



Proyecto de Fin de Máster.
Curso 2007-2008.

OBTENCIÓN DE DISTANCIAS Y
POSICIONES GEOGRÁFICAS MEDIANTE
TÉCNICAS DE VISIÓN ARTIFICIAL

Autor:

David Romero Laorden

Director del proyecto:

Jose Antonio López Orozco

Facultad de Informática.
Universidad Complutense de Madrid.

Obtención de distancias y posiciones geográficas mediante técnicas de Visión Artificial.

Memoria del proyecto de fin de Máster presentado por David Romero Laorden en la Universidad Complutense de Madrid, y realizado bajo la dirección de Jose Antonio López Orozco.

Se autoriza a la Universidad Complutense a difundir y utilizar con fines académicos, no comerciales, y mencionando expresamente a sus autores, tanto la propia memoria como el código, la documentación y/o el prototipo desarrollado.

Madrid, 22 de junio de 2008.

David Romero Laorden

Resumen

El objetivo de este proyecto ha sido utilizar técnicas de Visión por Computador para desarrollar un sistema capaz de medir distancias a objetos móviles. Para ello, el trabajo se centra en dos de ellas: la visión estereoscópica o *binocular* (el enfoque clásico) y la visión *monocular* o metrología. Por tanto, se han desarrollado en primer lugar, algoritmos de identificación de objetos capaces de procesar y extraer aquellas regiones (conocidas como *blobs*) de la imagen que son de interés. Una vez, conocemos la posición del objeto en la imagen, se calcula por ambos métodos la distancia a la que se encuentra, y además se realiza una conversión a coordenadas geográficas. Por último, se ha analizado la robustez y fiabilidad del sistema en ambas variantes, la precisión de los resultados obtenidos así como el impacto de las condiciones de luz en las distintas situaciones de ensayo.

Summary

The aim of this project has been using Computer Vision techniques to develop a system capable of calculating measures distances to objects. To this end, this paper focuses on two of them: the stereoscopic or *binocular* vision (the traditional approach) and the *monocular* vision or metrology. Therefore, we have developed identification algorithms of objects capable of processing and extracting those regions (known as *blobs*) of the image that are of our interest. Once we know exactly the position of the object in the image, it is calculated by both methods the distance where it is , and we also do a conversion to geographical coordinates. Finally, it has analyzed the robustness and reliability of the system in the two cases, precision of the results and the impact of the lighting conditions in different situations.

Lista de palabras clave

Técnicas de Visión Artificial, Visión por Computador, procesamiento y análisis, blobs, visión estereoscópica, metrología, calibración cámaras, coordenadas geográficas, MIL.

Keyword list

Computer Vision techniques, Computer Vision, processing and analysis, blobs, stereo vision, metrology, camera calibration, geographical coordinates, MIL.

Agradecimientos

Quiero mostrar mi gratitud a D. Jose Antonio López Orozco y a D. Bonifacio de Andrés y Toro por haberme dado la oportunidad de elaborar este proyecto y por la ayuda que me han proporcionado. Y por supuesto, a mi compañero de trabajo, Javier por sus conocimientos y ayuda en nuestros momentos críticos y sobre todo por los buenos momentos que hemos pasado durante este año.

También quiero aprovechar esta ocasión para agradecer a D. Gonzalo Pajares Martinsanz, su inestimable ayuda, la documentación amablemente cedida y el alto grado de interés mostrado por nuestros resultados.

Índice general

Índice de figuras	v
Índice de tablas	vii
1. Introducción	1
2. Visión Artificial	7
2.1. Introducción a la Visión Artificial	7
2.1.1. Disciplinas relacionadas con la Visión Artificial	9
2.1.2. Aplicaciones	10
2.2. Sistema Estéreo	11
2.2.1. Introducción	11
2.2.2. Geometría del Sistema Estereoscópico	13
2.2.3. Obtención de la distancia	15
2.2.4. Limitaciones de un Sistema Estéreo	16
2.2.4.1. Precisión de un Sistema Estéreo	17
2.2.4.2. Oclusiones	17
2.3. Metrología	17
2.3.1. Introducción	17
2.3.2. Cálculo de medidas y distancias	18
3. Análisis y Procesamiento de Imágenes	21
3.1. Representación de las imágenes	21
3.1.1. Resolución espacial y en intensidad	22
3.2. Procesamiento de imágenes digitales	23

3.2.1.	Introducción	23
3.2.2.	Modelos de Color	25
3.2.2.1.	El modelo RGB	25
3.2.2.2.	Otros modelos de color	26
3.2.3.	Suavizado	28
3.2.4.	Binarizado	29
3.2.5.	Operaciones sobre imágenes digitales	31
3.2.6.	Operaciones morfológicas	31
3.2.6.1.	Dilatación	32
3.2.6.2.	Erosión	32
3.2.7.	Apertura	32
3.3.	Análisis de imágenes digitales	33
3.3.1.	Identificación de regiones	33
3.3.2.	Segmentación de regiones	33
4.	Coordenadas Geográficas y GPS	35
4.1.	Fundamentos matemáticos	35
4.1.1.	Introducción	35
4.1.2.	Meridianos	37
4.1.3.	Paralelos	37
4.1.4.	Longitud	38
4.1.5.	Latitud	38
4.2.	Global Positioning System	39
4.2.1.	GPS MIDG II	42
4.3.	Conversión a longitud y latitud	43
5.	Implementación	47
5.1.	<i>Matrox Imaging Library</i>	47
5.2.	Cámaras y lentes	48
5.3.	Estructura general	49
6.	Resultados experimentales	51
6.1.	Inicio y preparación de las pruebas	51
6.2.	Calidad de los resultados	52

6.3. Repetitividad	55
6.4. Error relacionado con la distancia	58
6.5. Precisión	60
7. Conclusiones y trabajo futuro	63
7.1. Conclusiones	63
7.2. Trabajo futuro	65
A. Instalación y uso	67
A.1. Instalación	67
A.2. Uso	68
B. Matrox Imaging Library	75
C. MIDG II Display	79
C.1. Horizonte Artificial	80
C.2. Ventana de Estado	80
C.3. Ventana de Datos del Sensor de Navegación	80
C.4. Ventana de la Posición de Navegación y Velocidad	81
C.5. Ventana de Estimación de la Precisión de Navegación	82
C.6. Ventana de posición y velocidad dada por el GPS	82
C.7. Ventana de información de los satélites del GPS	83
Bibliografía	85

Índice de figuras

2.1. Distintas aplicaciones de la Visión Artificial	10
2.2. Sistema de visión estereoscópica biológico, con dos objetos . .	11
2.3. Superposición de las imágenes de ambas cámaras	12
2.4. Geometría de un sistema <i>binocular</i> con ejes ópticos paralelos .	14
2.5. Geometría de dos cámaras en estéreo desde una perspectiva superior	15
2.6. Sistema de ecuaciones hallado para el cálculo de la distancia .	16
2.7. Modelo del plano la cámara en un sistema <i>monocular</i>	19
3.1. Representación del proceso de captura de una escena	22
3.2. Representación matemática de una imagen digital	23
3.3. Tetraedro de color que representa el modelo RGB	26
3.4. Algunas de las propiedades de un <i>blob</i>	34
4.1. Sistema de referencia de tres dimensiones	36
4.2. Representación de un meridiano en la superficie terrestre . . .	37
4.3. Representación de un paralelo en la superficie terrestre	38
4.4. Visualización gráfica de la longitud	39
4.5. Visualización gráfica de la latitud	40
4.6. Ejes del sistema de conversión a coordenadas geográficas . . .	43
4.7. Representación gráfica del problema de las cámaras no alineadas	45
6.1. Gráfica que muestra la media de resultados para los tres puntos	55
6.2. Tablas que muestran la conversión de coordenadas geográficas y UTM datum WGS84 a datum ED50 (y viceversa)	56

6.3. Repetitividad en los datos obtenidos para 100 metros	57
6.4. Repetitividad en los datos obtenidos para 150 metros	57
6.5. Repetitividad en los datos obtenidos para 200 metros	58
6.6. Repetitividad en los datos obtenidos para 20 metros	59
6.7. Repetitividad en los datos obtenidos para 30 metros	59
6.8. Efecto de la distancia con respecto a los resultados obtenidos .	60
6.9. Precisión en el desplazamiento lateral hacia la derecha del ob- jeto en la escena.	62
6.10. Precisión en el desplazamiento lateral hacia la izquierda del objeto en la escena.	62
A.1. Ventana principal de Halcón.	68
A.2. Diálogo que permite modificar los parámetros de Halcón	70
A.3. Ventana que muestra el proceso de introducción de la distancia en la alineación	71
A.4. Diálogo que muestra la utilización de las intensidades de color	72
A.5. Diálogo que muestra los resultados obtenidos	73
A.6. Calibración de las cámaras, selección de puntos	74
A.7. Calibración de las cámaras, introducción de coordenadas. . . .	74
B.1. Paquete de librerías <i>Matrox Imaging Library</i>	75
B.2. Arquitectura software de MIL	78
C.1. Ventana principal MIDG II	79
C.2. Cuadro del horizonte artificial	80
C.3. Ventana de estado	81
C.4. Ventana de datos del sensor de navegación	81
C.5. Ventana de la posición de navegación	81
C.6. Ventana de estimación de la precisión	82
C.7. Ventana de posición y velocidad	82
C.8. Ventana de información de los satélites	83

Índice de tablas

6.1. Distancias entre los puntos elegidos en Cantarranas con GPS .	54
6.2. Distancias entre los puntos elegidos en Cantarranas con Halcón	54
6.3. Precisión del sistema en distancias cortas	61

Capítulo 1

Introducción

Este documento tiene el objetivo de dar una visión a distintos niveles sobre el proyecto de *Fin de Máster en Informática para la Industria*, pudiéndose encontrar en él información sobre los distintos aspectos que lo forman. En este capítulo se pretende introducir brevemente cuáles son nuestros objetivos, las herramientas que hemos utilizado y la estructura del resto de la memoria.

Uno de los principales objetivos de la Visión Artificial (también conocida como Visión por Computador del inglés *Computer Vision*) ha sido y es, la toma de medidas de una escena y la reconstrucción de su modelo tridimensional. El problema del uso de técnicas de visión para obtener y medir distancias en el mundo ha centrado la atención y recibido gran variedad de aplicaciones durante estos años [Cri99], [AC99], [GWW04], como medición de interiores, reconstrucciones desde imágenes, mediciones forenses e investigaciones en accidentes de tráfico.

La aproximación tradicional al cálculo de medidas ha sido tomar manualmente distancias usando cintas métricas, reglas o incluso mediante dispositivos especiales como sistemas de ultrasonidos o láseres. Estos procedimientos consumen una gran cantidad de tiempo, son proclives a errores e invasivos. Con los métodos basados en Visión Artificial, lo único imprescindible que debe hacerse es capturar varias fotografías o imágenes de la escena para después

tomar todas las medidas que se necesiten con mayor precisión, flexibilidad y eficiencia. Existen, por tanto un gran número de ventajas para este tipo de procedimientos [Cri99]. La adquisición de las imágenes es rápida, simple y mínimamente invasiva, ya que el único aspecto involucrado es la cámara que captura las imágenes del entorno.

