al mismo tiempo, la fuerza hacia abajo y la aceleración resulta en un desplazamiento hacia ese sentido de la posición que tendría el proyectil si no hubiera gravedad.

La trayectoria final de una bala de cañón es determinada entonces sólo por tres parámetros: la velocidad inicial, el ángulo de tiro y la fuerza de gravedad. En la figura 7.1 se puede observar un ejemplo de la trayectoria de un proyectil.

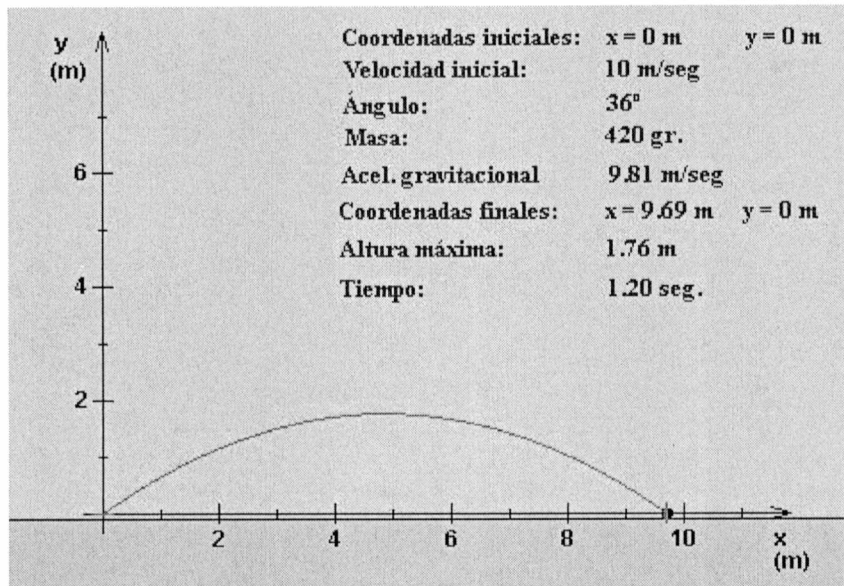


Figura 7.1 - Trayectoria de bala de cañón

Utilizando estos principios, se puede estimar la trayectoria de un tiro libre de fútbol. Sin embargo, si se desea lograr una trayectoria similar a la real, es necesario tener en cuenta una fuerza adicional para poder estimar la curvatura de la parábola del tiro ejecutado. Ésta es la denominada *fuerza magnus*. El efecto magnus es el fenómeno físico por el cual la velocidad de rotación que presenta un objeto afecta a la trayectoria del mismo a través de un fluido, en este caso, el aire.

En la aplicación implementada no se ha tomado en cuenta este efecto adicional, por lo que la trayectoria estimada se corresponde con la de una bala de cañón. En la figura 7.2 se puede observar tanto la trayectoria estimada (en color azul) como la real (en color rojo) de la ejecución de un tiro libre.

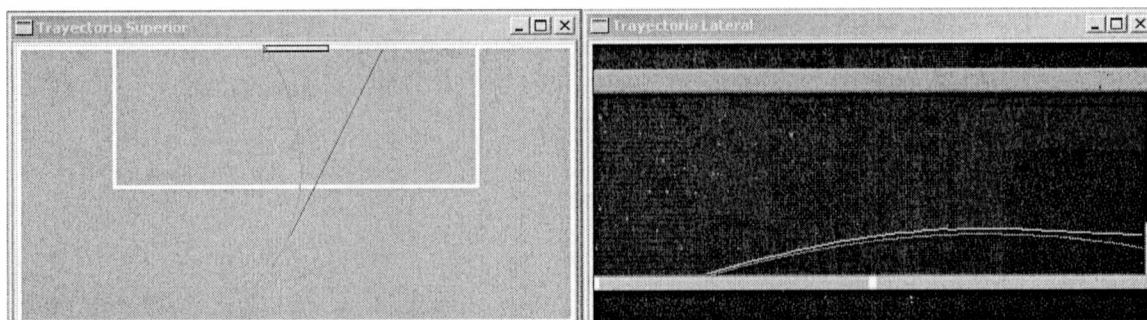


Figura 7.2 - Trayectoria real (rojo) y trayectoria estimada (azul)

Como se puede observar en la figura anterior, sólo la vista lateral es de utilidad, ya que para la vista superior (lo mismo sucede en la vista frontal), la parábola de la trayectoria estimada será muy diferente de la real. Queda planteado para un futuro, realizar una solución que considere la fuerza magnus y de esta forma lograr una estimación precisa en los tres ejes de coordenadas, basándose en la velocidad de rotación calculada.

3. Módulo de filtrado de la pelota

La aplicación implementada tiene como objetivo efectuar el seguimiento de la pelota sobre los videos y detallar la reconstrucción 3D de la trayectoria descrita por ésta. Dado esto, es necesario encontrar una forma de individualizar la pelota a lo largo de los videos capturados. El método adoptado para realizar este trabajo es el de filtrado por color.

En un principio se dispuso que las pelotas utilizadas para el entrenamiento de tiros libres debían ser en su totalidad de color rojo. Posteriormente se aumentó esta definición: las pelotas deben contener además pequeñas etiquetas de color amarillo, las cuales son utilizadas para el cálculo de la rotación de la pelota, que se analizará en detalle en el próximo capítulo.

Para realizar un correcto filtrado por color, es necesario configurar los valores límites por los que un píxel será considerado o no como perteneciente a la pelota. Esta tarea es la que se realiza en el módulo de “Configuración del umbral rojo”.

El módulo de configuración del umbral rojo requiere como entrada una imagen, que será sobre la que se configurarán los valores de filtrado del umbral rojo. La imagen provista deberá pertenecer a una de las cámaras de cualquiera de los dos sistemas estéreo y deberá ser tomada bajo las mismas condiciones en las que se efectuarán luego los tiros libres. Esto es debido a que un cambio en la iluminación de la escena puede alterar notablemente los valores del umbral configurado.

Es indispensable además que en la imagen provista aparezca al menos una pelota para poder entonces realizar la configuración de los umbrales sobre los píxeles rojos pertenecientes a la misma. Cabe destacar que, bajo esta metodología de identificación de la pelota, el único objeto de color rojo en la escena deberá ser la pelota. Tanto el sector de entrenamiento como las vestimentas de los jugadores no podrán contar con partes de este color.

Una vez seleccionada la imagen, y tomando en cuenta el modelo de color RGB, se procede a configurar un umbral de colores máximos y mínimos permitidos, tal como se detalla a continuación:

- Valor mínimo canal rojo: los píxeles que cuenten en su conformación con una intensidad de rojo por debajo de dicho valor no serán tomados en cuenta.
- Valor máximo canal verde: los píxeles que cuenten en su conformación con una intensidad de verde por encima de dicho valor no serán tomados en cuenta.
- Valor máximo canal azul: los píxeles que cuenten en su conformación con una intensidad de azul por encima de dicho valor no serán tomados en cuenta.

En la figura 7.3 se puede observar una configuración de estos valores y el filtrado resultante.

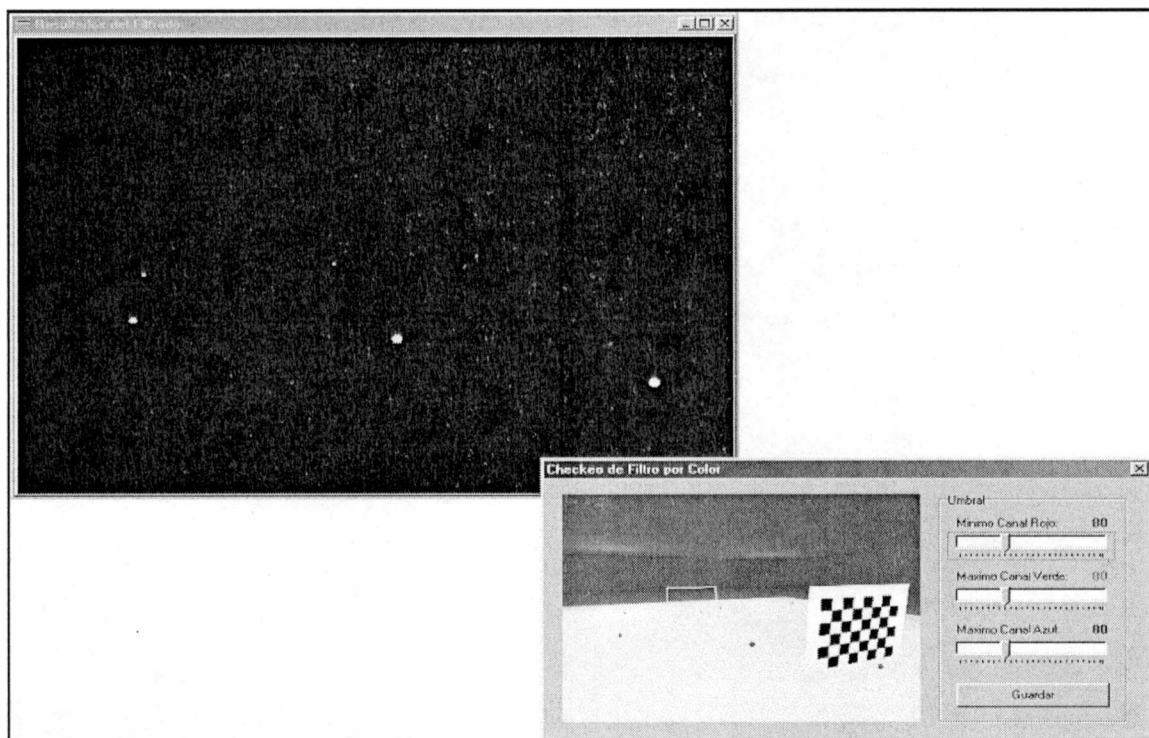


Figura 7.3 - Configuración del umbral rojo

4. Procesamiento de un tiro

A continuación se describirán los distintos pasos que constituyen el procesamiento de un tiro para la obtención del vector de posiciones en el espacio que adopta la pelota a lo largo de éste.

Se asume que antes de comenzar el mismo, ya se han realizado los procesos de calibración de ambos sistemas de cámaras estéreo y la configuración del filtrado por color, tanto para el color rojo de la pelota como para el color amarillo de las etiquetas.

Cabe mencionar que tanto la calibración como la configuración de umbrales para el filtrado son procedimientos que deben realizarse una única vez al iniciar el uso del sistema. Sólo es necesario realizar nuevamente alguno de ellos si las condiciones del

ambiente varían. Esto puede darse en caso de un cambio en la iluminación de la escena, modificación de la posición las cámaras, cambio de lentes de las mismas, etc.

En la figura 7.4 se puede observar el diagrama de procesamiento completo de un tiro, que consta de cuatro etapas principales. La primer etapa se corresponde con la captura de la secuencias de videos.

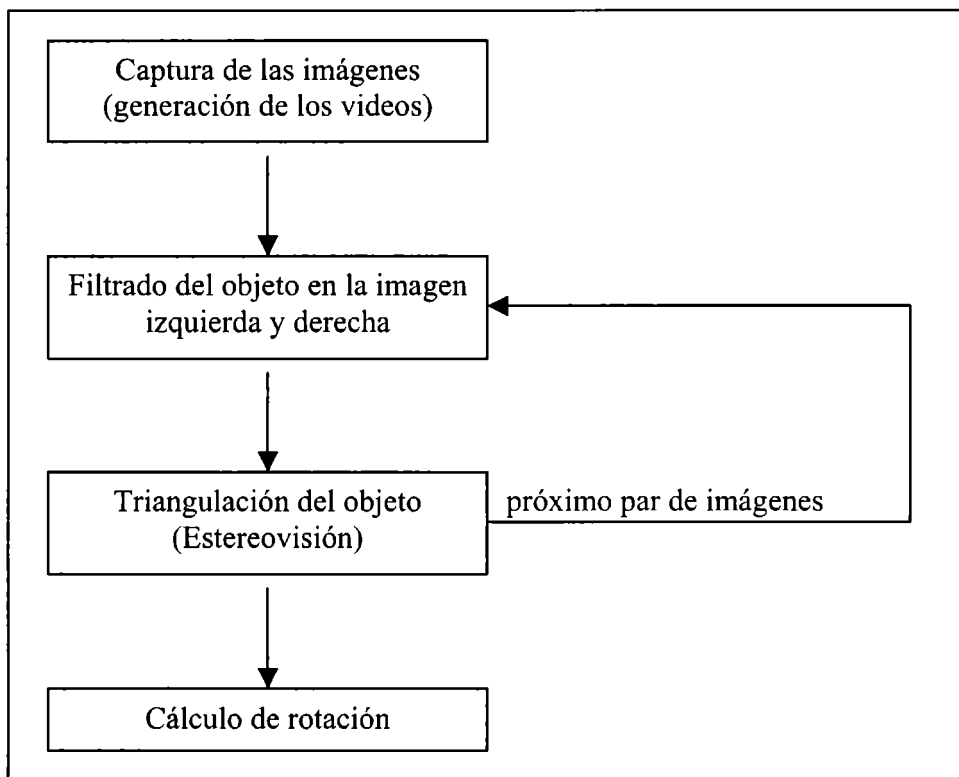


Figura 7.4 - Diagrama de procesamiento de un tiro

Una vez finalizada esta etapa, se cuenta con cinco videos en modo color RGB:

- Dos videos correspondientes a las cámaras izquierda y derecha del sistema estéreo 1
- Dos videos correspondientes a las cámaras izquierda y derecha del sistema estéreo 2
- Un video correspondiente a la cámara utilizada para el cálculo de rotación

Las siguientes dos etapas se encargan de la determinación de la posición de la pelota. Primeramente se toma una imagen de la cámara izquierda y su correspondiente de la cámara derecha pertenecientes al sistema estéreo 1, el cual se encarga de la captura de la primera parte del tiro. Con estas imágenes se realizará el siguiente procesamiento:

1. Se realiza el filtrado de la pelota en la imagen izquierda.
2. Se busca el centro $c_{izq} = (x_{izq}, y_{izq})$ de la zona filtrada (la pelota).
3. Se realiza el filtrado de la pelota en la imagen derecha.
4. Se busca el centro $c_{der} = (x_{der}, y_{der})$ de la zona filtrada (la pelota).
5. Con c_{izq} y c_{der} se realiza la triangulación obteniendo el punto tridimensional $P = (x, y, z)$.

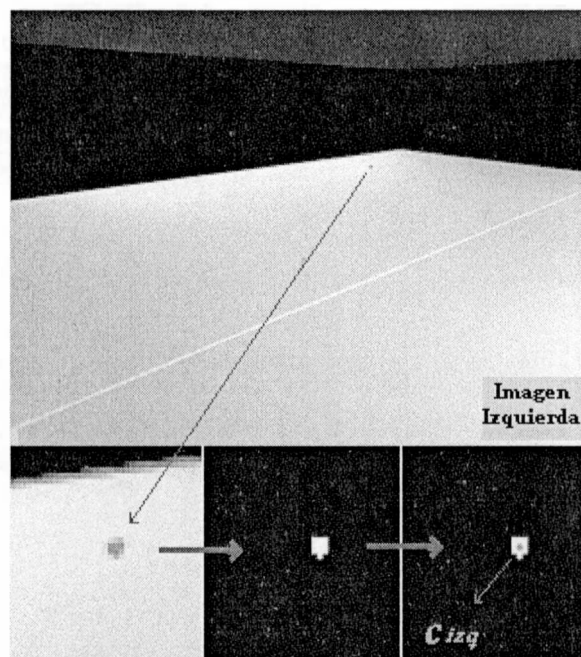


Figura 7.5 - Procesamiento de la imagen izquierda

En la figura 7.5 se pueden observar los pasos 1 y 2 del procesamiento. Análogamente se realizan los pasos 3 y 4 con la imagen derecha. Una vez que se cuenta con ambos puntos se realiza la triangulación utilizando los conceptos de visión estereoscópica vistos en el capítulo 2. De esta forma se obtiene el punto 3-D correspondiente a la

posición de la pelota en el primer cuadro de la captura de la escena. Este mismo procesamiento se realizará sobre todos los cuadros de los videos de los sistemas estéreo, logrando finalmente una serie de puntos 3-D que describen la trayectoria de la pelota.

Una vez finalizado el proceso de determinación de la trayectoria de la pelota se continúa con la estimación de la velocidad de rotación de la misma. En el siguiente capítulo se describirá en detalle el método desarrollado para realizar esta tarea.

Una vez computada la rotación de la pelota, es posible visualizar la reconstrucción de la trayectoria lograda. En la figura 7.6 se puede observar el módulo de visualización de tiros. En el mismo se presentan tanto los videos del tiro original como la trayectoria reconstruida a partir de los mismos. En este caso se detalla con amarillo el punto de mayor altura logrado por la pelota. La trayectoria y el punto de mayor altura también son detallados sobre los videos originales.

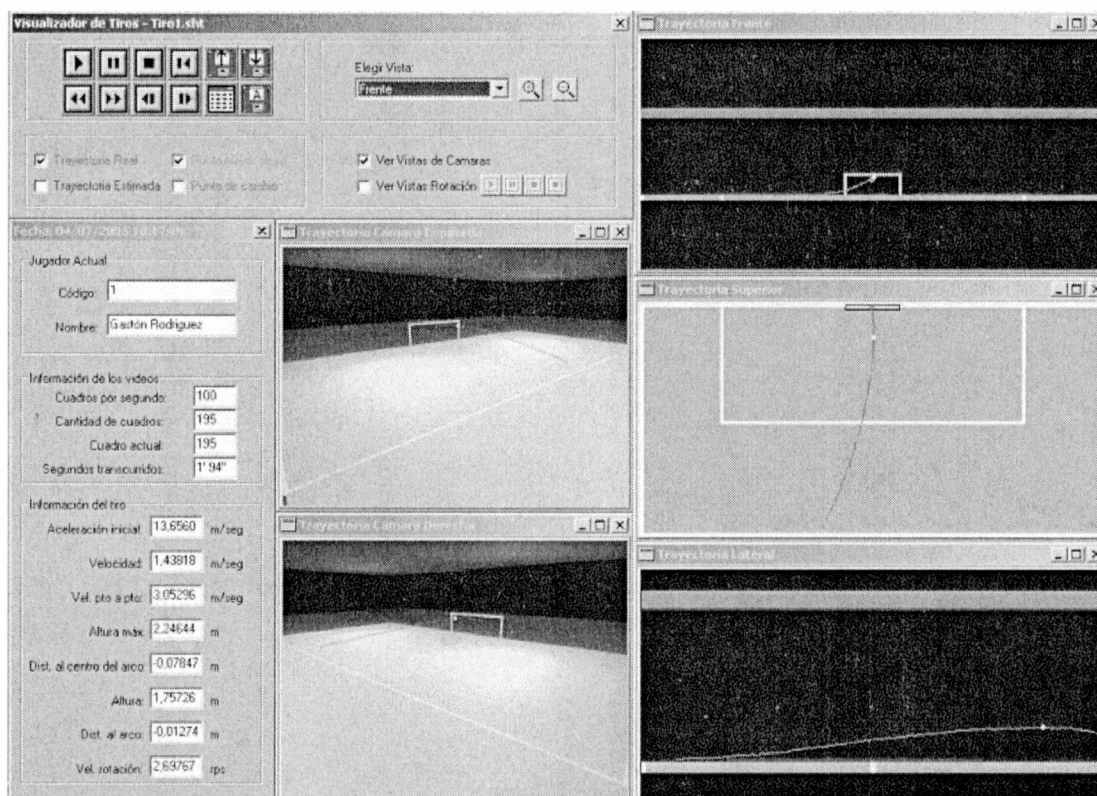


Figura 7.6 - Módulo de visualización de tiros

Además de la información visual, el módulo de visualización de tiros cuenta con diversa información numérica. El capítulo 9 analiza en detalle el entorno de visualización de tiros desarrollado.

5. Aspectos Técnicos

Al momento de realizar el procesamiento de los videos capturados, se deben efectuar una serie de tareas en cada uno de los cuadros pertenecientes a los mismos. Con el fin de llevar a cabo este procesamiento ordenadamente, realizar los cálculos necesarios y almacenar la trayectoria y demás datos computados, se hizo uso de una estructura de datos desarrollada especialmente con este objetivo. Esta estructura fue denominada ShotData y se analizará en detalle en la sección 7.1 del capítulo 9.

En la figura 7.7 se pueden observar cada una de las tareas que se efectúan sobre los videos de los tiros libres. Por simplicidad y mayor entendimiento, solo se reproducen los fragmentos relevantes y en forma de pseudocódigo. Es interesante detallar algunas de estas líneas con el fin de facilitar la comprensión de las mismas.

(1) Los videos tendrán una serie de cuadros iniciales en donde el jugador esta preparando la ejecución del tiro libre. Estos cuadros deben ser despreciados, y pueden ser identificados, ya que la pelota aún no se encuentra en movimiento.

(2) Un tiro finaliza cuando la pelota cruza la línea de meta o la misma invierte su dirección, es decir, comienza a retroceder.

(3) El origen de los cuadros a procesar dependerá del sistema de cámaras estéreo que se esté utilizando para calcular la trayectoria en cada punto. El sistema de cámaras estéreo a utilizar dependerá del punto de cambio de sistema definido.

```

procesamientoVideos()
{
    Mientras no se mueva la pelota ignoro los cuadros; (1)
    Mientras el tiro no finalice (2)
    {
        Tomar los cuadros de los videos que corresponde; (3)
        Escribir los cuadros en los videos comprimidos; (4)
        Buscar el centro de la pelota en ambos cuadros; (5)
        Almacenar los centros encontrados; (6)
        Triangular el punto utilizando los centros encontrados; (7)
        Almacenar el punto tridimensional calculado; (8)
        Calcular la velocidad punto a punto; (9)
        Determinar la aceleración inicial del tiro; (10)
        Determinar la mayor altura; (11)
        Determinar la velocidad promedio del tiro; (12)
        Determinar la trayectoria estimada; (13)
    }
    Generar los videos de rotación a partir del original; (14)
    Estimar la velocidad de rotación; (15)
    Almacenar los videos y los datos calculados; (16)
}

```

Figura 7.7 – Seudo-código de procesamiento de videos

(4) Estos videos se visualizarán en el entorno de visualización de tiros. Los mismos corresponden a las vistas preparadas para observar los videos originales.

(5) y (6) Además de servir para realizar el cálculo de la posición de la pelota en el espacio, estos puntos se utilizarán luego para realizar el trazado de la trayectoria sobre las vistas de los videos originales del tiro. Debido a esto, deben ser almacenados en la estructura ShotData.

(7) y (8) Utilizando los centros de la pelota encontrados en los cuadros del video izquierdo y el video derecho correspondientes, se procede a triangular el punto. Una

vez calculado, se almacena el resultado en la estructura ShotData. Con este punto se podrá luego trazar la trayectoria de la pelota en las vistas del tiro sin perspectiva definidas en el entorno de visualización de tiros.

(9) Tomando en cuenta la posición de la pelota en el punto anterior, la posición actual, y las características del video, se procede a calcular la velocidad punto a punto de la misma.

(10), (11) y (12) En base a los puntos triangulados, se podrá ir calculando todos los datos de relevancia, tales como la aceleración inicial, la altura máxima y la velocidad promedio. Para realizar el cálculo de la aceleración inicial, sólo se toman en cuenta los cuadros iniciales de los videos.

(13) La trayectoria estimada es determinada en base a la velocidad inicial, el ángulo de tiro y la fuerza de gravedad.

(14) A partir del video correspondiente a la cámara dispuesta para estimar la rotación de la pelota, se genera tanto el video comprimido de rotación, como el video comprimido del seguimiento de la misma.

(15) Se realiza la estimación de la velocidad de rotación en base al método ad-hoc desarrollado. Este método se explicará en detalle en el capítulo 8.

(16) Se almacenan los cuatro videos generados (cámara izquierda y derecha de la trayectoria del tiro, video de rotación y de seguimiento de rotación) y todos los datos calculados mencionados anteriormente.

Dado que este procesamiento puede tener una duración del orden de un minuto, en todo momento se informa el estado y el progreso del mismo. También es posible, si se desea, cancelar el procesamiento.

6. Conclusiones

Se ha desarrollado un algoritmo para el cálculo de la trayectoria de un tiro libre a partir de dos sistemas de cámaras estéreo. Para lograr esto se utilizaron diferentes técnicas de procesamiento de imágenes y video. También se hizo uso de funciones pertenecientes a la librería OpenCV.

Como se explicará en el capítulo 10, las pruebas realizadas arrojaron un balance positivo en cuanto a la precisión lograda por el algoritmo desarrollado, con un tiempo de respuesta mayormente aceptable.

Capítulo VIII: Determinación de la velocidad de rotación de la pelota

1. Introducción

Conocer la velocidad de rotación de la pelota lograda en un tiro libre es de vital importancia, ya que el movimiento de rotación de ésta provoca que se desvíe la trayectoria de la misma hacia un lado u otro mientras se encuentra en el aire debido a un fenómeno conocido como fuerza de Magnus. Esta fuerza es muy pronunciada al final del vuelo de la pelota, cuando el jugador ha aplicado mucho efecto de rotación en el tiro.

Se han propuesto varios métodos para el cálculo de velocidad de rotación. En general, aunque los mismos presentan buenos resultados, se basan en complejos etiquetados de la pelota, en hardware altamente especializado o en algoritmos de ejecución costosa. El método ad-hoc desarrollado busca simplificar los requerimientos de hardware y las técnicas de etiquetado de la pelota.

El presente capítulo detalla el análisis preliminar y la metodología utilizada para realizar el cálculo de la velocidad de rotación que adopta la pelota en la trayectoria hacia el arco.

2. Procesamiento de un tiro

Siguiendo con la secuencia de pasos necesaria para obtener la información detallada en los requerimientos, se llega a la instancia de realizar el cálculo de rotación.

La figura 8.1 muestra el estado del proceso una vez finalizado el procesamiento cuadro a cuadro de la secuencia de video para la triangulación de la pelota a lo largo del tiro. El paso restante es realizar el cálculo de rotación, el cual por su complejidad e implementación ad-hoc será detallado a continuación; detallando el análisis y desarrollo llevados a cabo.

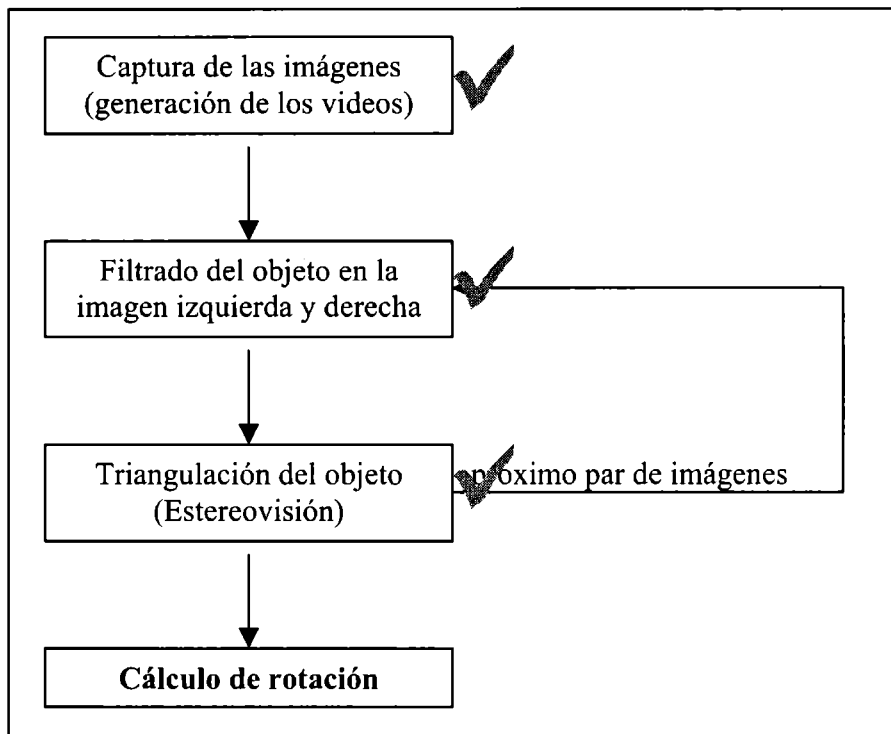


Figura 8.1 – Último paso de la secuencia: el cálculo de rotación

3. Análisis preliminar

Al ejecutar un tiro libre, las direcciones de giro rotacional que puede adoptar la pelota son sin ninguna duda casi ilimitadas. Esto significa que la misma es capaz de rotar en un sin fin de direcciones posibles. Cabe entonces preguntarse, se pueden clasificar las variantes que presenta la rotación de la pelota?.