El presente trabajo tiene como base fundamental la Visión Artificial, que es un subcampo de la inteligencia artificial y que como se ha comentado anteriormente, tiene el propósito de capturar la información visual del entorno físico para extraer características relevantes visuales, utilizando procedimientos automáticos. Las imágenes son adquiridas mediante una o más cámaras y tras la interpretación del mismo por los programas adecuados, la imagen se representa desde el punto de vista de su tratamiento computacional como una matriz numérica de dimensión $M \times N$, es decir, con M filas y N columnas. El contenido de esta matriz se refiere a valores enteros situados en localizaciones espaciales (x, y) cada uno de los cuales se denomina píxel (*"picture element"*).

Las técnicas de Visión Artificial tienen muy diversos campos de aplicación; se citarán algunos de los más importantes en la actualidad a modo de ejemplo (y sin pretender en modo alguno cubrir todas las posibilidades):

- En control de calidad se usan para inspección de productos, identificación de piezas, entre otras.
- En automoción es utilizada en gran parte para la medición, calibres, acabados y reconocimiento de matrículas.
- En envasado se aplican a la comprobación en láminas hojalata (vetas, agujeros, impurezas), detección de rayas, defectos, goma, contaje de producción y acabado de litografía. También para control de llenado, empaquetado y etiquetado, así como verificación de la fecha de caducidad e identificación de código de barras.
- En electrónica se utilizan para la comprobación de soldadura y de circuitos integrados (serigrafía, pines, golpes . . .), la medición de precisión

y conectores.

- En farmacia y equipos médicos se usan para comprobación de pipetas, presencia en blisters, marcas de seguridad ultravioleta, y medición de precisión (en herramientas médicas).
- En robótica se aplican a control de soldaduras, guiado y asistencia de robots.

Es ampliamente conocido en literatura científica, que los métodos de Visión Artificial pueden ser clasificados en términos generales en dos categorías. Concretamente, este trabajo se centra en esas dos técnicas de visión. En el campo que nos ocupa, existen distintos mecanismos de los que se dispone para localizar un objeto mediante métodos de visión por computador. El primer método se da cuando la localización se realiza mediante el uso conjunto de dos cámaras, técnica conocida como *visión estereoscópica* o *binocular*, mientras que el segundo de ellos engloba solamente el uso de una cámara, técnica conocida como *metrología*.

Por tanto, el objetivo de este proyecto ha sido utilizar estos métodos de visión para obtener distancias a objetos móviles (identificados por algún color distintivo) para realizar un seguimiento y control de manera automática.

La *visión estereoscópica* es un procedimiento para la obtención de la forma de los objetos en una escena que se determina a través de la distancia de los objetos en relación con un sistema de referencia [GW05], [DL98], [RH00], [IR96]. En estos sistemas el desplazamiento relativo de las cámaras (término conocido como *disparidad*) permite obtener la profundidad de los objetos o tercera dimensión mediante un simple proceso de triangulación a partir de las dos imágenes generadas por el mismo objeto de la escena 3D en cada cámara. Los requisitos en este procedimiento son tener un sistema de cámaras perfectamente alineado así como una buena calibración de las cámaras que determinen sus características reales. También se ha utilizado visión estéreo

en trabajos completamente distintos, como el uso de mapas *fuzzy* cognitivos y enfriamiento simulado en la correspondencia de características entre las imágenes estereoscópicas [GPM06], [GPM].

La *metrología* por otra parte, tiene como principal ventaja la no necesidad de realizar un proceso de calibración de la cámara, puesto que únicamente se necesitan cuatro puntos de correspondencia entre la imagen tomada y la imagen real para poder llevar a cabo el cálculo de la distancia. El procedimiento se concentra básicamente en la aplicación de operaciones matemáticas en perspectiva, de manera que conocidos los puntos antes mencionados se puedan medir distancias entre dos puntos cualesquiera de la imagen, y por consiguiente, medir longitudes y obtener distancias.

Conocido el sistema de referencia (la posición geográfica de las cámaras) y una vez calculada la distancia z al objeto por cualquiera de los dos métodos anteriores, los resultados son expresados en x e y y además, se realiza también una conversión a coordenadas geográficas (latitud y longitud).

Para el desarrollo de este sistema se han utilizado un conjunto de librerías de procesamiento de imágenes conocidas como MIL (*Matrox Imaging Library*) en su versión 8.0 [Mat]. Las cámaras utilizadas son de la marca *Basler*, concretamente el modelo *Scout scA1400-17fc* con un sensor *CCD* Sony que soporta gran variedad de resoluciones (hasta 1390 x 1038 píxeles) y con una interfaz *IEEE 1394b* [Bas].

Este trabajo se encuentra enmarcado dentro del proyecto “Plataforma de planificación, simulación y control para colaboración de múltiples vehículos autónomos marinos y aéreos” cuyos objetivos son: (1) desarrollar una plataforma software para facilitar a los diseñadores desarrollar, analizar, simular y testear algoritmos de coordinación y cooperación de vehículos autónomos marinos y aéreos, dentro de la filosofía MDA, (2) analizar y desarrollar algoritmos de “Asignación de Tareas y Planificación de Caminos” para un conjunto de UAVs y de UMVs, (3) desarrollo de una librería de vehículos UAVs y UMVs.

Cada UV llevará un controlador “básico” de *tracking*, (4) desarrollar escenarios de coordinación y colaboración de UAVs y UMVs para tareas de vigilancia, búsqueda y rescate y (5) estudiar algoritmos de control de formaciones de UV en tareas de coordinación. La entidad financiadora es PDI2006-15661-C02-01 con una duración que comprende las fechas desde 1\10\2006 hasta 30\09\2009.

Por último sólo queda comentar la estructura del trabajo. En el segundo capítulo se profundiza en la Visión Artificial, dando una introducción general de la misma y analizando en detalle las dos técnicas de cálculo de distancias que hemos comentado.

En el tercer capítulo se aborda la temática del procesamiento digital de imágenes así como su posterior análisis, comentando las etapas seguidas para preparar y mejorar la imagen para su posterior tratamiento. En el cuarto capítulo se introducen algunos aspectos de la implementación de manera global. En el quinto capítulo se hace un breve repaso de las coordenadas geográficas y la utilización de un GPS; y en el sexto capítulo se detallan los resultados obtenidos de manera experimental teniendo en cuenta las pruebas realizadas.

Entre las conclusiones detalladas en el séptimo, y último, capítulo, cabe destacar que nuestro sistema cumple con las expectativas y que el cálculo de distancias obtenidas para las dos técnicas de visión es realmente bueno, y lo que es igualmente importante, la calidad de los resultados no tienen nada que envidiar a obtenidos por otros sistemas de medición.

Capítulo 2

Visión Artificial

En este capítulo se explican los distintos mecanismos de los que se dispone para localizar un objeto mediante métodos de visión por computador. Concretamente nos centraremos en dos de ellos. El primer método se da cuando la localización se realiza mediante el uso conjunto de dos cámaras, esto es, se estará realizando un método de visión estereoscópica, mientras que el segundo de ellos engloba solamente el uso de una cámara para desarrollar un método de visión monocámara, sistema conocido como metrología.

La estructura del capítulo es la siguiente. En la sección 2.1 se hace una pequeña introducción a la Visión Artificial, comentando las disciplinas que engloba y algunas de sus aplicaciones. En la sección 2.2 se abordan los fundamentos de un sistema estéreo mientras que en la sección 2.3 se hace lo propio con el sistema monocámara.

2.1. Introducción a la Visión Artificial

Uno de los sentidos más importantes de los seres humanos es la visión. Ésta es empleada para obtener la información visual del entorno físico. Según Aristóteles, *‘Visión es saber que hay y donde mediante la vista’*. De hecho, se calcula que más de 70 % de las tareas del cerebro son empleadas en el análisis de la información visual. El refrán popular de *‘Una imagen vale más que mil’*

palabras” tiene mucho que ver con los aspectos cognitivos de la especie humana. Casi todas las disciplinas científicas emplean utillajes gráficos para transmitir conocimiento. Por ejemplo, en Ingeniería Electrónica se emplean esquemas de circuitos, a modo gráfico, para describirlos. Se podría hacer mediante texto, pero para la especie humana resulta mucho más eficiente procesar imágenes que procesar texto. La visión humana es el sentido más desarrollado y el que menos se conoce debido a su gran complejidad. Es una actividad inconsciente y difícil de saber cómo se produce. De hecho, hoy en día, se carece de una teoría que explique cómo los humanos perciben el exterior a través de la vista.

En el año 1826 el químico francés Niepce (1765-1833) llevó a cabo la primera fotografía, colocando una superficie fotosensible dentro de una cámara oscura para fijar la imagen. Posteriormente, en 1838 el químico francés Daguerre (1787-1851) hizo el primer proceso fotográfico práctico. Daguerre utilizó una placa fotográfica que era revelada con vapor de mercurio y fijada con trisulfato de sodio.

Desde que se inventó la fotografía se ha intentado extraer características físicas de las imágenes. La Fotogrametría dió sus primeros pasos desde imágenes capturadas en globos. La Astronomía avanzó enormemente con el análisis de imágenes recibidas por los telescopios. El análisis de radiografías transformó la Medicina. Se podrían citar muchos más ejemplos que durante décadas han transformado la percepción de la Ciencia con el procesamiento de las imágenes, alguna veces por separado y otras de forma multidisciplinar.

Sin embargo, el momento histórico que hace que estas técnicas confluyan y den un cuerpo de conocimiento propio, surge en la década de los 80. La revolución de la Electrónica, con las cámaras de vídeo CCD y los microprocesadores, junto con la evolución de las Ciencias de la Computación hace que sea factible la Visión Artificial.

Por tanto, la Visión Artificial o también llamada Visión por Computador, pretende capturar la información visual del entorno físico para extraer car-

acterísticas relevantes visuales, utilizando procedimientos automáticos. Según David Marr [Mar82], ‘*Visión es un proceso que produce a partir de imágenes del mundo exterior una descripción útil para el observador y no tiene información irrelevante*’.

Las imágenes para ser procesadas en el computador han sido adquiridas a través de una o más cámaras y puestas en su memoria. Esta señal es de carácter bidimensional y emplea variables discretas. Los elementos de la matriz se llaman píxeles. El acceso a esta elemental información se hace indicando la fila y la columna que ocupa. El origen de coordenadas de la imagen se encuentra en la esquina superior izquierda. El eje horizontal corresponde con las columnas y el eje vertical con las filas. Se emplearán índices enteros para posicionar el píxel. Se denotará el valor del píxel a través de una función, del tipo $f(x, y)$, siendo x el índice de la fila e y de la columna. Si la imagen es acromática, sólo se presenta la luminancia, esto es, los niveles de grises. La función $f(x, y)$ retornará el nivel de gris del píxel mencionado. En caso de que la imagen sea en color, $f(x, y)$ devolverá un vector. Normalmente, suele expresarse como una proyección del color sobre el sistema RGB (*Red-Green-Blue*), como se ha realizado en el presente trabajo.

2.1.1. Disciplinas relacionadas con la Visión Artificial

Generalmente, y como sucede en nuestro caso, un proyecto de Visión Artificial es de tipo multidisciplinar. La ejecución de las distintas etapas, mencionadas en el apartado anterior, requiere del manejo de las siguientes técnicas:

- Fotografía y Óptica: Crear el ambiente de iluminación adecuada en la adquisición de las imágenes, muchas veces requiere del uso de técnicas profesionales de fotografía y vídeo. La selección de la óptica y de la cámara, el uso de filtros y polarizadores, las técnicas de iluminación con pantallas y la elección de los tipos de focos son algunas habilidades que se pueden mencionar.
- Procesamiento Digital de las Imágenes (*Image Processing*): Hace referencia a los algoritmos de computación que convierte la imagen digital

adquirida en otra de mayor relevancia. Es muy difuso la separación entre el procesamiento de imágenes y la Visión Artificial.

- Reconocimiento de Patrones (*Pattern Recognition*): disciplina, dentro de la Inteligencia Artificial, dedicada a la clasificación de las señales y a la búsqueda de patrones existentes dentro de éstas. Se encuentran incluidas las técnicas de clasificadores estadísticos, Redes Neuronales, Sistemas Expertos, Lógica Borrosa, ...
- Computación Gráfica (*Computer Graphics*): Presenta el problema inverso de la Visión Artificial. Si en Visión se desea extraer las características físicas de las imágenes, la Computación Gráfica se dedica a la presentación visual de los modelos geométricos. Cada vez más, la Visión Artificial emplea la Computación Gráfica para representar las conclusiones extraídas del análisis de las imágenes adquiridas.

2.1.2. Aplicaciones

El número de aplicaciones relacionadas con la Visión Artificial aumenta cada día. En la tabla 2.1 se muestran algunos de los campos donde es empleada esta disciplina.

Control de calidad	Inspección de productos (papel, aluminio, acero...)	Astronomía	Exploración del Espacio
	Identificación de piezas	Reconocimiento de caracteres	Control de cheques, inspección de textos, ...
	Etiquetados (fechas de caducidad,...)	Control de tráfico	Matriculas de coches
	Inspección de circuitos impresos		Tráfico viario
	Control de calidad de los alimentos (naranjas,...)	Meteorología	Predicción del tiempo
Robótica	Control de soldaduras	Agricultura	Interpretación de fotografías aéreas
	Guiado de robots (vehículos no tripulador)		Control de plantaciones
Biomédicas	Análisis de imágenes de microscopía (virus, células, proteínas)	Militares	Seguimiento de objetivos
	Resonancias magnéticas, tomografías, genoma humano		Vigilancia por satélites

FIGURA 2.1: Distintas aplicaciones de la Visión Artificial.

2.2. Sistema Estéreo

La visión estereoscópica constituye uno de los procedimientos clásicos para la obtención de la tercera dimensión de la escena. Esta tercera dimensión en realidad se convierte en la obtención de distancias desde el sistema estéreo a los objetos. Por ello, ésta sección explica los fundamentos físicos de la visión estereoscópica que son parte fundamental del presente trabajo.

2.2.1. Introducción

La visión estereoscópica es un procedimiento para la obtención de la forma de los objetos en una escena que se determina a través de la distancia de los objetos en relación con un sistema de referencia. Toma como referencia el modelo estereoscópico biológico. En estos sistemas el desplazamiento relativo de los ojos permite obtener la profundidad de los objetos o tercera dimensión mediante un simple proceso de triangulación a partir de las dos imágenes generadas por el mismo objeto de la escena 3D en cada ojo.

El hecho de que los ojos estén desplazados entre sí hace que las imágenes de los objetos en sendos ojos se muestren desplazadas según la distancia de los objetos a los ojos. Observemos la figura 2.2:

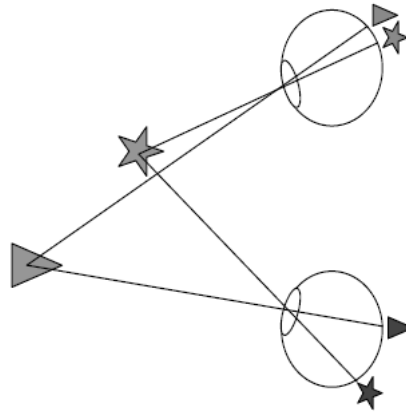


FIGURA 2.2: Sistema de visión estereoscópica biológico, con dos objetos en la escena 3D a distintos niveles de profundidad. Fuente: [MdlC08]

Se muestra un sistema estereoscópico biológico observando dos objetos (estrella y triángulo) a distintos niveles de profundidad, obteniéndose en las correspondientes retinas de los ojos las imágenes. Si solapamos ambos ojos obtenemos una nueva imagen, mostrada en la figura 2.3 en la que se observa que la separación relativa entre las imágenes de los dos triángulos es menor que la separación relativa entre las imágenes de las estrellas. Este fenómeno se explica por el hecho de que la estrella en la escena 3D se encuentra más próxima a los ojos que el triángulo. Estas separaciones relativas entre las imágenes de un mismo objeto de la escena 3D es lo que denominamos *disparidad* y que se comentará más adelante.



FIGURA 2.3: Superposición de las imágenes de ambos ojos. Fuente: [MdlC08]

En visión estereoscópica artificial se utilizan dos o más cámaras separadas entre sí una cierta distancia relativa con las que se obtienen las correspondientes imágenes. El procedimiento consiste en captar dos o más imágenes de una misma escena, cada imagen es capturada desde una posición de las cámaras ligeramente diferente a las anteriores y como las imágenes se presentan ligeramente desplazadas entre sí podremos obtener la distancia a la que se encuentra un objeto como veremos en la subsección 2.2.3.

Las cámaras tienen sus ejes ópticos perpendiculares a la línea base y sus líneas de exploración o epipolares paralelas a la línea base. Las líneas epipolares son líneas uniendo las imágenes izquierda y derecha de un mismo punto en la escena. En este sistema de ejes ópticos paralelos el desplazamiento entre los ejes ópticos de las dos cámaras es horizontal, por lo que las imágenes de un punto determinado de la escena captado por ambas cámaras difiere solamente en la componente horizontal.

Siguiendo la terminología de Barnard y Fischler (1982), el proceso de la visión estereoscópica consta de los siguientes pasos: (1) adquisición de imágenes, (2) geometría del sistema, (3) extracción de características, (4) correspondencia de características y (5) obtención de la distancia.

El proceso de adquisición de imágenes conlleva capturar más de una imagen procedentes de la misma escena, si bien con el mencionado desplazamiento relativo entre las cámaras. Sobre las imágenes estereoscópicas capturadas se llevan a cabo los siguientes pasos requeridos en el proceso, por lo que se requiere que tanto la cuantización espacial como temporal debe ser la misma en todas las imágenes estereoscópicas.

En el resto de esta parte del capítulo, se analizarán cada uno de los pasos restantes, tratando la geometría del sistema y la obtención de la distancia en las subsecciones 2.2.2 y 2.2.3. La correspondencia de características es el paso más complejo y difícil de todo el proceso, aunque en el presente trabajo y debido a la naturaleza del proyecto esto no será así. Por último en la subsección 2.2.4 se expondrán algunas de las limitaciones presentes en sistemas de éste tipo

2.2.2. Geometría del Sistema Estereoscópico

Un sistema convencional está caracterizado por un par de cámaras con sus ejes ópticos mutuamente paralelos y separados por una distancia horizontal que denominamos *línea base*, y que en la figura 2.4 es el parámetro b . Las cámaras tienen sus ejes ópticos perpendiculares a la línea base y sus líneas de exploración o epipolares paralelas a la línea base. Las líneas epipolares son líneas uniendo las imágenes izquierda y derecha de un mismo punto en la escena.

Es importante destacar que en este sistema de ejes ópticos paralelos el desplazamiento entre los centros ópticos de las dos cámaras es horizontal, esto se traduce en el hecho de que las imágenes de un punto determinado de la escena

captado por ambas cámaras difiere solamente en la componente horizontal.

Supongamos el siguiente esquema para la geometría de un par de cámaras en estéreo con ejes ópticos paralelos:

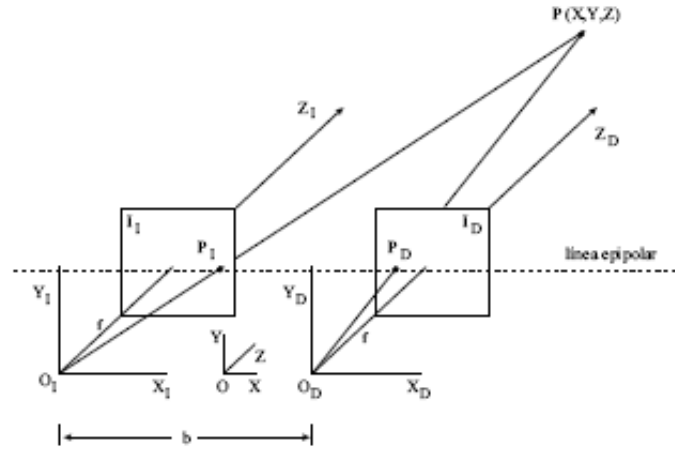


FIGURA 2.4: Geometría de un par de cámaras con ejes ópticos paralelos. Fuente: [MdlC08]

Así, el origen del sistema de coordenadas de referencia o del mundo está en O , siendo la longitud focal efectiva de cada cámara f , y la línea base b como ya hemos definido anteriormente. Los ejes de coordenadas del mundo X , Y , y Z se sitúan entre los ejes de ambas cámaras. Supongamos que $P_I(x_I, y_I, z_I)$ y $P_D(x_D, y_D, z_D)$ son las proyecciones del punto $P(X, Y, Z)$ de la escena 3D.

Los rayos de proyección POI y POD definen el plano de proyección del punto de la escena 3D llamado plano epipolar. Para un punto dado P_I en la imagen izquierda, su punto de correspondencia P_D en la imagen derecha debe caer en la línea de intersección del plano epipolar y el plano de la imagen, es decir, sobre la línea epipolar.

La línea epipolar en la imagen derecha correspondiente a un punto P_I en la imagen izquierda define el espacio de búsqueda dentro del cual se situará el correspondiente punto de emparejamiento P_D . Como consecuencia de la geometría de la imagen se obtiene la denominada restricción epipolar, que ayuda

a limitar el espacio de búsqueda de correspondencias, de manera que en el sistema de ejes paralelos convencional todos los planos epipolares originan líneas horizontales al cortarse con los planos de las imágenes.