Inicialmente, tomemos como referencia tres ejes principales sobre los que la pelota puede realizar el movimiento de rotación. Como se muestra en la figura 8.2, tendríamos tres posibilidades: que la pelota gire sobre el eje X, el eje Y o el eje Z. Naturalmente, el movimiento absoluto sobre un eje anula la posibilidad de movimiento de los restantes.

Pero claro está, el número de ejes que atraviesa el centro de la pelota es infinito. De acuerdo a la posición en que éstos se tomen, podría decirse que la misma adopta un movimiento absoluto sobre un eje o una combinación de fuerzas sobre más de uno, con lo cual no es posible lograr una clasificación útil.

Esta problemática se debe a que el análisis realizado se ha basado en tomar como punto de interés el centro de masa de la pelota. Es necesario entonces corregir la pregunta anteriormente planteada: se pueden generalizar o clasificar las variantes que presenta la rotación de la pelota de acuerdo a un punto de vista en particular?.

Para los fines de la aplicación, el punto de vista de interés será la cámara que capture la pelota en movimiento. Bajo este contexto, existen tres casos principales, tal como se muestra en la figura 8.3. El primero de estos es el que contempla las dos posibilidades de rotación aparente lateral. La segunda alternativa son los casos en que la pelota realiza un movimiento aparente de rotación frontal, en cualquiera de sus direcciones (en la figura sólo se muestran cuatro). La tercer posibilidad contempla la opción en la cual la pelota no presenta movimiento alguno de rotación.

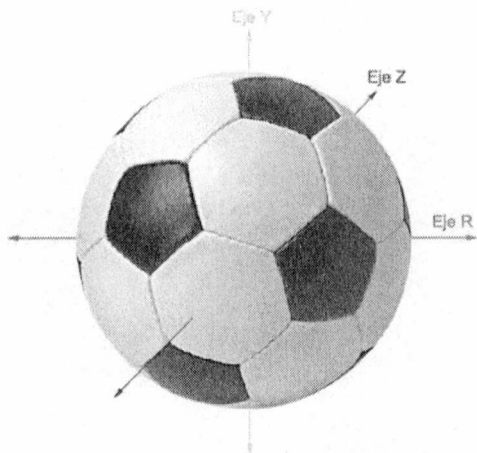


Figura 8.2 – Ejes de rotación basados en el centro de masa de la pelota

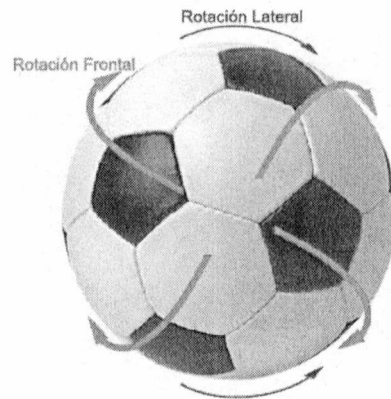


Figura 8.3 – Casos de rotación según movimiento aparente respecto del punto de vista de interés

Basándose en esta generalización de casos, se puede desarrollar una solución que explote esta característica con el fin de determinar la velocidad en que la pelota realiza el movimiento de rotación, ya sea rotación lateral aparente, frontal aparente o una combinación de ambas fuerzas.

4. Marcación de la pelota

El principal problema a resolver es el de determinar una marcación óptima sobre la superficie de la pelota con el fin de poder conocer en todo momento el movimiento que ésta realiza.

El análisis realizado presenta las bases bajo las cuales se proyectó la forma de marcado de la pelota. Se buscó maximizar la sencillez de diseño del mismo.

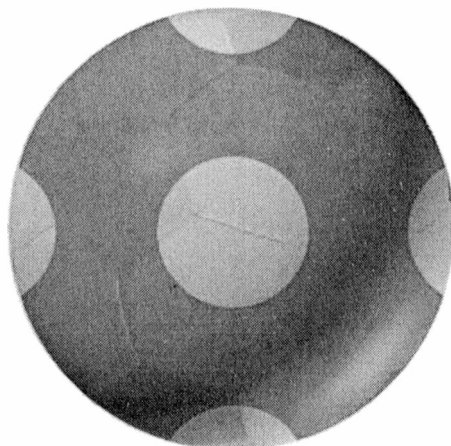


Figura 8.4 – Etiquetado de la pelota

La configuración de marcas puede observarse en la figura 8.4. La misma presenta seis marcaciones sobre la superficie de la pelota, dos por eje axial. Los colores que se muestran, tanto de la pelota como de las marcas fueron los utilizados para las pruebas, los mismos podrían no necesariamente poseer dicha configuración de color.

Como se detalla posteriormente, el algoritmo desarrollado realiza el seguimiento de las marcas a lo largo de la secuencia de video con el fin de determinar la velocidad y dirección de rotación de la pelota. La detección de las marcas en cada uno de los cuadros que componen la captura se realiza mediante una técnica de filtrado.

5. Filtrado

Existe el problema de identificar individualmente cada una de las marcas de la pelota. Esto es, dada una imagen y una marca, determinar la nueva ubicación de dicha marca en la imagen anterior o siguiente. El filtrado de las marcas amarillas es el primer paso y se realizó mediante una umbralización por niveles de intensidad lumínica en los canales RGB.

La lógica de filtrado es análoga a la realizada sobre el canal rojo para identificar la pelota en el procesamiento de determinación de la trayectoria. En este caso, dado que las etiquetas son amarillas, el criterio de umbralización es el que se muestra en el módulo de filtrado de marcas de la figura 8.5.

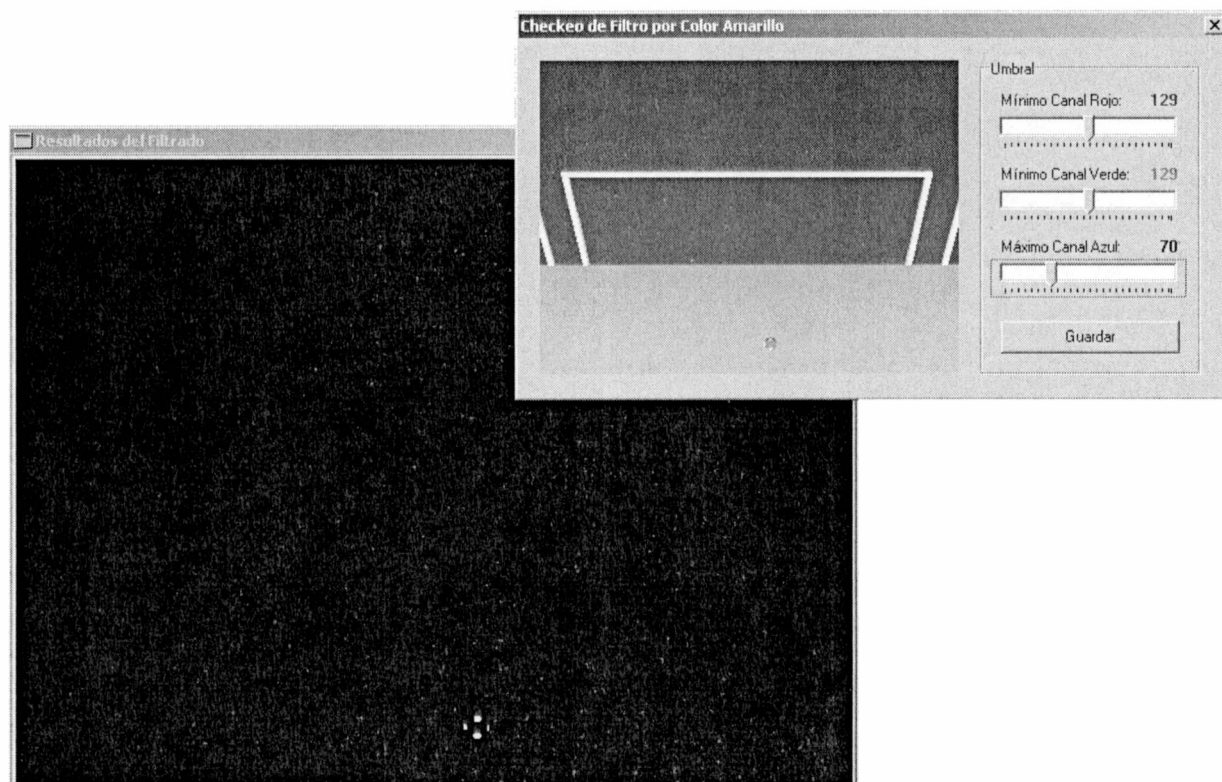


Figura 8.5 – Modulo de filtrado de marcas

Una vez especificados los niveles de intensidad adecuados para cada canal, se procesa la imagen realizando el filtrado con dichos umbrales.

Con el fin de lograr mayor robustez, el filtrado se realiza de forma que se ignoren los píxeles aislados que no pertenezcan a las marcas que componen la pelota (lo cual es necesario contemplar en condiciones adversas, como por ejemplo, baja luminosidad).

Para llevar a cabo esto se desechan píxeles lejanos a la posición de la pelota y píxeles que no estén ligados a la misma, es decir, se tienen en cuenta los 4-vecinos y 8-vecinos

para determinar el filtrado. De esta forma el método determina sin errores cada una de las marcas de la pelota. La figura 8.6 muestra la pelota una vez filtrada.

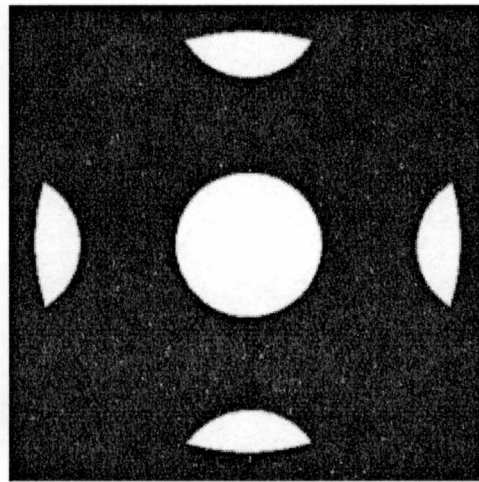


Figura 8.6 – Filtrado de la pelota

El objetivo del seguimiento de las marcas es el de determinar la posición de cada marca en las imágenes anterior y siguiente a una que será tomada como referencial.

6. Seguimiento de las marcas

Como se ha explicado anteriormente, la pelota puede adoptar una rotación lateral aparente o frontal aparente. Si incorporamos además el modelo de marcación expuesto; se puede llegar a la conclusión de que la rotación de la pelota de la figura 8.3 puede manifestarse de las siguientes formas:

1. La pelota realiza una rotación lateral aparente, con lo cual (a) puede llegar a presentarse una marca central aparentemente inmóvil, (b) la marca central atraviesa parcialmente el centro de la pelota, o (c) puede no presentarse una marca en el centro de la misma.

2. La pelota realiza una rotación frontal aparente, con lo cual es posible que (a) una marca central atraviese el centro de la pelota, o que (b) ninguna marca central atraviese el centro de la pelota.

3. La pelota no realiza rotación aparente alguna, esto es: el tiro ha sido efectuado de manera tal que la pelota no ha sufrido ninguna fuerza de rotación.

Por lo tanto, para obtener un valor preciso será necesario realizar el seguimiento de la marca central y laterales, y, según sea el caso basarse en una o ambas para realizar el cálculo de rotación. La figura 8.7 presenta los esquemas que resumen las alternativas expuestas.

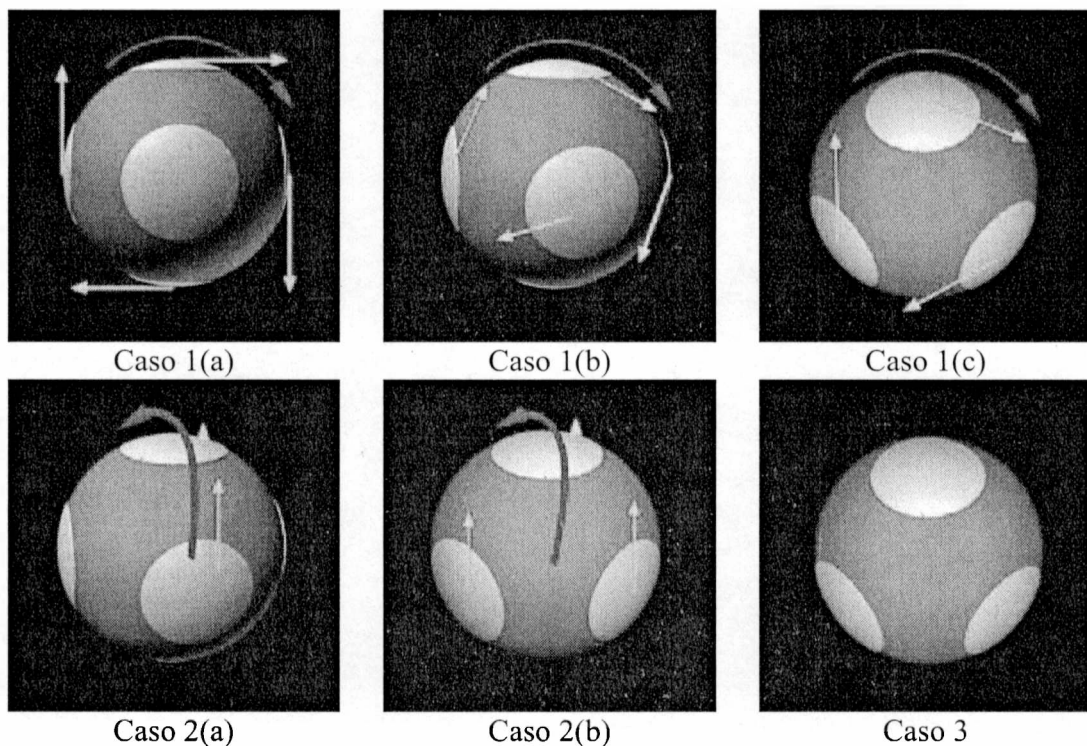


Figura 8.7 – Formas de manifestación de rotación de la pelota

A continuación se explicará en detalle el método desarrollado, el cual se basa en el análisis descripto hasta el momento.