En un sistema con la geometría anterior se obtiene un valor de disparidad d , para cada par de puntos emparejados $P_I(x_I, y_I)$ y $P_D(x_D, y_D)$ dado por:

$$d = x_I - x_D$$

2.2.3. Obtención de la distancia

Vista con detalle en la sección anterior la geometría de un sistema estéreo, el problema de obtener la distancia se ha convertido en una tarea bastante sencilla. Si consideramos una relación geométrica de semejanza de triángulos como la mostrada en la figura 2.5 vemos que las coordenadas del punto de la escena $P(X, Y, Z)$ se pueden deducirse fácilmente.

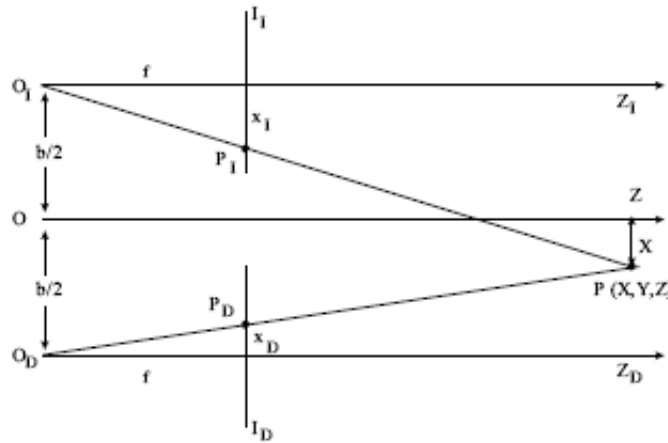


FIGURA 2.5: Geometría de dos cámaras en estéreo con ejes ópticos paralelos desde una perspectiva superior. Fuente: [MdIC08]

Obteniendo los resultados dados por la ecuación 2.6. Cabe prestar una mención especial a la distancia focal f . En todo momento se ha considerado que es idéntica en ambas cámaras lo que permite despejar la Z sin ninguna dificultad. Se deduce a partir de esta ecuación que cuando se utiliza esta ge-

ometría, la profundidad Z , es inversamente proporcional a la disparidad de la imagen y para una profundidad dada, a mayor b mayor d , lo que sugiere que la línea base puede incrementarse para mejorar la exactitud de la profundidad medida, pero ello lleva consigo el hecho de que ahora ambas imágenes tienen menos características comunes, es decir menos bordes o regiones procedentes de los objetos de la escena, debido a las desapariciones y oclusiones de las imágenes de dichos objetos.

$$\left. \begin{array}{l} O_I: \frac{b+X}{Z} = \frac{x_I}{f} \\ O_D: -\frac{b-X}{Z} = \frac{x_D}{f} \end{array} \right\} \Rightarrow \left. \begin{array}{l} x_I = \frac{f}{Z} \left(X + \frac{b}{2} \right) \\ x_D = \frac{f}{Z} \left(X - \frac{b}{2} \right) \end{array} \right\} \Rightarrow d = x_I - x_D = \frac{fb}{Z} \Rightarrow Z = \frac{fb}{d}$$

FIGURA 2.6: Sistema de ecuaciones hallado para el cálculo de la distancia.

Como contraposición a los sistemas con dos cámaras estáticas, el sistema estereoscópico propuesto en [Mor77] supuso en su momento una interesante innovación. En un robot móvil incorpora una única cámara de forma que en cada pausa del robot dicha cámara se desplaza de izquierda a derecha tomando nueve imágenes en intervalos controlados de tiempo. El desplazamiento de la cámara se realiza de forma paralela al eje horizontal del sistema de coordenadas. Con los pares de imágenes capturados realiza una estima de la posición de la característica de la imagen. Dicha estima se modela como una curva Gaussiana cuya media es igual a la distancia estimada y con desviación estándar inversamente proporcional a la longitud de la línea base.

2.2.4. Limitaciones de un Sistema Estéreo

A la hora de diseñar un sistema estereoscópico han de tenerse en cuenta una serie de consideraciones que afectan al mismo. Existen dos limitaciones de cierta importancia, a saber: (1) de naturaleza física y (2) impuesta por la geometría del sistema. En el primer caso hablaríamos de precisión y en el segundo de oclusiones y ambigüedades.

2.2.4.1. Precisión de un Sistema Estéreo

Un aspecto importante que ha de tenerse en cuenta a la hora de diseñar un sistema estereoscópico es el relativo a las limitaciones impuestas por los dispositivos físicos y que afectan tanto a la precisión como a la operatividad del mismo. Parte de estas limitaciones vienen impuestas por la disposición limitada de las celdas del sensor CCD.

2.2.4.2. Oclusiones

La oclusión es un fenómeno dependiente del punto de observación y surge porque ciertas características que son visibles desde un punto de vista dejan de serlo desde otro. Esto hace que las posiciones espaciales en la escena 3D de las características afectadas por la oclusión no se puedan determinar mediante el procedimiento de triangulación en el que se basa la visión estereoscópica.

2.3. Metrología

La visión *monocular* (o *metrología*) es otro procedimiento alternativo válido para la obtención de la tercera dimensión de la escena. En esta escena, al igual que cuando vemos un cuadro o una fotografía, existen una serie de factores que, de una forma intuitiva, aportan información sobre aspectos tridimensionales como la distancia o profundidad. Por ello, en ésta sección se explican los fundamentos generales de la metrología que son, al igual que la visión *binocular*, una parte importante del presente trabajo.

2.3.1. Introducción

La metrología es la ciencia de la medida. Tiene por objetivo el estudio de los sistemas de medida en cualquier campo de la ciencia. Tiene dos características muy importantes que son el resultado de la medición y la incertidumbre de medida. Los físicos y la industria utilizan una gran variedad de instrumentos para llevar a cabo sus mediciones. Desde objetos sencillos como reglas y cronómetros, hasta potentes microscopios, medidores de láser e incluso acel-

eradores de partículas.

En el caso que nos ocupa, la visión monocular trabaja directamente con una única imagen sin ningún tipo de calibración [AC99], [GWW05], [vdH98]. Es bien conocido en este campo, que una sola imagen no puede dar la suficiente información para una completa reconstrucción 3D. Sin embargo, el cálculo de medidas y distancias pueden ser inferidos directamente desde una imagen a partir del conocimiento de algunas constantes geométricas de la escena como el paralelismo de líneas, planos y el efecto de la perspectiva.

En algunos trabajos previos como [ACZ99], se relata el uso de cuatro puntos específicos de control en el plano, para obtener la estimación de la distancia en homografía. Además, en [GWW04] este mismo problema ha sido discutido, estudiando otras alternativas y métodos para calcular medidas en la escena. En [MW01], [TK98] y [PG00] otras técnicas dedicadas a la reconstrucción de objetos para tomar medidas han sido investigadas en las comunidades de Visión Artificial y Fotogrametría. Estas técnicas se basan en las características de los objetos a ser reconstruidos, como esquinas, bordes, perpendicularidad o paralelismo.

En este trabajo, la aplicación de éste método guarda una estrecha relación con el uso de las librerías de procesamiento de imágenes en las que nos hemos apoyado para realizar el proyecto. Es por ello, que únicamente se dará una visión general de esta técnica de manera menos teórica.

2.3.2. Cálculo de medidas y distancias

La visión monocámara es un sistema que nos permite realizar el posicionamiento de un objeto con la ayuda de las imágenes provenientes de una sola cámara. En la figura 2.7 podemos apreciar el modelo de la cámara en el plano.

Al contrario que en la visión estereoscópica, en este método solo disponemos

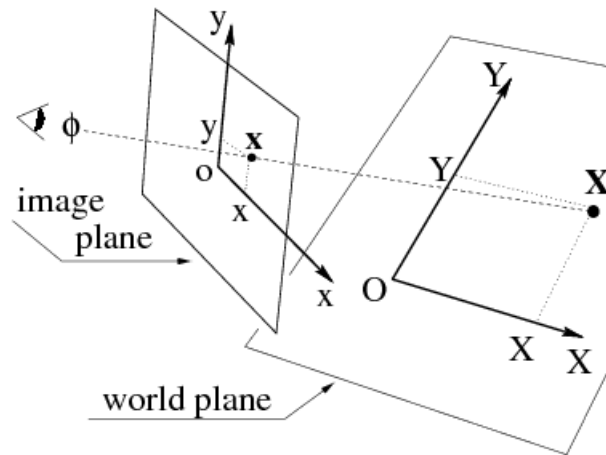


FIGURA 2.7: Modelo del plano la cámara: un punto X del plano es dibujado en la imagen como x . Las coordenadas euclídeas $X - Y$ y $x - y$ son utilizadas para los planos del mundo y de la imagen, respectivamente. El centro de la cámara es ϕ .

de una imagen, con lo que no lo podemos comparar con nada para que nos de una disparidad o una distancia relativa, por lo que tendremos que buscar un método para referenciar un punto del mundo real con su equivalente en píxeles.

Esta tarea no es tan fácil como pudiera parecer, ya que al capturar una escena de la realidad, que obviamente está en tres dimensiones, la estamos pasando a dos dimensiones, por lo que perdemos la profundidad. No tiene la misma longitud algo que esté cerca del objetivo que algo que esté lejos. Pero esta pérdida de profundidad se puede suplir si tenemos en cuenta la perspectiva. Es decir, ir variando la distancia del mundo real a la que equivale un píxel de la imagen, en función de lo cerca o lejos que esté del objetivo.

Por tanto, considerando el hecho de que conocemos exactamente la posición de cuatro puntos de la realidad, respecto al eje de coordenadas de la cámara, es posible obtener distancias a cualquier punto de la imagen, y también medidas relativas entre dos puntos cualesquiera de la misma.

Capítulo 3

Análisis y Procesamiento de Imágenes

Una vez introducidos los dos métodos de Visión Artificial utilizados en este trabajo, en este capítulo se explican los conceptos fundamentales relacionados con la representación de las imágenes digitales tras su captura tanto en su disposición espacial como en su contenido por cualquiera de los dispositivos de adquisición de imágenes. Además, se abordan también las etapas de procesamiento y análisis, profundizando en los detalles más relevantes de cada una de ellas.

La primera parte del capítulo (sección 3.2) se centrará en el procesamiento digital de imágenes, es decir, el conjunto de técnicas que se aplican a las imágenes digitales con el objetivo de mejorar su calidad y facilitar la búsqueda de información. Por otro lado, en la segunda parte (sección 3.3) abordará el problema del análisis de las regiones interesantes de las imágenes para el caso que nos ocupa.

3.1. Representación de las imágenes

Las imágenes digitales son capturadas mediante los dispositivos apropiados para su almacenamiento en una computadora. La figura 3.1 representa un

proceso de captura de una escena tridimensional.