7. Modelo computacional

El seguimiento de las marcas para el posterior cálculo de rotación debe llevarse a cabo de manera tal que las marcas de interés no se pierdan a lo largo de los cuadros seleccionados como óptimos para la medición.

7.1. Marca central definida

La forma de determinar cual es el cuadro óptimo se basa en buscar la marca amarilla central más definida dentro de la secuencia de video completa. Para llevar a cabo esta tarea, para cada cuadro, se determina el centro de la pelota y se obtiene el nivel de intensidad amarillo dentro de una ventana de búsqueda V de tamaño acorde al diámetro de la pelota en píxeles. La figura 8.8 muestra este proceso.

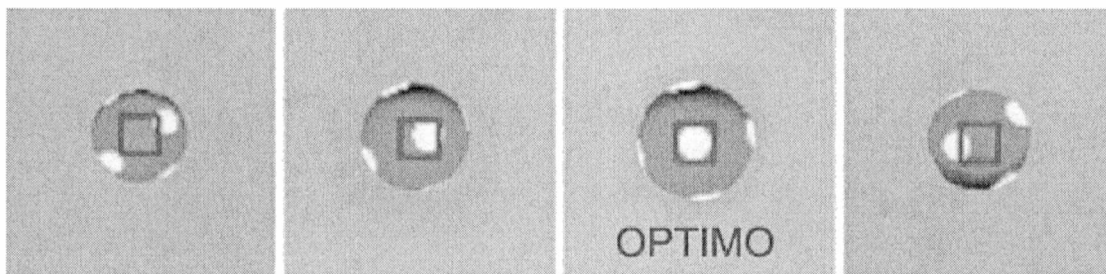


Figura 8.8 – Selección del cuadro óptimo.

Sean $i_1..i_n$ el conjunto de cuadros que componen la secuencia de video de captura de un tiro, V el nivel de luminosidad amarilla de la ventana de búsqueda dentro de la zona central de la pelota, U el valor de luminosidad amarilla que debe superar la búsqueda sobre la ventana V .

Entonces, se debe encontrar el cuadro i_c tal que,

$$i_c = \max \{ V(i_k) > U, k = 1..n \}$$

Como se verá posteriormente, el algoritmo desarrollado contempla los casos en el que se presenten las alternativas 1(c) o 2(b) de la figura 8.5, en las que puede ocurrir que i_c no llegue a existir.

Una vez encontrado i_c , se utilizará la información de los cuadros i_{c-1} , i_c e i_{c+1} para el cálculo de rotación. Un algoritmo de etiquetado será aplicado a cada una de las marcas de los cuadros seleccionados. El etiquetado de marcas consiste en asignarle un identificador (etiqueta) a cada una de las marcas obtenidas luego del filtrado. Esta tarea se realiza mediante un procesamiento de rasterización de la imagen y el resultado se puede observar en la figura 8.9.

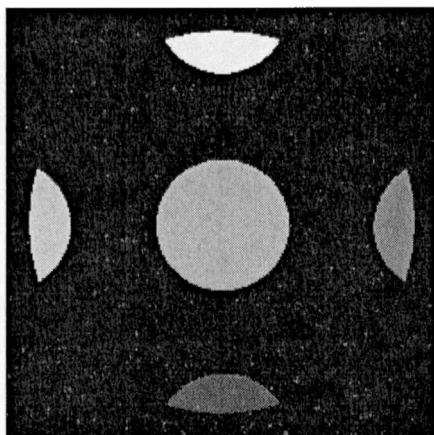


Figura 8.9 – Etiquetado de la pelota

En dicha imagen, el algoritmo de etiquetado detectó 5 marcas diferentes: una superior, una izquierda, una central, una derecha y una inferior. A cada una de éstas se le es asignado un valor diferente a fin de identificarlas inequívocamente.

Dado que la pelota se encuentra en movimiento (esto es inclusive notorio en la corta secuencia de tres cuadros), las etiquetas asignadas a las marcas en cada cuadro pueden no llegar a ser las mismas. Esto se debe a la característica que presentan los algoritmos de etiquetado, en los cuales el procesamiento de la imagen se realiza en forma de raster o barrido a lo largo de la imagen.

Por ejemplo, si se realiza un raster horizontal desde la esquina superior izquierda hasta la inferior derecha, y si la pelota de la figura 8.9 girase un par de grados en sentido anti horario, la marca de la izquierda podría ser encontrada luego de la marca de la derecha. En ese caso, el etiquetado sería incorrecto para la secuencia seleccionada y no existiría una consistencia de las etiquetas a lo largo de los cuadros. A fin de contrarrestar este problema, la forma de determinar la posición de cada etiqueta a lo largo de la secuencia de video se detalla a continuación.

Sean i_{c-1} , i_c e i_{c+1} los cuadros seleccionados, se tendrán un conjunto de marcas por cuadro. Por lo tanto, para el cuadro central, existirán i_c^t , $t = 1..4$ etiquetas t , destacándose además la etiqueta central c , por la cual se obtuvo i_c . Esto es, sean

i_{c-1}^t , $t = 1..5$ las cinco (como máximo) marcas del cuadro anterior a i_c

i_c^t , $t = 1..4$ las cuatro (como máximo) marcas del cuadro i_c , más la marca central i_c^c

i_{c+1}^t , $t = 1..5$ las cinco (como máximo) marcas del cuadro posterior a i_c

Se requiere entonces determinar la posición de la etiqueta central i_c^c dentro de las etiquetas de los cuadros anterior y posterior al central. Esto significa que se deberán encontrar las etiquetas distinguidas i_{c-1}^c y i_{c+1}^c dentro de i_{c-1}^t , $t = 1..5$ y i_{c+1}^t , $t = 1..5$.

Para esto se realiza una búsqueda en forma de espiral de una etiqueta desde el punto central de la etiqueta i_c^c en los cuadros i_{c-1} y i_{c+1} . Dadas las características de determinación de i_c , la primera etiqueta encontrada será i_{c-1}^c y i_{c+1}^c respectivamente. La figura 8.10 ilustra esta situación.

Conociendo la distancia que las marcas han realizado entre un cuadro y el siguiente de la secuencia de video, la velocidad de captura de la misma (FPS) y el tamaño en milímetros que representa un píxel, se puede determinar así la velocidad de rotación de la pelota. El último parámetro se obtiene conociendo el diámetro de la pelota en

milímetros y determinando cuántos la píxeles componen la misma en la zona ecuatorial, como se explica en apartados posteriores.

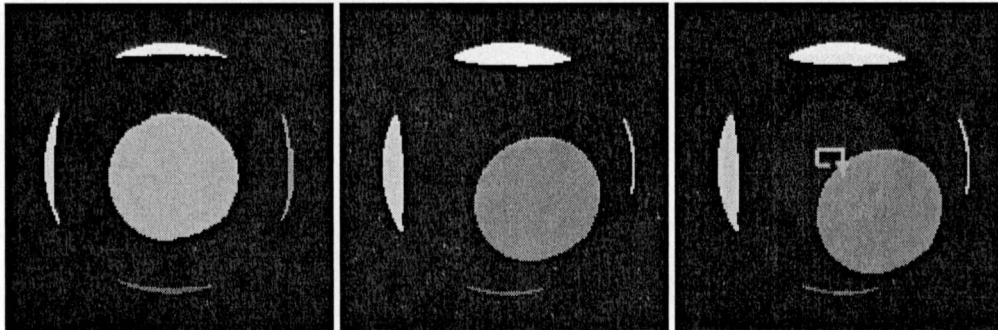


Figura 8.10 – (a) Cuadro i_c (b) Cuadro i_{c+1}
(c) Búsqueda de etiqueta i_{c+1}^c en i_{c+1} respecto del centro de i_c^c

Podríamos concluir que el algoritmo presentado hasta el momento es totalmente válido cuando la pelota adopta rotación frontal aparente completa, tal es el caso 2(a) de la figura 8.7. Pero si la pelota presenta una combinación de rotación frontal aparente y lateral aparente (caso 1(b)), los resultados obtenidos no serán completamente correctos, ya que se está despreciando una fuerza de rotación. Mas aún, en el caso de que se presente una rotación lateral aparente completa (caso 1(a)), el cálculo de rotación sería totalmente incorrecto, ya que la posición de i_c^c será igual a la de i_{c-1}^c y i_{c+1}^c .

Es por esto que las etiquetas i_{c-1}^l ($t = 1..5$), i_c^l ($t = 1..4$) y i_{c+1}^l ($t = 1..5$) deben ser contempladas en el modelo computacional desarrollado. Realizando un procedimiento análogo al explicado, y basándose en las marcas perimetrales se obtiene entonces la velocidad de rotación lateral. Ponderando ambas fuerzas se determina así la velocidad de rotación total.

7.2. Marca central no definida

Como se ha mencionado anteriormente, puede ocurrir el caso en el que no existe i_c , tal como se expone en los casos 1(c) y 2(b), con lo cual o la pelota no gira o ésta lo hace en sentido completamente normal al sentido de visión de la cámara sin que ninguna marca atraviese la zona central de la pelota.

Para estos casos se realiza el cálculo basándose únicamente en las marcas perimetrales. Al no existir i_c , los cuadros seleccionados son los primeros de la secuencia de video en donde se inicia la aparición de la pelota. En dichos cuadros, la misma posee la mayor relación centímetros/píxel; con lo que se gana así en precisión.

7.3. Realizando el cálculo

Se detalla a continuación el paso final en la tarea de calcular la velocidad de rotación de la pelota. Una vez obtenida la ubicación de las marcas en los cuadros correspondientes, realizar el cálculo final es relativamente sencillo. De manera resumida, la serie de pasos para determinar las revoluciones por segundo se sintetiza en:

- Determinar el centro de masa en coordenadas (X, Y) de la imagen de cada una de las etiquetas de la pelota en los tres cuadros a utilizar. Esto se realiza promediando separadamente en X y en Y el área de coordenadas que componen cada marca.
- Calcular distancia euclídea en píxels entre los centros de cada etiqueta a lo largo de los frames i_{c-1} , i_c e i_{c+1} .
- Determinar relación píxel / milímetros en la zona del video donde se realiza el procesamiento. Este paso es sencillo ya que se conoce de antemano el diámetro en milímetro de la pelota.
- Realizar la conversión de las distancias en píxeles a milímetros mediante la relación obtenida en el punto anterior.
- Utilizar las distancias en milímetros y los FPS del video a fin de obtener las velocidades frontal aparente y periférica aparente. Conociendo los milímetros por revolución que efectúa la pelota (calculando el diámetro de la pelota en milímetros por π), se determina de esta manera el porcentaje de rotación efectuado a lo largo de los cuadros seleccionados.
- Finalmente, se ponderan ambas velocidades de acuerdo a los casos 1a, 1b, 1c, 2a, 2b o 3; anteriormente explicados.

8. Conclusiones

La técnica detallada en el presente capítulo mantiene analogías con otros métodos para realizar estimaciones de movimiento en una secuencia de video. Sin embargo, el análisis de posibles casos y su posterior explotación para el desarrollo de la solución implementada presentan un método novel que permite calcular la velocidad de rotación de la pelota.

Como se explicará posteriormente, las pruebas realizadas arrojaron un balance positivo en cuanto a la precisión lograda por el algoritmo desarrollado con un tiempo de respuesta mayormente aceptable.

Capítulo IX: Entorno de visualización de tiros

1. Introducción

Basándose en las pautas de diseño discutidas en el capítulo 4, el presente capítulo se centra en el desarrollo de la interfaz gráfica para el módulo principal del sistema, el entorno de visualización de tiros.

La implementación del mismo fue llevada a cabo intentando respetar los principios de un correcto desarrollo de una interfaz gráfica de usuario. Es por esto que se buscó en el diseño el buen uso de la visibilidad y comprensión intuitiva de los objetos que la componen, así como maximizar la consistencia para aumentar la facilidad de aprendizaje en el uso de la misma.

2. La GUI principal

Dado que el volumen de información a mostrar en pantalla es elevado y variado (se debe mostrar información tanto gráfica como numérica), el primer problema a resolver fue determinar la distribución óptima de los componentes que conforman la vista. Se debe desplegar en pantalla una serie de *grupos visuales* de información.

El primer grupo son las tres vistas principales de la trayectoria del tiro sin perspectiva, para lograr una fácil interpretación del recorrido que ha realizado la pelota.

El segundo grupo debe mostrar en pantalla la información numérica correspondiente consistente en el tiempo con los datos mostrados en el primer grupo, tal como velocidad, altura, distancia al arco, etc.

El tercer grupo debe presentar los videos reales capturados a lo largo de la trayectoria del tiro, permitiendo así el contraste entre vistas sin y con perspectiva, además de contar con la posibilidad de observar la técnica desplegada en la ejecución del tiro por parte del deportista.

El cuarto y último grupo es el conjunto de componentes visuales que controlan la ejecución del entorno de visualización de tiros, como es la reproducción, carga y almacenamiento de los mismos.

Dado que además se cuenta con el video capturado para el cálculo de rotación, se optó por incluir el mismo como una alternativa a la reproducción de los videos originales de cada tiro, con lo cual el usuario podría alternar entre una vistas o la otra.

La figura 9.1 muestra la pantalla principal del entorno de visualización de tiros. A la derecha se encuentra el primer grupo descrito, a la izquierda el segundo, en el centro el tercero y por último en la parte superior izquierda el cuarto.