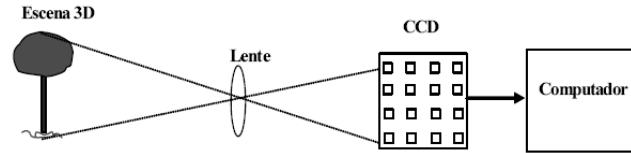


FIGURA 3.1: Captura de una imagen 3D por un dispositivo CCD. Fuente: [MdlC08]

El proceso termina con el almacenamiento de la imagen en la computadora a través de algún formato característico de imágenes (BMP, JPEG). Tras la interpretación del mismo por los programas adecuados, la imagen se representa desde el punto de vista de su tratamiento computacional como una matriz numérica de dimensión $M \times N$, es decir, con M filas y N columnas. El contenido de esta matriz se refiere a valores enteros situados en localizaciones espaciales (x, y) cada uno de los cuales se denomina píxel o pel (“*picture element*”) y cuyos valores posibles van del 0 (negro) al 255 (blanco).

Sobre cada matriz se establece un sistema de coordenadas con origen normalmente en la esquina superior izquierda y con los ejes x e y tales que su orientación es positiva hacia la derecha en el caso del eje x y hacia abajo en el caso del eje y . De modo esquemático, este hecho se puede observar en la figura 3.2. Cuando la imagen es de color, existen tres valores de intensidad asociados a cada localización espacial que corresponden a los canales espectrales Rojo, Verde y Azul, propios de la nomenclatura RGB (Red, Green, Blue). Es decir existen tres imágenes con la misma representación antes mencionada.

3.1.1. Resolución espacial y en intensidad

La resolución espacial hace referencia al hecho de que una misma escena puede representarse con más o menos filas o columnas. Cuanto mayor sea el número de filas y columnas mayor es la resolución espacial.

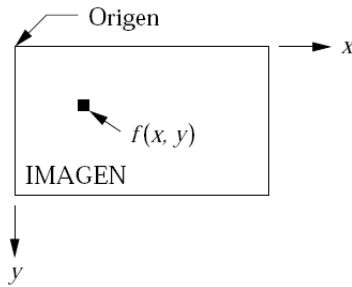


FIGURA 3.2: Representación de una imagen digital. Fuente: [MdlC08]

Se habla también de resolución en amplitud, que hace referencia al rango de valores con los que se representa la escena, este rango se conoce como número de niveles de intensidad.

3.2. Procesamiento de imágenes digitales

El término procesamiento digital de imágenes versa sobre la manipulación y análisis de imágenes por medio de una computadora. De aquí en adelante se usará el término procesado de imágenes refiriéndose a procesamiento digital de imágenes.

El procesamiento de imagen puede considerarse como un tipo especial del procesamiento digital en dos dimensiones, el cual se usa para revelar información sobre imágenes y que involucra hardware, software y soporte teórico.

En la anterior sección 3.1 se ha dado la definición de imagen. En esta parte del capítulo se describen las áreas en que se divide el procesamiento de imágenes y se da una amplia visión de cada una de ellas.

3.2.1. Introducción

El origen del procesamiento digital de imágenes puede ser rastreado hasta principios de la década de los sesenta. Fue en este tiempo cuando la *NASA* daba seguimiento al programa de ciencia lunar, en un intento de caracteri-

zar la superficie de la luna para apoyar el posterior programa *Apolo*. Una vez tomadas las grabaciones de la superficie lunar, y tras varios intentos fallidos de enviarlas a la Tierra, el proyecto Ranger 7 logró enviar dichas grabaciones a la tierra y, una vez aquí, fueron convertidas de su forma análogica a digital. Una vez hecho esto, fue necesario evaluar las imágenes y eliminar distorsiones geométricas y de respuesta. Este proceso dió inicio al empleo de computadoras para procesamiento de imágenes.

A pesar de que en sus inicios el procesamiento de imágenes estaba restringido a campos de investigación de ciencias de la computación, matemáticas y astronomía, hoy en día se ha convertido en una disciplina en sí misma. El objetivo es visualizar o evaluar estadísticamente algunos aspectos de la imagen que no se encuentran en su forma original. Dado que las imágenes monocromáticas y las de color pueden ser representadas matemáticamente como funciones determinísticas espacio-temporales, se desarrolló una formulación matemática completa y compleja, además de equipo especializado, incluyendo procesadores digitales de propósito específico y periféricos de computadora.

Tanto la ingeniería electrónica como la ingeniería en sistemas son ramas de un extenso campo de aplicación, como robótica, procesamiento de imágenes, desarrollo de software o telecomunicaciones. Como ejemplo, Matlab es uno de los programas de matemáticas más populares y accesibles en el área de matemáticas y a la vez consta de una herramienta de procesamiento de imágenes que facilita la tarea de construir una aplicación específica. Mediante el procesamiento digital de imágenes es posible manipular imágenes digitales en un computador con el fin de obtener información objetiva de la escena captada por una cámara. Son dos las tareas fundamentales: (1) mejorar una imagen digital con fines interpretativos, (2) automatizar la toma de decisiones de acuerdo al contenido de la imagen digital.

Podemos citar como aplicaciones típicas la detección de presencia de objetos, la inspección visual automática, medición de características geométricas y de color de los objetos, clasificación de objetos, restauración de imágenes y

mejora de la calidad de las imágenes. En este trabajo en particular se han centrado los esfuerzos para identificar un objeto de un color determinado, aislarlo y obtener algunas propiedades del mismo.

3.2.2. Modelos de Color

Desde la perspectiva del ojo humano, todos los colores son vistos como combinaciones variables de los tres colores primarios: rojo (*Red*), verde (*Green*) y azul (*Blue*). Las características utilizadas generalmente para distinguir un color de otro son: brillo, matiz y saturación. El brillo incorpora la noción cromática de intensidad. El matiz es un atributo asociado con la longitud de onda dominante en la mezcla de longitudes de onda de la luz. La saturación se refiere a la pureza relativa o a la cantidad de luz blanca mezclada con un matiz.

El matiz y la saturación cuando se toman conjuntamente se denominan cromaticidad. Y, por tanto, un color se puede caracterizar por su brillo y cromaticidad. Las cantidades de rojo, verde y azul requeridas para formar un color particular se denominan colores triestímulos, denominados X , Y y Z respectivamente. En esta parte del capítulo se introduce el modelo de color utilizado en el trabajo así como un breve repaso sobre algunos otros modelos de color existentes hoy en día.

3.2.2.1. El modelo RGB

En el modelo RGB cada color aparece en sus componentes espectrales primarias: rojo, verde, azul. Este modelo está basado en el sistema de coordenadas cartesianas. El subespacio de color de interés es el tetraedro mostrado en la figura 3.3, en el cual los valores RGB están en tres vértices; cyan, magenta y amarillo se sitúan en otros tres vértices, el negro corresponde al origen y el blanco se sitúa en el vértice más alejado del origen. En este modelo, la escala de grises se extiende desde el negro al blanco a lo largo de la diagonal que une esos dos puntos, y los colores son puntos dentro del tetraedro, definidos por vectores desde el origen.

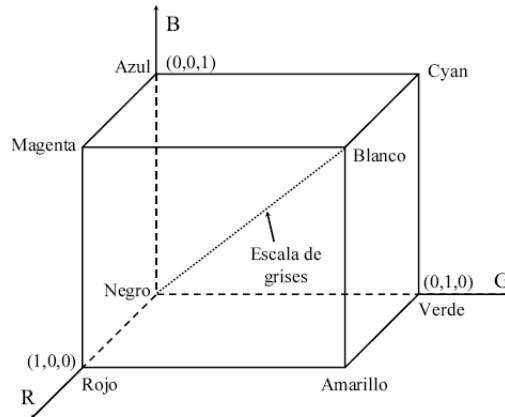


FIGURA 3.3: Tetraedro de color RGB. Los puntos a lo largo de la diagonal principal tienen valores de gris, desde el negro en el origen al punto $(1,1,1)$. Fuente: [MdlC08]

3.2.2.2. Otros modelos de color

Aunque el presente trabajo trata con imágenes en el modelo RGB, en esta sección se comentan de manera general otros modelos de color presentes en la actualidad y se comparan entre sí.

El modelo CMY trabaja mediante la absorción de la luz (colores secundarios). Los colores que se ven son de parte de la luz que no es absorbida. En CMY magenta más amarillo producen rojo, magenta más cian producen azul, cian más amarillo generan verde y la combinación de cian, magenta y amarillo forman negro. Debido a que el negro generado por la mezcla de colores primarios sustractivos, no es tan denso como el color negro puro (uno que absorbe todo el espectro visible). Es por esto que al CMY original se ha añadido un canal clave (*key*) que normalmente es el canal negro (*black*) para formar el espacio CMYK o CMYB. Actualmente las impresoras de cuatro colores, utilizan un cartucho negro además de los colores primarios de este espacio, lo cual genera un mejor contraste. Sin embargo el color que una persona ve en una pantalla de computador difiere del mismo color en una impresora, debido a que los modelos RGB y CMY son distintos. El color en RGB es hecho por la reflexión o emisión de luz, mientras que el CMY mediante la absorción de ésta.

El modelo YIQ se usa en la televisión comercial. Es una combinación de los valores RGB o viceversa. La representación YIQ se emplea a veces en transformaciones de procesamiento digital de imágenes en color. Por ejemplo, aplicando una ecualización del histograma directamente a los canales en una imagen RGB se alterarían los colores unos en relación con otro, resultando una imagen con colores que no tienen sentido. En vez de ello, si la ecualización del histograma es aplicada al canal Y de la representación YIQ de la imagen, sólo se normalizan los niveles de brillo de la imagen. Fue una recodificación realizada para la televisión americana (*NTSC*), la cual tenía que ser compatible con la televisión blanco y negro que solamente requiere del componente de iluminación. Los nombres de los componentes de este modelo son Y por luminancia (*luminance*), I fase (*in-phase*) y Q cuadratura (*quadrature*). Estas últimas generan la cromaticidad del color. Los parámetros I y Q son nombrados en relación al método de modulación utilizada para codificar la señal portadora. Los valores de RGB, son sumados para producir una única señal Y' que representa la iluminación o brillo general de un punto en particular. La señal I luego es creada al restar el Y' de la señal azul de los valores RGB originales y luego el Q se realiza restando la señal Y' del rojo.

El modelo *HSV* es un espacio cilíndrico, pero normalmente asociado a un cono o cono hexagonal, debido a que es un subconjunto visible del espacio original con valores válidos de RGB.

- Tonalidad (*Hue*): Se refiere a la frecuencia dominante del color dentro del espectro visible. Es la percepción de un tipo de color, normalmente la que uno distingue en un arcoiris, es decir, es la sensación humana de acuerdo a la cual un área parece similar a otra o cuando existe un tipo de longitud de onda dominante. Incrementa su valor mientras nos movemos de forma antihoraria en el cono, con el rojo en el ángulo 0.
- Saturación (*Saturation*): Se refiere a la cantidad del color o a la “pureza” de éste. Va de un color “claro” a un color más vivo (azul cielo - azul oscuro). También se puede considerar como la mezcla de un color con blanco o gris.

- Valor (*Value*): Es la intensidad de luz de un color. Dicho de otra manera, es la cantidad de blanco o de negro que posee un color.

3.2.3. Suavizado

El suavizado de imágenes se utiliza normalmente bajo dos supuestos: dar a una imagen un difuminado o efecto especial y para la eliminación de ruido.

El suavizado o filtrado espacial paso bajo borra los detalles más finos de una imagen, es decir, conlleva una atenuación de las altas frecuencias, mientras se mantienen las bajas y medias frecuencias. Tiene un buen número de aplicaciones: algunas veces se emplea para simular una cámara desenfocada, o para restar énfasis a un fondo; mientras los fotógrafos usan un filtro de cámara para conseguir ese efecto, los artistas informatizados emplean filtros digitales.

El suavizado se alcanza mediante la convolución, y es fácil ver en la máscara de convolución que el suavizado es simplemente el promedio del vecindario. Promediar tiende a eliminar los valores extremos de un grupo, así los píxeles extremadamente claros u oscuros pueden hacerse más grises dependiendo de los vecinos del píxel. Cuanto más grande es la máscara, mayor es el efecto de suavizado y mayor el tiempo de cómputo requerido.

Entendemos por ruido en imágenes digitales cualquier valor de un píxel de una imagen que no se corresponde exactamente con la realidad. En el caso que nos ocupa, cuando se adquiere una imagen digital, ésta puede estar contaminada por ruido. El ruido se debe, la mayoría de las veces al equipo electrónico utilizado en la captación de las imágenes (ruido de cuantificación de la imagen, efecto de niebla en la imagen) y al ruido añadido en los tramos de transmisión (posibles interferencias o errores al transmitir los bits de información).

Vamos a distinguir dos clases diferentes de ruido:

- Ruido gaussiano: Se caracteriza por tener un espectro de energía constante para todas las frecuencias. Cuando se presenta este problema, el

valor exacto de cualquier píxel es diferente cada vez que se captura la misma imagen. Este efecto, suma o resta un determinado valor al nivel de gris real y es independiente de los valores que toma la imagen.

- Ruido impulsivo: Se caracteriza por la aparición de píxeles con valores arbitrarios normalmente detectables porque se diferencian mucho de sus vecinos más próximos.

Es por ello, que en este trabajo se ha hecho un primer proceso de suavizado para eliminar ese ruido procedente por naturaleza en las imágenes. Y aunque un filtro paso bajo difumina los bordes de los objetos de las imágenes, para nuestro caso es más que suficiente puesto que lo que es realmente interesante es conocer el centro de gravedad del objeto que intentamos aislar.

3.2.4. Binarizado

Una vez conseguido limpiar de impurezas la imagen, se hace necesario un proceso especial de binarizado para conseguir aislar las zonas interesantes de estudio.

La técnica de binarizado consiste en modificar la imagen original (en color) a una imagen binaria (compuesta únicamente por ceros y unos). En nuestro caso, la tarea fué identificar las zonas de un color específico (rojo) de la imagen y una vez hecho este proceso, binarizarla para proceder a su análisis. Las imágenes capturadas poseen una gran resolución y es por ello que la rapidez de los algoritmos y el poco gasto de memoria es algo fundamental cuando queremos reconocer objetos por su tamaño, forma o cualquier detalle en los que no intervienen ni los grises ni el color.

Por tanto, el proceso que se siguió fue el siguiente:

1. Descomposición de la imagen. Se extraen de la imagen en RGB, las bandas correspondientes al rojo y verde. Estas bandas de color son matrices numéricas, por lo que podemos realizar operaciones de comparación sobre ellas.

2. Primer filtro tomando la matriz correspondiente al rojo. El valor de intensidad correspondiente al rojo puro y el blanco será 255. Diversos factores como la iluminación de la escena provocan que el rojo no sea en muchas ocasiones el rojo puro. Por tanto, se hace necesario un proceso de binarización de acuerdo a un umbral fijado ε_{rojo} , de manera que si consideramos $f_{rojo}(x, y)$ el valor de un pixel en la imagen roja original, el nuevo valor en la imagen modificada $f_{bin}(x, y)$:

$$f_{bin}(x, y) = \begin{cases} 1 & \text{si } f_{rojo}(x, y) \geq \varepsilon_{rojo} \\ 0 & \text{si } f_{rojo}(x, y) < \varepsilon_{rojo} \end{cases}$$

3. Segundo filtro tomando la matriz correspondiente al verde. La imagen resultante de la etapa anterior será uno si el pixel original era blanco o rojo y cero en caso contrario. En esta fase lo que se pretende es eliminar los tonos blancos para que consigamos tener una imagen en la que solo tengamos objetos de color rojo. De manera similar a la etapa anterior, se realiza un proceso de binarización de acuerdo a un umbral fijado ε_{verde} , donde la imagen final resultante es la misma $f_{bin}(x, y)$ de acuerdo a la siguiente ecuación:

$$f_{bin}(x, y) = \begin{cases} f_{bin}(x, y) & \text{si } f_{verde}(x, y) \geq \varepsilon_{verde} \\ 0 & \text{si } f_{verde}(x, y) < \varepsilon_{verde} \end{cases}$$

Así tras este proceso, obtenemos una imagen en la que los objetos de tonalidad roja están presentes (con valor uno, y por tanto en blanco) sobre un fondo negro, que efectivamente era lo que se quería conseguir. A partir de aquí, solo queda realizar una operación morfológica de apertura sobre la imagen para eliminar las distorsiones que han podido producirse durante toda la etapa de binarizado. Este operación se detalla en la sección 3.2.7.

3.2.5. Operaciones sobre imágenes digitales

Desde el punto de vista del tratamiento de imágenes y una vez que las mismas son consideradas como matrices numéricas, podemos realizar las mismas operaciones que se pueden efectuar con las matrices, eso sí, con las debidas limitaciones derivadas del hecho de la cuantización de los niveles de intensidad.

Esta cuantización se refiere al hecho de que una imagen sólo va a tomar valores en el rango preestablecido, como ya sabemos generalmente en el intervalo $[0, 255]$. Esas operaciones a las que nos referíamos anteriormente son las operaciones aritméticas habituales de suma, resta, multiplicación, división o producto. Además, dado que los datos en un computador finalmente se pueden representar en forma binaria, siempre es posible realizar cualquier tipo de operación lógica binaria sobre ellos, dando lugar a las denominadas operaciones lógicas elementales: and, or, xor, not, ...

Pero solamente nos vamos a centrar en las operaciones morfológicas, muy utilizadas en imágenes binarias como es nuestro caso, después de haber aplicado una binarización especial a las imágenes originales.

3.2.6. Operaciones morfológicas

Las operaciones morfológicas han sido muy utilizadas en imágenes binarias. En primer lugar hay que determinar el origen de la imagen desde el punto de vista de las coordenadas (x, y) . El origen se identifica gráficamente con un punto situado en uno de los píxeles.

Con la imagen dada y el origen fijado se construye un conjunto cuyos elementos son los pares de coordenadas (x, y) que indican las posiciones de los “unos” lógicos. Una transformación morfológica φ viene dada por la relación de la imagen (conjunto de puntos X) con otro pequeño conjunto de puntos B , llamado operador o elemento estructural. B se expresa también con respecto a un origen local O (llamado punto representativo o elemento director).

3.2.6.1. Dilatación

La transformación morfológica de la dilatación \oplus combina dos conjuntos utilizando la adición de vectores (o adición de conjuntos de *Minkowski*). La dilatación $X \oplus B$ es el conjunto de puntos de todas las posibles adiciones vectoriales de pares de elementos, uno de cada conjunto X y B .

$$X \oplus B = \{d \in E^2 : d = x + b \text{ para cada } x \in X \text{ y } b \in B\}$$

3.2.6.2. Erosión

La transformación morfológica de la erosión \otimes combina dos conjuntos utilizando la sustracción de vectores. Es dual de la dilatación.

$$X \otimes B = \{d \in E^2 : d + b \in X \text{ para cada } b \in B\}$$

Esta expresión dice que cada punto d del conjunto X , que para nosotros será la imagen, es comprobado; el resultado de la erosión está dado por los puntos d para los cuales todos los posibles $d + b$ están en X .

3.2.7. Apertura

La erosión y dilatación son transformaciones no invertibles. Si una imagen es erosionada y luego dilatada, la imagen original no se recupera. En efecto, el resultado es una imagen más simplificada y menos detallada que la imagen original. La erosión seguida de una dilatación crea una transformación morfológica importante llamada apertura. La apertura de una imagen X por un elemento estructural B se denota por $X \circ B$ y se define como

$$X \circ B = (X \otimes B) \oplus B$$

La apertura en el caso que nos ocupa es de gran utilidad, puesto que permite en gran medida eliminar el exceso de ruido que queda en una imagen, y lo más importante, sin eliminar de la imagen los objetos que nos interesan.

3.3. Análisis de imágenes digitales

Una vez, explicados los métodos de procesamiento empleados, la etapa de análisis comprende la extracción de las regiones que nos interesan (el objeto) y la obtención de algunas de sus propiedades.

3.3.1. Identificación de regiones

Las capacidades de análisis de regiones permiten identificar y extraer características de regiones de píxeles conectados entre sí (comúnmente concidos como *blobs*) dentro de una imagen. Para llevar a cabo, un análisis de cuantos píxeles pertenecen a un *blob* determinado de la imagen se necesitan algunas técnicas que están implementadas con éxito en las librerías MIL. A partir de un identificador de *blob*, se pueden extraer características que involucran al completo la región de la imagen original.

El módulo de análisis que los píxeles que se están tocando entre sí forman parte del mismo *blob*. Por consiguiente, lo que es fácilmente identificable por el ojo humano también lo es por parte de las MIL. Además, cualquier parte de la región que esté en contacto con el fondo, a causa de la luz o los reflejos, es considerada como parte del fondo durante el análisis.

Para reducir el preprocesamiento, el identificador de *blob* debería ser adquirido bajo las mejores circunstancias posibles. Esto quiere decir, que distintas regiones no se superpongan y si es posible que no se toquen. También quiere decir, que las condiciones de luz sean óptimas con unos valores de gris distintos entre el fondo y los objetos.

3.3.2. Segmentación de regiones

Una vez que todos los *blobs* son claramente identificables, ya puede empezar el proceso de calcular características sobre ellos. Sin embargo, en algunos casos, no queremos que el tiempo de ejecución se exceda más de lo necesario, calculando propiedades sobre cada uno de los *blobs* de la imagen. En nuestro

caso, eso rara vez ocurrirá puesto que solo disponemos de un objeto a analizar.