En el caso de la figura, se está mostrando la trayectoria de un tiro que se encuentra en una instancia cercana a la línea de fondo. En lugar de reproducir los videos de trayectoria, se está mostrando el video para el cálculo de rotación y un procesado de éste mostrando en detalle el seguimiento de la pelota.

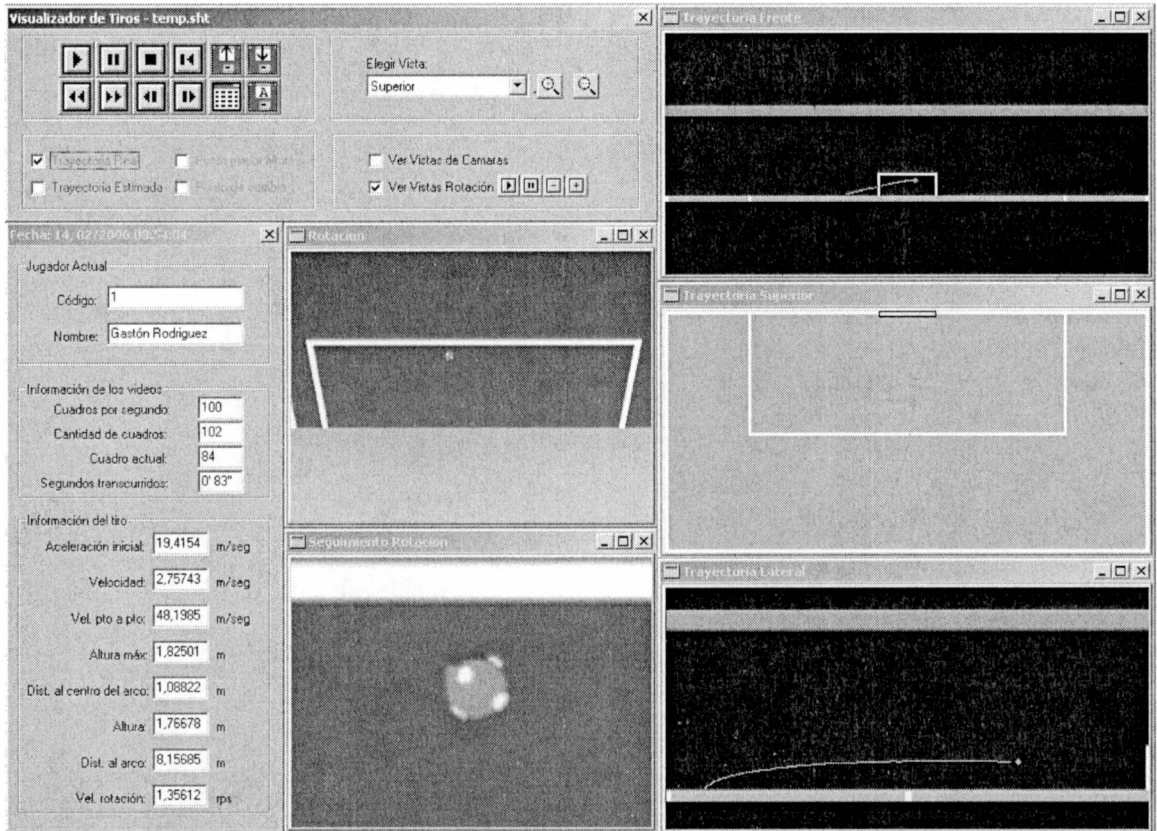


Figura 9.1 – Entorno de visualización de tiros

La figura 9.2 muestra en detalle el *grupo visual* de los componentes de control de la interfaz, los que serán explicados a continuación.



Figura 9.2 – Componentes de control

3. Componentes funcionales principales

En la parte superior izquierda se encuentran los componentes de reproducción principales.

Los ocho botones de la izquierda (zona A) permiten reproducir un tiro, pausar, rebobinar, adelantar, y observar cuadro por cuadro. Afortunadamente, debido a la funcionalidad de los mismos, típicos de un reproductor de video, la facilidad de aprendizaje y la familiaridad son implícitos en este caso.

Los restantes cuatro botones de la derecha quizás no sean tan intuitivos, aunque con el fin de disminuir la carga cognitiva se pensó en minimizar las variaciones gráficas de los mismos, a fin de reducir el tiempo de aprendizaje y fijación de éstos.

En los componentes de la zona B poseen la particularidad del diskette como imagen de fondo, el primero con una flecha “saliente” (desde el centro del diskette hacia fuera), el segundo con una “entrante” (hacia el centro del diskette) y el tercero con una letra A. Es de esperar que la funcionalidad de los primeros dos botones ya esté entendida: abrir y almacenar un tiro respectivamente.

La funcionalidad del último botón de la zona B es quizás la única que requiera una explicación aparte, dado que la representación simbólica que permite intuir el efecto del mismo puede ser cuestionable en este caso. Este componente permite almacenar un tiro de manera automática, sin tener que especificar la ruta y nombre de archivo; de acuerdo a la especificación realizada en el módulo de configuración de almacenamiento automático de tiros, que se detallará posteriormente.

Por último, el botón de la zona C, permite reordenar automáticamente las ventanas a su disposición original. Dado que se buscó maximizar la flexibilidad, los elementos que componen del entorno de visualización de tiros son ventanas que pueden ser ubicadas

en el lugar que el usuario desee, permitiéndole llevar la configuración visual a su estado inicial mediante la funcionalidad de este componente.

4. Selección de vista

En la parte superior derecha de la figura 9.2 se encuentran los componentes que permiten realizar zoom sobre alguna vista en particular para una mejor apreciación del tiro.

Dos botones representan la funcionalidad de zoom para aumentar y disminuir la vista, mientras que la selección del tipo de vista es realizada mediante la lista desplegable que se encuentra a su izquierda. La intuitividad en este caso es implícita, sin carga cognitiva asociada.

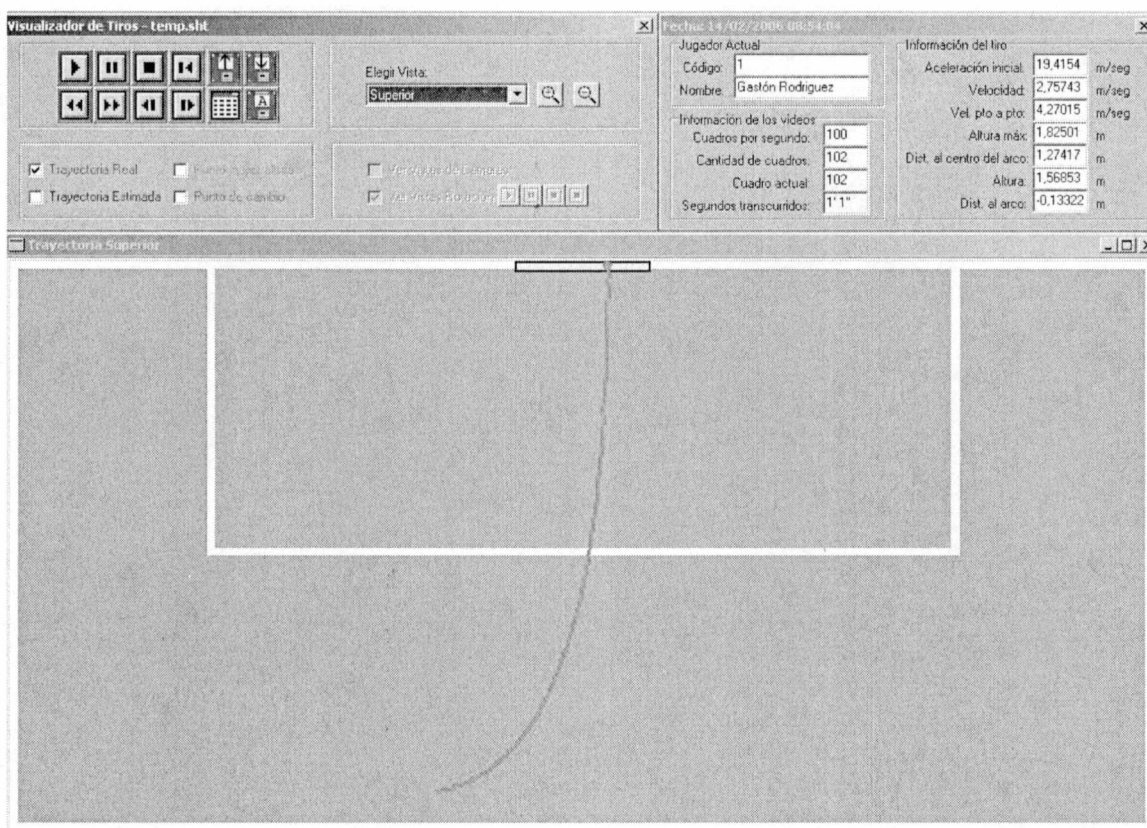


Figura 9.3 – Reproducción de un tiro en modo zoom. Vista superior

La imagen 9.3 muestra la vista superior de un tiro en su cuadro final, ampliada en modo zoom. Aunque los controles mantienen su posición original, dada la relación de aspecto de la imagen a mostrar el tiro, la información numérica debió ser ubicada en la zona superior derecha de la pantalla. El concepto de consistencia en este caso es relegado por el bien de la correcta visibilidad y el buen gusto.

5. Reproducción de videos

En la parte inferior derecha de la figura 9.2 se muestran dos opciones de reproducción de videos: los capturados por las cámaras para la triangulación o el capturado por la cámara para el cálculo de rotación.

La primer opción mostrará de manera sincronizada en el tiempo los videos de captura junto con las tres vistas.

El tiempo en el que la pelota atraviesa la zona de captura de la cámara de rotación es considerablemente menor que el tiempo total del tiro, esto lleva a separar la reproducción de la trayectoria con respecto de la reproducción de la rotación. Es por esto que la segunda opción cuenta con un reproductor exclusivo para este fin.

Cuatro botones permiten reproducir, detener, aumentar o disminuir la velocidad de reproducción del video de rotación y de seguimiento. Éstos se habilitan únicamente al seleccionar la segunda opción, con lo que la visibilidad y comprensión intuitiva de los mismos es mayormente adecuada.

5.1. Información gráfica a renderizar

En la parte inferior izquierda de la figura 9.2 se presentan cuatro opciones de renderización sobre las vistas principales. Las mismas permiten mostrar en pantalla, además de la ubicación de la pelota, la trayectoria realizada por la misma, la trayectoria

estimada según el ángulo y la aceleración inicial, el punto de cambio entre ambas, y el punto de mayor altura que adopta la pelota. Con el fin de facilitar la asociación control-renderización, los colores de las leyendas de cada opción concuerdan con la correspondiente información visual.

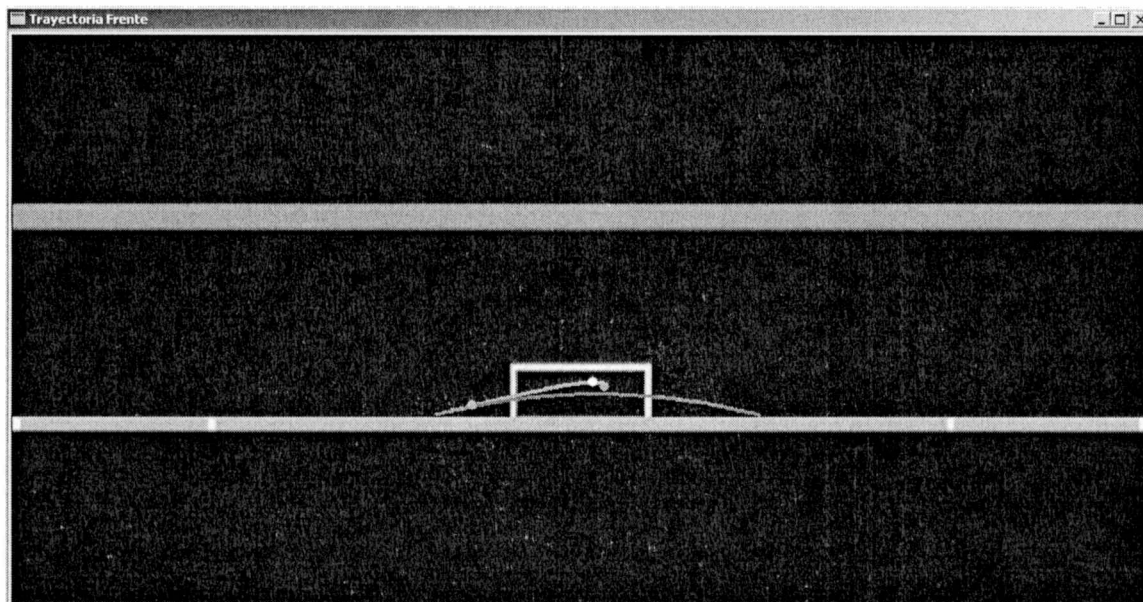


Figura 9.4 – Activadas todas las opciones de renderización

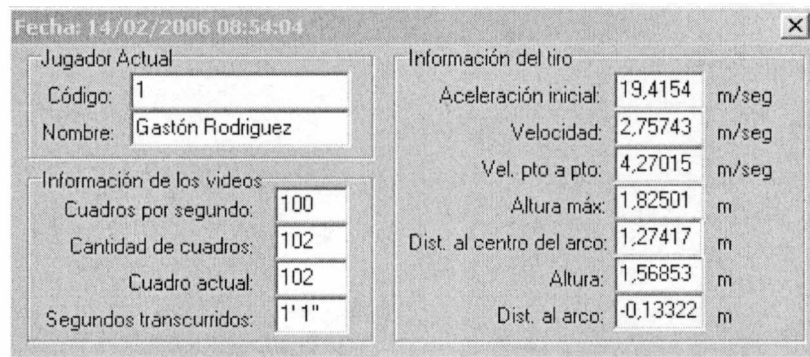
La figura 9.4 muestra la vista frontal del mismo tiro que en el mostrado en la figura 9.1, pero en este caso, con todas las opciones de renderización activadas.

5.2. La información numérica

Las trayectorias renderizadas en las vistas requieren ser acompañadas por los valores numéricos en los cuales están basadas. La figura 9.5 muestra la ventana de información en modo zoom.

Además de contener los datos relacionados con la velocidad del tiro, altura, distancia al arco, etc. se muestra la información inherente al video. Esto es, los cuadros por segundo, la cantidad total de cuadros, la posición actual en la reproducción en cuadros y en segundos.

Por último se muestran los datos de la persona ejecutante del tiro, junto con la fecha y hora de la realización del mismo.



The screenshot shows a window titled 'Fecha: 14/02/2006 08:54:04'. It is divided into two main sections. The left section, 'Jugador Actual', contains fields for 'Código: 1' and 'Nombre: Gastón Rodríguez'. Below this is 'Información de los videos' with fields for 'Cuadros por segundo: 100', 'Cantidad de cuadros: 102', 'Cuadro actual: 102', and 'Segundos transcurridos: 1'1"'. The right section, 'Información del tiro', contains numerical data for various shot parameters: 'Aceleración inicial: 19,4154 m/seg', 'Velocidad: 2,75743 m/seg', 'Vel. pto a pto: 4,27015 m/seg', 'Altura máx: 1,82501 m', 'Dist. al centro del arco: 1,27417 m', 'Altura: 1,56853 m', and 'Dist. al arco: -0,13322 m'.

Jugador Actual	
Código:	1
Nombre:	Gastón Rodríguez

Información de los videos	
Cuadros por segundo:	100
Cantidad de cuadros:	102
Cuadro actual:	102
Segundos transcurridos:	1'1"

Información del tiro	
Aceleración inicial:	19,4154 m/seg
Velocidad:	2,75743 m/seg
Vel. pto a pto:	4,27015 m/seg
Altura máx:	1,82501 m
Dist. al centro del arco:	1,27417 m
Altura:	1,56853 m
Dist. al arco:	-0,13322 m

Figura 9.5 – Información numérica

La información mostrada en ambos modos es mayormente la misma, a fin de lograr un alto grado de consistencia.

6. Adiciones funcionales que facilitan el uso de la aplicación

A fin de maximizar la usabilidad, se buscó automatizar ciertos aspectos repetitivos de la aplicación con el fin de reducir los tiempos de respuesta e interacción.

Dos aspectos que se consideran repetitivos en este tipo de aplicación es la carga de los videos y el almacenamiento de los tiros. Especificar en cada ocasión la ubicación de los mismos atenta contra el control del usuario y la flexibilidad.

La figura 9.6 muestra el módulo de configuración de almacenamiento automático de tiros. El mismo permite indicar la ruta y nombre de archivo que deberá tener cada tiro al momento de almacenarlos. Esta funcionalidad reduce considerablemente el tiempo de interacción con la aplicación.

Además de poder especificar la ruta en la que se deben almacenar los archivos, se puede indicar si se desea incorporar en el nombre del archivo, la fecha y hora, el código del jugador y el nombre del mismo. Inclusive el orden de éstos puede ser modificado mediante las flechas correspondientes.

La separación entre cada parte que compone el nombre del archivo puede ser separada mediante algún carácter especial, el cual también puede ser especificado. Cada vez que se modifica la configuración, el resultado es mostrado inmediatamente en la vista preliminar. Esto hace de la aplicación una herramienta versátil y de gran flexibilidad.

La figura 9.7 muestra el módulo de configuración de carga automática de videos. En este caso, se pueden especificar las rutas y nombres de archivo de cada uno de los videos que componen un tiro capturado. Mediante esta facilidad, la carga de los mismos para su posterior procesamiento se reduce simplemente a invocar la opción correspondiente.

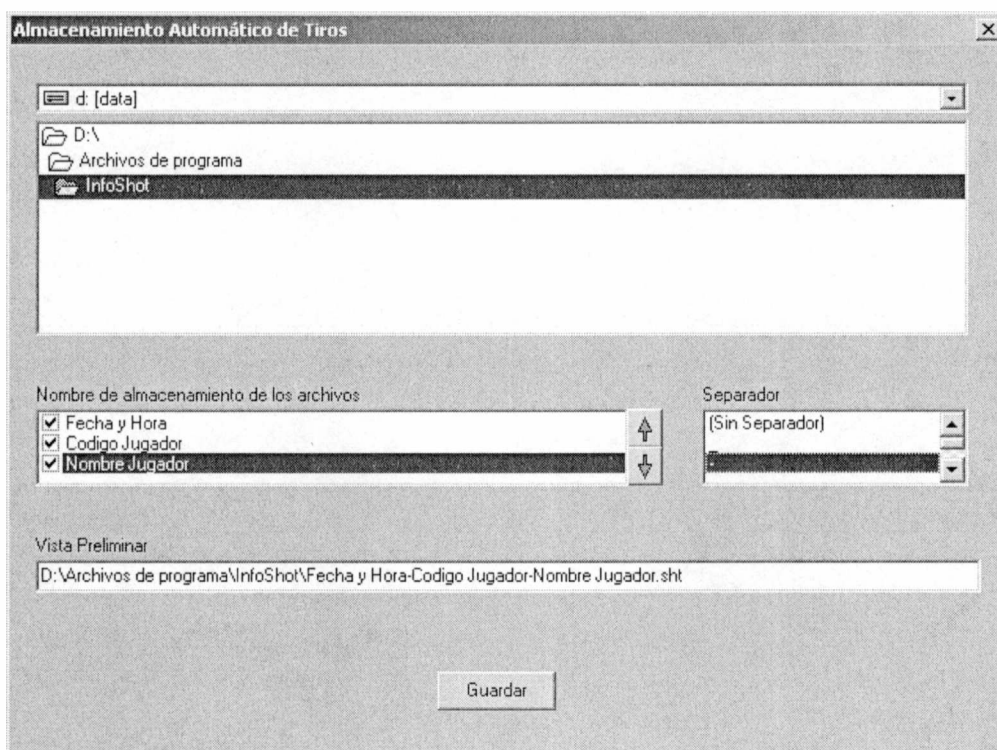


Figura 9.6 – Módulo de configuración. Almacenamiento automático de tiros.

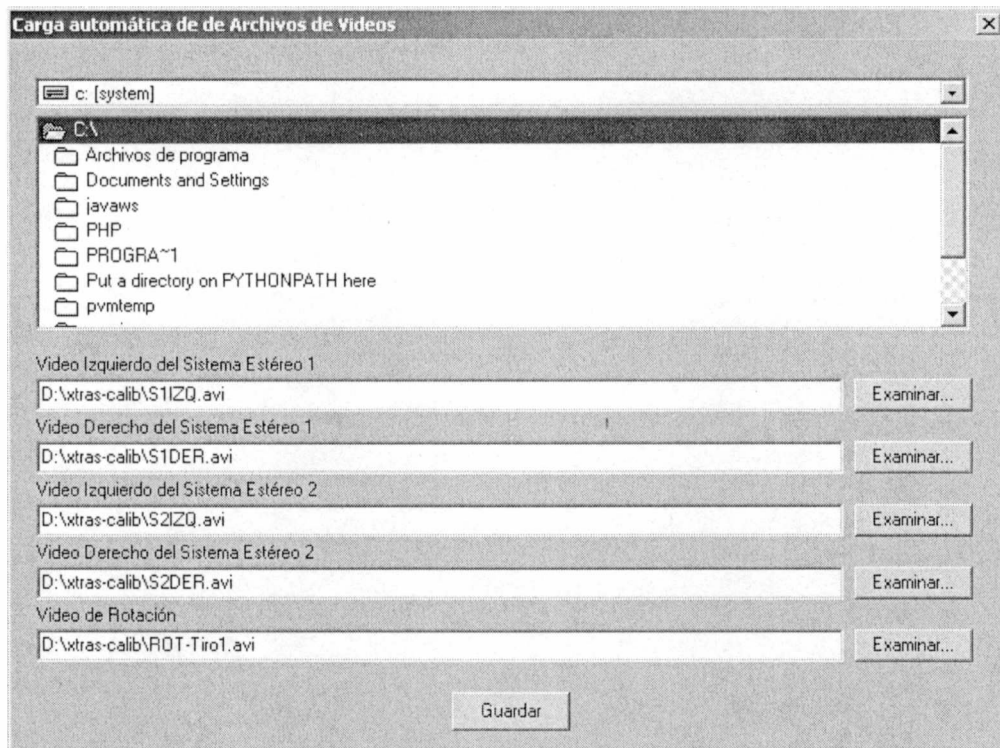


Figura 9.7 –Módulo de carga automática de videos.

7. Aspectos técnicos

7.1. Almacenamiento y carga de tiros

Una vez procesados los videos, la información que se despliega en el entorno de visualización de tiros puede ser almacenada en un formato desarrollado ad-hoc para la aplicación.

Los archivos, definidos con extensión .sht poseen la estructura de la figura 9.8. La misma es detallada a continuación, indicando qué referencia cada dato de ésta. Es interesante detallar algunos de los datos con el fin de facilitar la comprensión de los mismos.

El punto (1) almacena la posición en el vector de puntos que contiene la información con la mayor altura que adopta la pelota.

El punto (2) almacena la posición en el vector de puntos en la cual se realizó el cambio de captura del sistema 1 al sistema 2.

El punto (3) almacena la distancia en milímetros desde el arco en la que se configuró el cambio de sistema.

El punto (4) almacena la cantidad total de cuadros que conforman los videos de captura. La aplicación se encarga de realizar el recorte, esto es: eliminar los cuadros iniciales que no contengan movimiento, y eliminar los cuadros posteriores a que la pelota llegue a la línea de fondo.

El punto (5) almacena los cuadros por segundo que poseen las capturas.

El punto (6) almacena la velocidad promedio del tiro en milímetros por segundo.

El punto (7) almacena la aceleración inicial del tiro en milímetros por segundo.

El punto (8) almacena las revoluciones por segundo que adoptó la pelota.

Los puntos (9) y (10) almacenan las longitudes en bytes de los videos de la cámara izquierda y derecha respectivamente, utilizando primero el sistema 1 y posteriormente el sistema 2.

Los puntos (11) y (12) almacenan las longitudes en bytes del video de rotación y de seguimiento respectivamente.

Los puntos (13) y (14) almacenan el identificador del jugador que realizó el tiro y la fecha y hora en que se ejecutó.

```

struct datosStruct
{
    int mayorAltura;                (1)
    int puntoDeCambio;             (2)
    int cambioSistema;             (3)
    int cantFrames;                (4)
    int FPS;                       (5)
    float velocidad;               (6)
    float acelInicial;             (7)
    float RPS;                     (8)
    int longitudAviIzq;            (9)
    int longitudAviDer;            (10)
    int longitudAviRotacion;       (11)
    int longitudAviSeguimiento;    (12)
    int idJugador;                 (13)
    TDateTime fechaHora;           (14)
    float velocidades[300];         (15)
} datos;

struct puntos3DStruct
{
    CvPoint3D32f puntos3DReal[300]; (16)
    CvPoint3D32f puntos3DEst[300];  (17)
} puntos3D;

struct puntos2DStruct
{
    CvPoint2D32f puntosIzq[300];    (18)
    CvPoint2D32f puntosDer[300];    (19)
} puntos2D;

```

Figura 9.8 – Estructura del archivo de tiro

El punto (15) almacena las velocidades punto a punto en milímetros por segundo que adoptó la pelota en el recorrido.

Los puntos (16) y (17) almacenan la posición real y estimada respectivamente de la pelota en formato (X, Y, Z) a lo largo del tiro. Estos valores son indicados con respecto al centro de coordenadas (0, 0, 0) especificado en capítulos anteriores (centro del arco, a nivel del piso). Ésta es la información (usada de a pares (X,Y), (Y,Z), (X,Z)) para renderizar la trayectoria de la pelota sobre las tres vistas principales sin perspectiva.

Los puntos (18) y (19) almacenan la posición de la pelota en píxeles (X, Y) sobre los videos capturados a lo largo de la trayectoria del tiro. Ésta información es utilizada para renderizar la trayectoria de la pelota sobre las capturas, como se muestra en la figura 9.9.

Cabe mencionar que se definió un array de 300 posiciones ya que un tiro de estas características no supera los dos segundos de duración, con lo cual, capturando 100 cuadros por segundo, el espacio es más que suficiente.



Figura 9.9 – Renderización sobre la captura de la cámara derecha del sistema 2

Los videos capturados en un tiro poseen la particularidad de contar con alta redundancia temporal. Principalmente existe un único objeto en movimiento (la pelota) y el fondo de cada video es el mismo todo el tiempo, (despreciando la información del jugador ejecutante del tiro).

En este caso, las técnicas de compresión temporal son perfectas para reducir el tamaño de los mismos a una ínfima parte de los originales. Es por esto que los videos almacenados en los archivos de tiro se encuentran comprimidos mediante codecs DivX.

El tamaño de un archivo de este tipo se encuentra en los 170 KiloBytes. Hay que recordar que los mismos pueden llegar a almacenar hasta 1200 cuadros de video, además de los vectores mencionados y metainformación necesaria para la utilización de éstos.

Los tipos de datos no convencionales utilizados (por ejemplo CvPoint3D32f) corresponden a los definidos en la librería de visión por computadora OpenCV.

7.2. Renderización sobre vistas y videos

Utilizando la información almacenada en los vectores, se realiza la renderización de la pelota y su correspondiente trayectoria sobre la imagen de fondo correspondiente a cada vista, la cual es mostrada en pantalla en cada nuevo cuadro a mostrar y antes de iniciar la renderización.

La pelota es dibujada mediante un círculo en cada una de las vistas superior, lateral y frontal basándose en la información (X, Z) , (Z, Y) y (X, Y) respectivamente, correspondiente al vector de posiciones 3D en la ubicación del mismo cuyo índice es el cuadro que debe ser mostrado.