Una vez conocida la localización del *blob* en cuestión, necesitamos conocer su posición en la imagen, así como otros factores con los que comparar, como el área. Este aspecto es importante, puesto que permite desechar aquellas regiones que no son del área esperada, si es que conocemos a priori ese valor. Sumado al cálculo del centro de gravedad (en la figura 3.4 se pueden ver algunas de sus características), ya podremos obtener tanto sus coordenadas x e y como realizar una conversión adecuada a coordenadas geográficas.

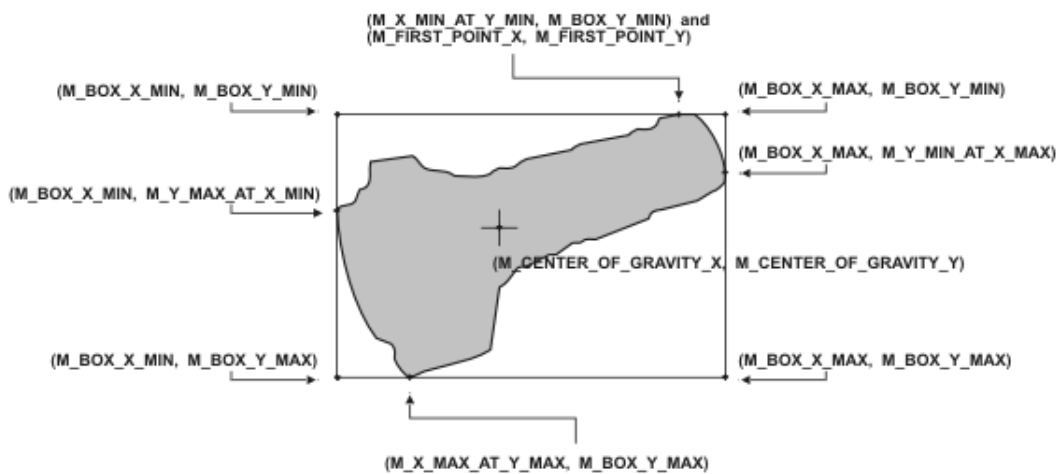


FIGURA 3.4: Algunas de las propiedades de un *blob*.

Capítulo 4

Coordenadas Geográficas y GPS

En este capítulo se explicarán los fundamentos matemáticos del posicionamiento geográfico de un punto cualquiera de la Tierra y a su vez se dará un breve repaso de un sistema global de posicionamiento (GPS) y de las características específicas del GPS utilizado para el presente trabajo.

4.1. Fundamentos matemáticos

4.1.1. Introducción

Básicamente la localización geográfica de un punto se puede hacer especificando cualquiera de estos parámetros:

- Coordenadas geográficas dadas en Longitud y Latitud.
- Coordenadas UTM (*Universal Transversal Mercator*).

Sea cuál sea la forma de localizar un punto sobre la superficie terrestre, se deben cumplir los siguientes requisitos:

1. Que el punto sea único.
2. Que quede perfectamente identificado el sistema de proyección empleado al localizar el punto.
3. Que permita referenciar la coordenada Z del punto.

Dadas la características del proyecto, se le va a prestar una especial atención a las coordenadas geográficas.

La localización de un punto en coordenadas geográficas viene expresado como la separación en grados ($^{\circ}$), minutos ($'$) y segundos ($''$) con respecto a un paralelo y un meridiano de referencia (el ecuador de la Tierra y el meridiano de *Greenwich* respectivamente).

Este hecho supone la creación de un sistema de referencia de tres dimensiones, donde el punto P está definido por r , λ y Φ , y puede verse de manera gráfica en la figura 4.1.

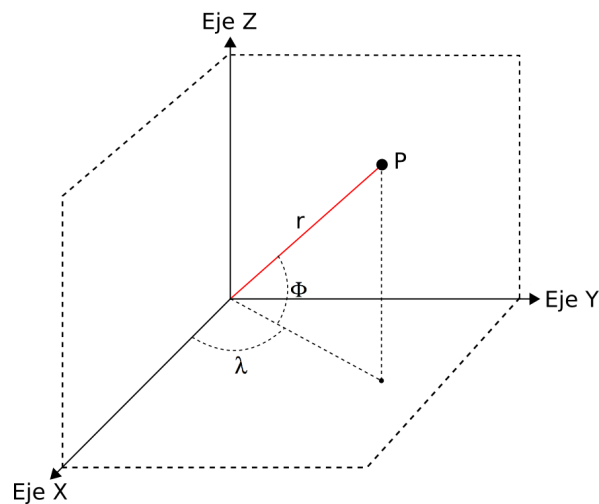


FIGURA 4.1: Sistema de referencia de tres dimensiones X , Y y Z .

Así, si consideramos al eje de la Tierra como la recta ideal de giro de la esfera terrestre en su movimiento de rotación, es decir, la recta que une los dos polos geográficos (Polo Norte y Polo Sur), se definen los meridianos y los paralelos, descritos en las subsecciones siguientes 4.1.2 y 4.1.3. Una vez creada una red de meridianos y paralelos, la situación geográfica de un punto estará definida por su Longitud y Latitud, ambas explicadas con detalle en las subsecciones ??.

4.1.2. Meridianos

Se definen los meridianos como las líneas de intersección de los infinitos planos que contienen al eje de la Tierra con la superficie terrestre.

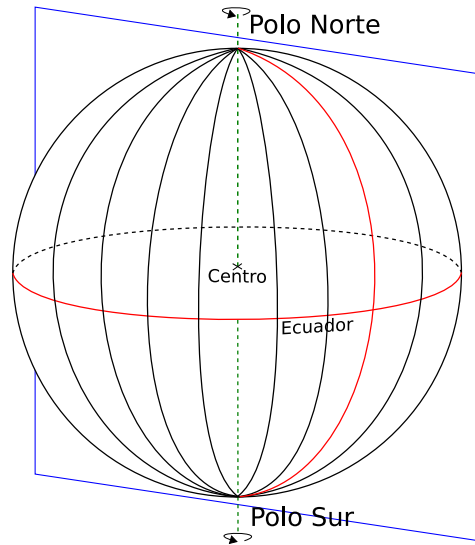


FIGURA 4.2: Representación de un meridiano en la superficie terrestre.

El sistema toma como origen de coordenadas el meridiano que pasa por la ciudad inglesa de *Greenwich*. Ese meridiano es el “meridiano 0”.

Así tenemos la esfera terrestre dividida en dos zonas, las situadas al Oeste (*W*) del meridiano 0 y las situadas al Este (*E*). La Tierra quedaría dividida aproximadamente de la manera observada en la figura 4.2.

4.1.3. Paralelos

Se definen los paralelos como las líneas de intersección de los infinitos planos perpendiculares al eje terrestre con la superficie de la Tierra.

El sistema toma como centro de referencia el paralelo que está a mayor distancia del centro de la Tierra, es decir, el Ecuador. Así tenemos dividido el Globo en dos hemisferios, el Norte (*N*) y el Sur (*S*).

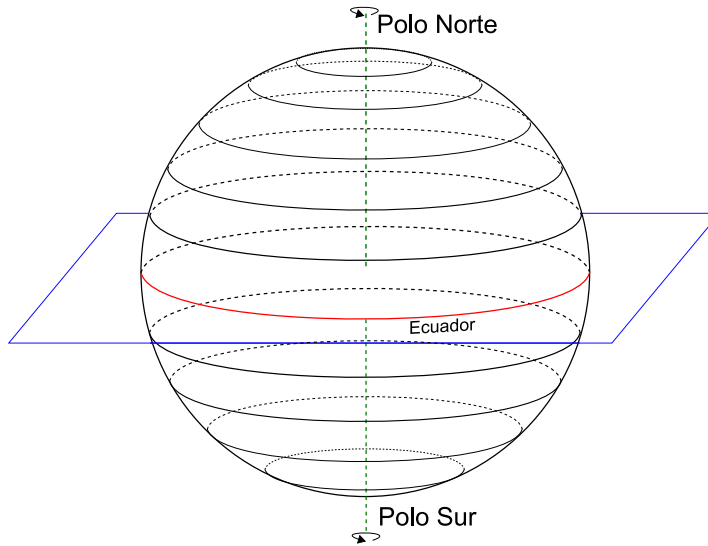


FIGURA 4.3: Representación de un paralelo en la superficie terrestre.

4.1.4. Longitud

Se define la Longitud λ de un punto P como el valor del diedro formado por el plano meridiano que pasa por P y el meridiano origen, o meridiano de Greenwich.

La longitud es el ángulo formado por AOB .

Al ser el origen el meridiano de Greenwich debemos dar la posición de nuestro punto respecto de dicho origen. Así, si nuestra posición está a la izquierda del origen designamos la posición Oeste (W) y si está a la derecha designamos la posición Este (E).

La longitud presenta un mínimo de 0° y un máximo de 180° .

4.1.5. Latitud

Se define la latitud geográfica Φ de un punto P al ángulo formado por la vertical a la Tierra que pasa por dicho punto con el plano ecuador. Siendo la vertical la unión del punto con el centro de la tierra, se obtiene la latitud midiendo el ángulo sobre el meridiano que pasa por el punto P .

La latitud será el ángulo formado por OAP .

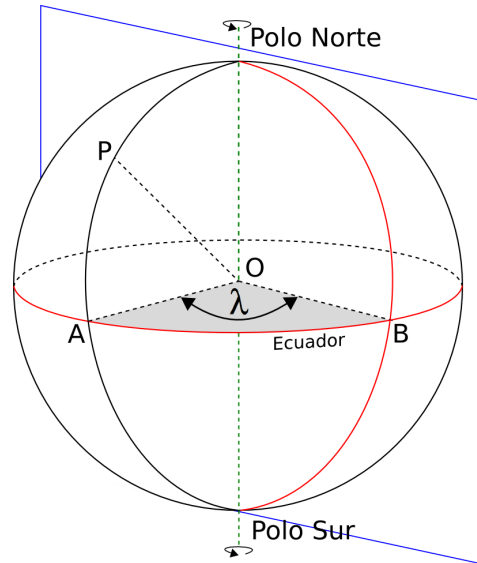


FIGURA 4.4: Visualización gráfica de la longitud.

Al ser el origen de referencia el Ecuador de la Tierra, tendremos que dar la posición de nuestro punto referido a este origen. Así todo punto que esté en el hemisferio Norte será designado en la posición Norte (N) y todo punto que esté en el hemisferio Sur será designado en la posición Sur (S).

La latitud mínima es de 0° y la máxima es de 90° .

Como curiosidad destacar que España (incluyendo las Islas Baleares y Ceuta pero no las Islas Canarias ni Melilla) está situada entre las longitudes $4, 33^\circ$ W y $9, 29^\circ$ W, y las latitudes $35, 82^\circ$ N y $43, 80^\circ$ N.