La renderización de las trayectorias requiere un procesamiento más complejo. Dada la gran cantidad de puntos triangulados a lo largo de un tiro, la posición de la pelota entre un cuadro y el siguiente difiere en aproximadamente unos 30cm. Esto permite dibujar líneas entre cada triangulación, ya que visualmente se obtiene una curva. Para lograr esto sobre las tres vistas, se recorre el vector de posiciones 3D (basándose nuevamente en los pares (X, Z) , (Z, Y) y (X, Y)) desde el inicio del tiro hasta el cuadro k -ésimo que debe ser mostrado, dibujando las líneas correspondientes entre los frames $[1, 2]$, $[2, 3]$... $[k-1, k]$.

En cada caso debe realizarse la conversión de los datos en milímetros del vector de posiciones a píxeles en pantalla según el factor de escala, de acuerdo al modo en que se este realizando la reproducción del tiro (modo normal o zoom).

La renderización sobre los videos es similar. Primeramente se obtiene el cuadro del video correspondiente a ser mostrado en pantalla, y luego se dibuja la información requerida basándose en el vector de posiciones 2D. En estos casos no es necesario realizar conversión alguna.

7.3. Generación del video de seguimiento

El video de seguimiento resulta de procesar la captura realizada para el cálculo de rotación. Para cada cuadro de la secuencia, una vez determinado el centro de la pelota, se realiza un recorte alrededor de la misma. Éste es redimensionado mediante interpolación bicúbica. La imagen resultante es incorporada a la nueva secuencia de video.

El efecto obtenido produce al usuario la sensación de contar con una cámara motorizada que realiza el seguimiento de la pelota en tiempo real. La facilidad de observación de las marcas de la pelota en el video de seguimiento es notoriamente superior con respecto a la captura de rotación original.

8. Conclusiones

Bajo los lineamientos discutidos en los capítulos de la primera parte, se llevó a cabo de manera satisfactoria la correcta implementación del prototipo, junto con su interfaz gráfica de usuario.

A lo largo de los testeos realizados (detalladas en el capítulo X), no se presentaron mayores complicaciones en la utilización del sistema. Las herramientas de

automatización que ésta provee permitieron además reducir notablemente los tiempos operativos de las pruebas llevadas a cabo.

Capítulo X: Pruebas realizadas

1. Introducción

El proceso de pruebas es esencial para asegurar la calidad del sistema. El testing realizado fue llevado a cabo en cada etapa del desarrollo del sistema, con la intención de evitar propagar errores a etapas posteriores. Se realizaron pruebas en cada uno de los módulos desarrollados de manera independiente.

Sin embargo, una vez finalizado el desarrollo fue necesario probar el sistema como un todo a fin de determinar la precisión del mismo bajo diferentes circunstancias variables de la escena.

Dado que no se cuenta con el hardware necesario para realizar las pruebas en un ambiente real, las mismas fueron realizadas utilizando el modelo desarrollado en 3D Studio.

Se realizaron una serie de pruebas que determinaron las condiciones óptimas necesarias para que el sistema obtenga la mayor precisión posible. A continuación se revisan los aspectos relevantes al momento de la utilización de la aplicación desarrollada.

2. Simulaciones

En primera instancia se analizó en detalle la precisión del sistema bajo diferentes estados de la escena. Se realizaron pruebas con la mayoría de las configuraciones de escena posibles, tomando en cuenta los siguientes factores:

- Disposición de las cámaras
- Diferentes grillas de calibración de cámaras.
- Iluminación de la escena.
- Resolución de los videos.
- Velocidad de captura.

A continuación se detalla la relevancia de cada uno de estos factores, que incide considerablemente en los resultados obtenidos.

2.1. Disposición de las cámaras

Como se explicó en el capítulo 5, en primera instancia se utilizó un único sistema estéreo en la parte posterior de la zona de captura. Luego se decidió utilizar dos sistemas estéreo, uno para la parte inicial del tiro libre y el restante para la parte final del mismo. Utilizando esta especificación de hardware se logra mayor precisión en la triangulación y se evita problemas de oclusión por parte de los jugadores.

Una vez establecido el uso de dos sistemas estéreo, fue necesario determinar cual es la mejor ubicación de las cámaras dentro de cada sistema. En otras palabras, se necesitó determinar el ángulo que deben formar las cámaras entre sí, para obtener el menor error posible en la reconstrucción de la trayectoria del tiro libre.

Sin embargo, dicho ángulo influye directamente sobre el tamaño del campo de visión del sistema estéreo. Se debe tener en cuenta que el tiro libre puede ser ejecutado desde cualquier punto cercano (externo) al área de meta. Por lo tanto, el ángulo adoptado

debe permitir un campo de visión en el que sea posible visualizar al menos el ancho del área de meta desde cada una de las cámaras. El área de meta tiene una medida aproximada de 16 metros de largo por 40 metros de ancho.

Teniendo en cuenta estos factores se realizaron pruebas con diferentes ángulos: 75, 90 y 105 grados. En la figura 10.0 se observan dichas configuraciones de posicionamiento de las cámaras. La mejor disposición resultó ser la que forma un ángulo aproximado de 90 grados entre las cámaras, ya que logra un campo de visión suficiente para capturar el área de meta y, al mismo tiempo, presenta una mayor precisión al momento de la triangulación. Dicha disposición es la que finalmente se adoptó en el presente trabajo. En la sección “Datos obtenidos” del presente capítulo se encuentra el detalle de la precisión de la triangulación con la disposición adoptada.

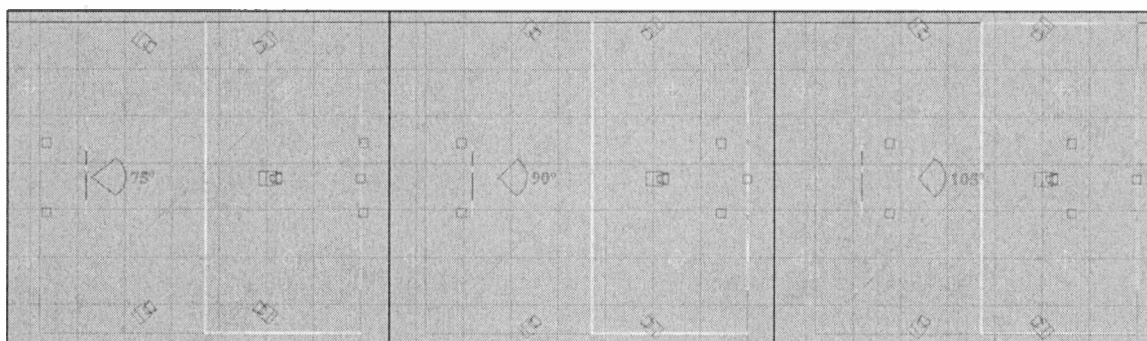


Figura 10.0 - Pruebas de disposición de las cámaras

2.2. Grilla de calibración

La grilla utilizada en el proceso de calibración presenta una incidencia directa sobre la posterior triangulación de los puntos en el espacio tridimensional. Existen tres factores que alteran la precisión de la calibración: el tamaño de los cuadros de la grilla de calibración, la cantidad de cuadros y la posición de la misma.

El tamaño de los cuadros de la grilla depende de la profundidad de la escena que se trata de capturar. En el presente trabajo, el sector de entrenamiento es un cuadrado de 34

metros de lado. Sin embargo, al contar con dos sistemas estéreo, cada uno de ellos debe tener en cuenta un sector de captura rectangular de 34x17 metros aproximadamente.

Teniendo en cuenta estas medidas, se realizaron una serie de calibraciones utilizando grillas con diferentes tamaño de cuadros, desde los 30cm hasta los 100 cm. Los mejores resultados se obtuvieron usando las grillas con cuadrados de un tamaño comprendido entre los 75cm y los 100cm.

El siguiente punto a tener en cuenta fue la cantidad de cuadros internos necesarios. Teniendo en cuenta el tamaño de los cuadros, la cantidad de cuadros posible se vio acotada. Esto se debe a que una gran cantidad de cuadros hace difícil la manipulación de la grilla debido a su gran tamaño.

Por otro lado, el sistema de calibración se basa en las esquinas internas, sin tener en cuenta las del contorno. Se necesita además una cantidad razonable de esquinas internas para poder realizar una calibración precisa.

Luego de realizar pruebas variando la cantidad de cuadros, se llegó a la conclusión de que es necesario mínimamente una grilla rectangular que contenga 24 esquinas internas (o sea, una grilla de 7 cuadros de ancho y 5 cuadros de alto).

El último aspecto a tener en cuenta es la posición de la grilla dentro de la escena. La misma fue posicionada a diferentes distancias de la cámara y se determinó que se logra una precisión mayor si ésta se ubica a una distancia promedio de profundidad. Teniendo en cuenta que cada sector de captura de un sistema estéreo posee 17 metros de profundidad, la grilla se debe situar a los 8,5 metros aproximadamente.

En la figura 10.1 se puede observar una grilla con las características especificadas. La misma contiene 8 cuadros de ancho y 6 cuadros de alto, donde cada cuadro mide 75cm. La grilla se encuentra además en una posición óptima, es decir, a una profundidad promedio del sector de captura.

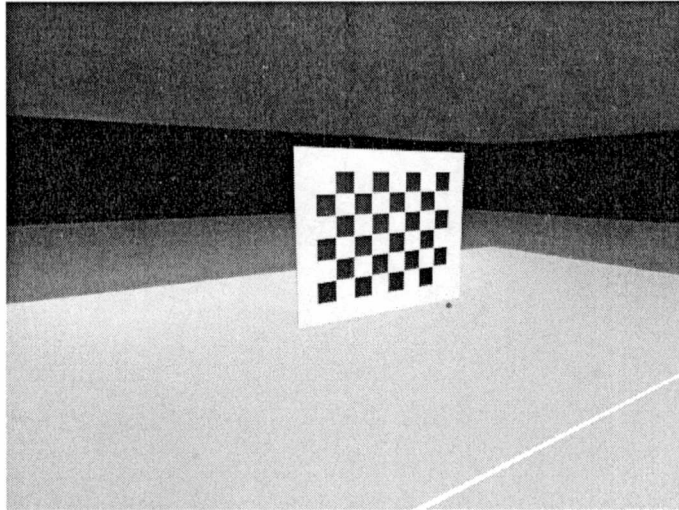


Figura 10.1 - Una grilla de calibración adecuada para el sistema desarrollado

2.3. Iluminación

Se probaron diferentes iluminaciones de la escena, tanto utilizando spots como luz omnidireccional. Idealmente se debe contar con una iluminación uniforme de la escena.

Un aspecto crucial en la iluminación es que ésta debe ser lo suficientemente potente como para permitir capturas a una velocidad de obturación de $1/1000$ de segundo. Los reflectores utilizados por ejemplo en canchas de fútbol 5 son suficientes para cumplir con esta condición. Esto es aproximadamente, 6 reflectores de 2kW cada uno.

3. Datos obtenidos

A continuación se puede observar una lista de los datos obtenidos a partir de la serie de pruebas que se han efectuado. Primeramente se especifican las pruebas y resultados de la reconstrucción de la trayectoria. Posteriormente se hace lo propio con las pruebas y resultados del cálculo de velocidad de rotación de la pelota.

Para las pruebas de triangulación de la posición de la pelota se generaron una serie de supuestos tiros libres en el ambiente virtual 3D. Dado que la posición de la pelota en dicho entorno es conocida en cualquier instante de tiempo, es posible verificar el error cometido en la triangulación de cualquier punto de la trayectoria del tiro libre.

Debido a que se realizaron cerca de cien tiros de prueba variando el punto de origen y el punto final del tiro, la velocidad de traslación y rotación de la pelota, se presenta una lista ejemplo de los datos obtenidos en algunos de los tiros de prueba virtuales.

Nº Tiro	Error máximo de triangulación (punto obtenido a mayor distancia de las cámaras)	Error promedio de triangulación (basado en la totalidad de los puntos obtenidos)
Tiro 1	37 mm	25 mm
Tiro 2	34 mm	22 mm
Tiro 3	29 mm	17 mm
Tiro 4	31 mm	20 mm
Tiro 5	33 mm	20 mm
Tiro 6	28 mm	15 mm

Para obtener estos valores se calculó la distancia euclídea entre el punto original (posición de la pelota en el ambiente 3D) y el punto triangulado (posición de la pelota reconstruida).

Mediando los errores máximos de triangulación de todos los tiros de prueba efectuados, se obtuvo un error máximo general de 34,31 mm de promedio. Mediando los errores promedio de triangulación de todos los tiros de prueba efectuados, se obtuvo un error promedio de 20,68 mm de promedio.

Para las pruebas del cálculo de velocidad de rotación de la pelota se realizaron pruebas similares en el ambiente virtual 3D. En este caso, se generaron tiros con diferentes tipos de rotación sobre los ejes X , Y , Z con el fin de verificar la precisión del método de

cálculo desarrollado en todos los casos posibles. Los diferentes casos de rotación fueron analizados en detalle en el capítulo 8.

A continuación se brinda una lista de los datos obtenidos en algunos de los tiros de prueba virtuales. A modo de ejemplo se presenta una muestra por cada caso de rotación posible.

Nº Tiro	Tipo de Rotación	Error
Tiro 1	Rotación lateral, con marca central inmóvil	0,2 rps
Tiro 2	Rotación lateral, con marca central atravesando parcialmente el centro de la pelota	0,9 rps
Tiro 3	Rotación lateral, sin marca en el centro de la misma	0,5 rps
Tiro 4	Rotación frontal, con marca central atravesando el centro de la pelota	0,8 rps
Tiro 5	Rotación frontal, sin marca en el centro de la misma	0,6 rps
Tiro 6	Sin rotación aparente	0,3 rps

Para obtener estos valores, se calculó la diferencia entre la rotación real (que es conocida en el ambiente 3D) y la calculada por la aplicación. Mediando los errores de cálculo de todos los tiros de prueba efectuados, se obtuvo un error promedio de 0,74 rps.

4. Tiempos de procesamiento

Los tiempos de procesamiento no son despreciables, dada la cantidad de información que es utilizada tanto para calcular la trayectoria lograda como para la velocidad de rotación de la pelota.

Un tiro libre típico tiene una duración no mayor a 1,5 segundos. Sin embargo, los videos que actúan como entrada del sistema contienen tanto las imágenes del disparo, como un conjunto de imágenes que corresponden a los instantes previos y posteriores al mismo.

La solución desarrollada trata de minimizar los tiempos de procesamiento con el objetivo de tornar mas dinámica la utilización de la aplicación. Esto se logra desechando los cuadros de los videos que no pertenezcan al tiro en si mismo, mediante un pequeño procesamiento que determina el inicio y el final del tiro según la posición y movimiento de la pelota.

De esta forma, la cantidad de cuadros a procesar en cuanto a los videos correspondientes al cálculo de trayectoria, será cercano a los 150 (1,5 segundos * 100 cuadros por segundo).

A esta cantidad se le debe adicionar el procesamiento de los cuadros pertenecientes al video de rotación. Si bien el cálculo de la velocidad de rotación se realiza en base a sólo tres cuadros, el mismo conlleva tareas adicionales como encontrar los cuadros ideales para la estimación y realizar el seguimiento de las marcas en el tiempo.

A continuación se puede observar una tabla con los tiempos de procesamiento finales logrados, acorde a los casos vistos en el apartado anterior.

Nº Tiro	Tiempo de procesamiento (segundos)
Tiro 1	45,3
Tiro 2	51,4
Tiro 3	43,2
Tiro 4	43,7
Tiro 5	48,3
Tiro 6	47,0

Para obtener estos valores, se cronometró cada uno de los tiempos de procesamiento de los tiros simulados en el ambiente 3D. Mediando los tiempos de procesamiento de todos los tiros se logró un tiempo de procesamiento promedio de 44,28 segundos.

5. Conclusiones

Las pruebas realizadas en el ambiente virtual 3D fueron de gran utilidad para obtener las características imprescindibles con las que debe contar la escena. Las mismas corroboraron la correctitud de los algoritmos de procesamiento.

Como futura tarea, resta principalmente mejorar los tiempos de respuesta de cada uno de los algoritmos de procesamiento, con el fin de lograr un sistema de mayor agilidad o incluso, de ser posible, una aplicación en tiempo real.

Conclusiones

1. Introducción

El presente capítulo compendia los objetivos inicialmente planteados y las metas alcanzadas. Posteriormente se plantea un conjunto de posibles mejoras y ampliaciones del sistema desarrollado. Finalmente se presenta una visión a futuro en cuanto a aplicaciones alternativas relacionadas con las técnicas utilizadas en el presente trabajo.

2. El prototipo

El desarrollo realizado tuvo como resultado un sistema totalmente funcional. El mismo es una solución cerrada, la cual recibe como entrada cinco secuencias de video de un tiro efectuado, y devuelve la información numérica y gráfica correspondiente a la traslación y rotación que sufrió la pelota en el mismo.

Mas aún, los módulos de configuración del sistema (calibración, filtrado, etc.) proveen la portabilidad del mismo. Mediante una simple reconfiguración, el sistema se encuentra listo a ser utilizado en diferentes predios de captura de tiros, los cuales pueden variar en posicionamiento de cámaras, luminosidad, etc.

La visualización en pantalla de la trayectoria realizada carente de perspectiva simplifica la interpretación para el ejecutante del tiro y facilita las explicaciones al entrenador. Ésto combinado con la información de rotación de la pelota permite interrelacionar la

incidencia de ésta sobre la trayectoria y determinar así las mejores alternativas de tiro que el ejecutante puede lograr.

La información numérica puede ser usada para realizar comparaciones entre tiros a lo largo del tiempo, corroborando la evolución del jugador en cuanto a precisión y velocidad de los tiros efectuados. A su vez, la comparación también es posible entre tiros de diferentes ejecutantes; determinando así cual de éstos se encuentra en mejor estado.

3. OpenCV

La utilización de una herramienta de trabajo potente y versátil como es la librería OpenCV permitió abstraerse de los problemas típicos en el procesamiento de imágenes y facilitó la tarea de enfocarse de lleno en las tareas a resolver.

Sin embargo, a los fines de ésta aplicación, se realizaron ciertas modificaciones a la librería a fin de adaptarla a los requerimientos específicos del sistema. Las mismas sirvieron para mejorar los resultados obtenidos en la calibración y mostrar los resultados de la misma de una manera más intuitiva.

Aunque en teoría la calibración de cámaras puede ser utilizada a cualquier escala, dada la gran distancia entre las cámaras y el tablero de calibración (y el tamaño de este último), el límite empírico del perímetro mínimo permitido para los cuadrados del tablero y la distancia mínima entre los puntos del patrón de la imagen fueron cambiados a valores más convenientes con las dimensiones utilizadas.

En caso contrario, la librería podría llegar a interpretar que la calibración realizada había sido efectuada de manera errónea, dada la distancia obtenida como resultado en cuanto al origen de coordenadas situado en la grilla de calibración.

Otra modificación realizada fue la indicación del inicio de coordenadas sobre la grilla de calibración mediante la palabra “ORIGEN”. De esta manera, conocer el (0,0,0) se hace más intuitivo para el usuario que está utilizando el sistema. La figura 11.1 muestra esta modificación en un sistema recién calibrado.

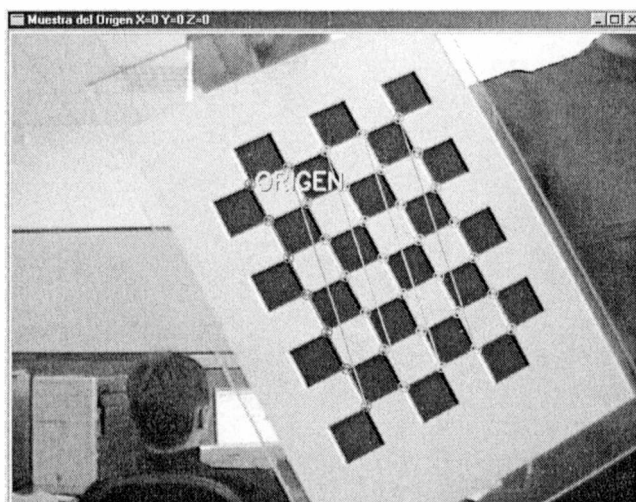


Figura 11.1 Indicación del inicio de coordenadas espaciales

4. Ampliaciones

Se prevee ampliar el sistema con nuevas funcionalidades que se espera facilitarán aun más la utilización del sistema y aumentar la potencialidad gráfica del módulo de visualización de tiros.

La primera de las ampliaciones se relaciona con el modo en que los videos son incorporados al sistema. Actualmente el sistema no se encarga de realizar la interacción con las cámaras; con lo cual una aplicación aparte debe ser usada para realizar la captura de los videos. Permitiendo el inicio y la detención de las capturas desde el mismo sistema facilita la tarea al usuario e inclusive permitiría por ejemplo limitar un intervalo fijo de captura, dado que el tiempo de cada tiro no supera los 2 segundos. Una

vez realizada la captura del tiro, automáticamente se iniciaría la tarea de procesamiento de los videos.

Otra ampliación se relaciona con los módulos de filtrado. Actualmente los mismos aceptan imágenes para realizar la configuración de los umbrales. Sin embargo sería ideal poder verificar que dichos valores sean los óptimos a lo largo de toda la zona de captura. Utilizando videos en lugar de imágenes, se podría verificar fácilmente que se cumpla esta condición, mostrando en tiempo real los resultados del filtrado junto con la reproducción del video original.

La última ampliación, y tal vez la más ambiciosa, es adicionar un módulo de renderización 3D de los tiros. De esta forma, se podrían visualizar los mismos desde un sin fin de posiciones. Por ejemplo, el usuario podría indicar la posición de la cámara en el espacio, rotarla, acercarse o alejarse de la escena. La reproducción de un tiro bajo este módulo también permitiría las funciones de pausa, reproducir en cámara lenta, etc.

Esta ampliación no intercede en absoluto con la implementación realizada hasta el momento. Mas aún, las estructuras de datos desarrolladas permiten adiciones como la planteada con un mínimo de esfuerzo. Dado que los archivos de tiro poseen la información ya procesada, incorporar el módulo en cuestión solo requeriría la correcta lectura de dichos datos a fin de utilizarlos posteriormente para la renderización del entorno 3D de visualización.

Un aspecto interesante de esta funcionalidad sería por ejemplo observar los tiros desde el punto de vista del arquero. Este entrenamiento visual permitiría a los mismos notar los tiempos necesarios de reacción, visibilidad de la pelota desde ciertos ángulos, incidencia de una posible barrera de jugadores, etc.

5. A futuro

Aunque no se encuentra relacionado con actividades de entrenamiento deportivo, se ha analizado una posible utilización en una aplicación específica de los módulos concernientes con la estereoscopia y reconstrucción 3D.

Determinar la ubicación en el espacio (X, Y, Z) de la pelota mediante dos imágenes estereo, en las cuales la posición de la misma se da a través de dos pares de puntos ((X_1, Y_1) y (X_2, Y_2) respectivamente) es, de manera sintética, la base del presente desarrollo en cuanto al seguimiento de la trayectoria que efectúa la pelota en dirección al arco.

En este caso, la cantidad de puntos a triangular en cada par de imágenes es mínima: tan solo uno, el centro de masa de la pelota. Pero extrapolando este criterio se podrían triangular varios puntos de una captura estereo. Por lo tanto, los fundamentos para la reconstrucción tridimensional de la pelota a lo largo del tiempo pueden ser utilizados como base para realizar una reconstrucción espacial 3D de una superficie basándose en un par de imágenes estereo.

Los scanners 3D realizan este tipo de tareas, pero en la mayoría de los casos se utilizan técnicas basadas en medición de distancias mediante láser o luz estructurada. La alternativa propuesta se apoyaría exclusivamente en dos cuestiones principales: la triangulación de puntos (la cual ya ha sido discutida a lo largo del presente trabajo) y la correlación de puntos entre las dos imágenes estereo.

Este último punto se presentaba implícitamente en el seguimiento de la pelota, ya que una vez filtrada la misma en ambas imágenes, los centros obtenidos en cada una indicaban la asociación espacial de las tomas izquierda y derecha.

En el caso de triangular varios puntos de una imagen, la complejidad de correlación de puntos aumenta. Determinar qué punto (X_2, Y_2) de la imagen derecha corresponde a al punto (X_1, Y_1) de la imagen izquierda puede realizarse únicamente mediante técnicas de flujo óptico¹ o de asociación por bloques², las cuales fueron explicadas al inicio de este trabajo.

Determinando una gran cantidad de puntos (X, Y, Z) , se obtiene una nube de puntos, la cual puede ser utilizada con el fin de generar el mallado correspondiente. Una vez obtenido, es posible visualizar este último mediante un entorno de reproducción 3D con técnicas similares a la planteada para la visualización 3D de tiros.

¹ Del inglés optical flow

² Del inglés, block matching

Bibliografía

Digital Video Processing. A. Murat Tekalp. Prentice Hall. 1995

Visión por Computador, Fundamentos y Métodos. Arturo de la Escalera. Prentice Hall. 2001

Introductory Techniques for 3D computer vision. Emanuele Trucco. Alessandro Verri, Prentice Hall 1998.

Epipolar Geometry in Stereo, Motion and Object Recognition, A Unified Approach. Xu Zhang. Kluwer Academic Publishers 1996.

Multiple View Geometry in computer vision. Richard Hartley. Andrew Zisserman. Cambridge University Press 2000.

Machine Vision. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck. Mac Graw. Hill Internacional Edition. 1995

Seguimiento tridimensional usando dos cámaras. Pablo Barrera González, José María Cañas Plaza. 2004.

Search Algorithms for Block-Matching in Motion Estimation, Deepak Turaga, Mohamed Alkanhal. Mid-Term project 18-899 Spring, 1998.

http://www.ece.cmu.edu/~ee899/project/deepak_mid.htm

Optical Flow. Raúl Varas Martínez.

<http://www.elai.upm.es/spain/Investiga/GCII/personal/rvaras/opticalflow.htm>

Optical Flow. David Marshall.

http://www.cs.cf.ac.uk/Dave/Vision_lecture/node45.html

Modelo de color RGB y CMYK

<http://www.fotonostra.com/grafico/rgb.htm>

La historia de las Interfaces Gráficas de Usuario (GUI)

<http://www.pilu.com>

Fundamentos de la cámara fotográfica

<http://mecfunnet.faii.etsii.upm.es/difraccion/camera.html>

Bayer mosaic definition

http://www.matrix-vision.com/info/articles/pdf/art_bayermosaic_e.pdf

Digital camera basics

<http://www.bobatkins.com/photography/digital/digitalbasics.html>

Distancia focal y tamaño de la imagen

<http://fisicarecreativa.net/comofunciona/comofunciona06.html>

The Design of Graphic User Interfaces Home Page

www.sju.edu/~jhodgson/gui/guihome.html

The Core of Information Technology

www.cne.gmu.edu/itcore/userinterface/index.html

A Summary of Principles for User-Interface Design

<http://www.sylvantech.com/~talin/index.shtml>

Designing a Graphical User Interface

<http://www.medicalcomputingtoday.com/0agui.html>

Estimación de la Geometría en Visión por Computador

<http://iie.fing.edu.uy/investigacion/grupos/gti/cursos/egvc/>

Projectile Trajectory of a Cannonball

<http://www.physics247.com/physics-tutorial/projectile-trajectory.shtml>

Projectile motion applet

<http://www.walter-fendt.de/ph11e/projectile.htm>

Técnicas de reconstrucción 3D

<http://agamenon.uniandes.edu.co/~pfiguero/Rec3D-05III/>

OpenCV - Open Source Computer Vision Library Community.

<http://groups.yahoo.com/group/OpenCV>

Camera Calibration Toolbox for Matlab.

www.vision.caltech.edu/bouguetj/calib_doc/index.html



BIBLIOTECA
FACULTAD DE INGENIERIA

DONACION..... FACULTAD.....
\$.....
Fecha..... 12-3-08
Inv. E..... Inv. B.....
003103

TES
07/15
003103

TES
07/15
DIF-03103
SALA



UNIVERSIDAD NACIONAL DE LA PLATA
Facultad de
Educativa
50 y 120 La Plata
cathec@info.unlp.edu.ar
biblioteca@info.unlp.edu.ar



DIF-03103