4.2. Global Positioning System

El GPS o Global Positioning System es un Sistema Global de Navegación por Satélite (GNSS) que permite determinar en todo el mundo la posición de un punto, desarrollado por el Departamento de Defensa de Estados Unidos (DOD).

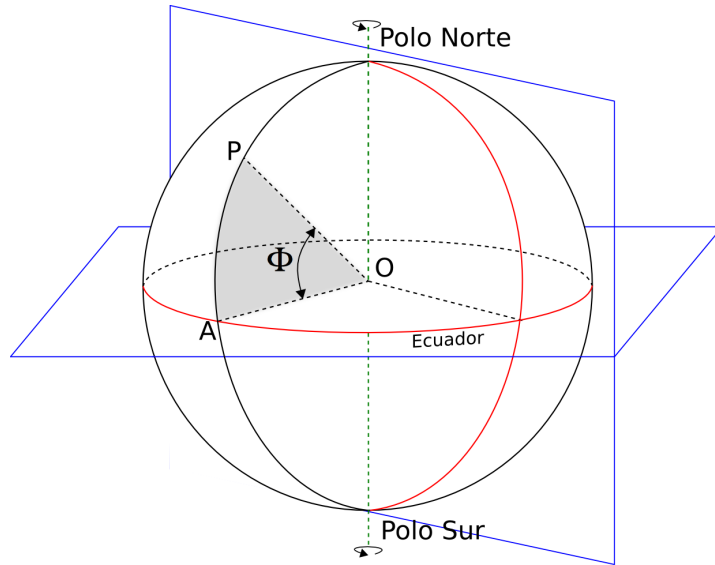


FIGURA 4.5: Visualización gráfica de la latitud.

Funciona mediante una constelación de 27 satélites NAVSTAR (24 operativos y 3 de reserva) en órbita sobre la Tierra, a una altura de 20.200 km., con trayectorias sincronizadas para cubrir toda la superficie de la Tierra.

Para localizar el punto en el que está un punto, el receptor localiza como mínimo a 3 de esos satélites que le mandan una señal indicando la posición y el reloj de cada uno de ellos. Así, midiendo cuanto tarda en llegarle la señal puede saber la distancia a la que está cada satélite y mediante triangulación saber su posición geográfica.

Este sistema de posicionamiento lo componen:

1. Sistema de satélites. Está formado por 24 unidades con trayectorias sincronizadas para cubrir toda la superficie del globo terráqueo. Más concretamente, repartidos en 6 planos orbitales de 4 satélites cada uno. La energía eléctrica que requieren para su funcionamiento la adquieren a partir de dos paneles compuestos de celdas solares adosados a sus costados.
2. Estaciones terrestres. Envían información de control a los satélites para

controlar las órbitas y realizar el mantenimiento de toda la constelación.

3. Terminales receptores: Indican la posición en la que están; conocidas también como Unidades GPS, son las que podemos adquirir en las tiendas especializadas.

Especificaciones del sistema GPS:

- Satélites en la constelación: 24 (4 X 6 órbitas)
 - Altitud: 20.200 km
 - Período: 11 h 56 min
 - Inclinação: 55 grados (respecto al ecuador terrestre).
 - Vida útil: 7,5 años
- Segmento de Control (estaciones terrestres):
 - Estación principal: 1
 - Antena de tierra: 4
 - Estación monitora (de seguimiento): 5
- Señal RF:
 - Frecuencia portadora:
 - Civil: 1 575.42 MHz (L1). Utiliza el Código de Adquisición Aproximativa (C/A).
 - Militar: 1227.60 MHz (L2). Utiliza el Código de Precisión (P), cifrado.
 - Nivel de potencia de la señal: -160 dBW (en superficie tierra).
 - Polarización: circular dextrógira.
- Exactitud:
 - Posición: aproximadamente 15 m (el 95 %).
 - Hora: 1 ns.

- Cobertura: mundial
- Capacidad de usuarios: ilimitada
- Sistema de coordenadas:
 - Sistema Geodésico Mundial 1984 (WGS84)
 - Centrado en la Tierra, fijo

Así, después de conocer el funcionamiento del sistema GPS vamos a centrarnos en el receptor que hemos utilizado en el desarrollo de este proyecto.

4.2.1. GPS MIDG II

El sistema utilizado en el presente trabajo es un sistema de navegación inercial ayudado por GPS, esto es, que aparte del GPS el sistema consta de unos giroscópos que nos dan información también de la posición del aparato respecto al horizonte y del ángulo que lo separa del Norte, como características más destacadas.

El sistema en cuestión se trata de un MIDG II de la casa Microbotics, Inc. Los requerimientos del sistema son los siguientes:

- Un PC corriendo Microsoft Windows.
- Un conversor de RS-422 a RS-232, para conectar el MIDG II al PC.
- Una fuente de corriente que provea una tensión de continua entre 9 y 36 voltios y una potencia de 2 vatios mínimo.
- Una antena GPS con un conector SMA.

El acceso a la información en este dispositivo es de fácil acceso gracias al uso de su display, que permite no tener que hacer uso de algún otro tipo de software. Un uso detallado del mismo se encuentra descrito en el apéndice C.

4.3. Conversión a longitud y latitud

Afrontemos ahora el problema de pasar las coordenadas de un punto, que tenemos referenciado respecto de un origen particular nuestro, a coordenadas geográficas.

Es decir, suponiendo que tenemos el sistema de cámaras en una posición fija, la cual tomaremos como origen, el barco estará a una distancia x en el eje de abscisas y a una distancia y en el eje de ordenadas de dicho origen (la posición de nuestras cámaras).

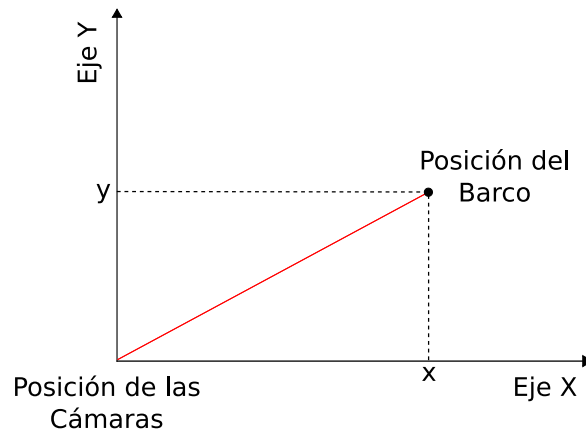


FIGURA 4.6: Ejes del sistema de conversión a coordenadas geográficas.

Esa posición (x, y) , que es la que hayamos con nuestros sistemas de visión, nos puede interesar darla en coordenadas geográficas (longitud y latitud), para lo cual tendremos que hacer algunos cálculos.

Como la tierra no es una esfera perfecta, todos los sistemas de coordenadas que existen hacen una aproximación en forma de elipsoide, la cual tendrá un semieje mayor y un semieje menor. El sistema de coordenadas que usa el GPS es el WGS84 (World Geodetic System) el cual usa los siguientes semiejes:

- Semieje mayor: $a = 6378137$ metros.
- Semieje menor: $b = 6356752,3142$ metros.

Con estos datos podemos hallar la primera excentricidad mediante la fórmula:

$$e_1 = \frac{a^2 - b^2}{a^2}$$

También podemos hallar la normal principal, lo cual haremos mediante la expresión:

$$N = \frac{a^2}{\sqrt{a^2 \cdot \cos^2 \Phi + b^2 \cdot \sin^2 \Phi}}$$

Donde Φ es, como ya sabemos, la latitud. Con ayuda de la excentricidad y de los semiejes podemos hallar el radio principal de curvatura, de la siguiente manera:

$$R = \frac{a \cdot (1 - e_1^2)}{(1 - e_1^2 \cdot (\sin^2 \Phi))^{3/2}}$$

Así, finalmente, el radio de la Tierra (realmente del elipsoide) para un punto cualquiera sabiendo su latitud, es la componente cuadrática del radio principal de curvatura y de la normal principal, esto es:

$$R_T = \sqrt{R \cdot N}$$

Entonces, si sabemos que un arco de circunferencia está dado por la expresión $arco = radio \cdot angulo$, y tomamos el arco como la posición x o la y, podremos hallar fácilmente las coordenadas geográficas si hacemos:

$$latitud = \frac{x}{R_T} \quad longitud = \frac{y}{R_T}$$

Aunque surge un problema, ya que esto sería así de fácil si los ejes de ordenadas y abscisas de mi sistema de cámaras estuvieran alineados con un paralelo y un meridiano, lo cual puede no ocurrir, dando lugar a situación mostrada en la figura 4.7.

Así pues, en esta situación, tendremos que hallar la proyección de x e y sobre el meridiano y el paralelo que pasen por la posición de las cámaras. Para ello necesitamos saber el ángulo α que separa a ambos ejes, el cual es fácilmente medible con una brújula o con el INS, ya que sabemos que todos los meridianos apuntan al Norte, por lo cual solo necesitaremos saber que ángulo

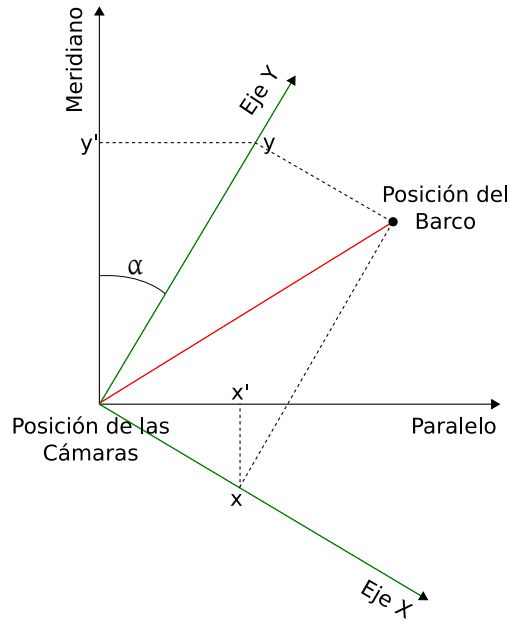


FIGURA 4.7: Representación gráfica del problema de las cámaras no alineadas.

nos separa del Norte.

Podemos considerar el meridiano y el paralelo como rectas ya que el campo de movimiento de nuestro barco va a estar limitado a las dimensiones del pantano, las cuales, comparadas con la superficie de la Tierra, podemos considerar que son planas, en 2D.

Así, una vez halladas estas proyecciones x' e y' y aplicando las expresiones antes usadas, hallaríamos la longitud y la latitud deseadas.

$$latitud = \frac{x'}{R_T} \quad longitud = \frac{y'}{R_T}$$

Capítulo 5

Implementación

Una vez explicados las técnicas y fundamentos de la visión por computador así como el procesamiento de imágenes y la motivación de las coordenadas geográficas, se continuará exponiendo algunos detalles de la implementación.

5.1. *Matrox Imaging Library*

Matrox Imaging Library (MIL) es una librería de 32bits modular e independiente del hardware. Es por esa razón que nosotros hemos utilizado gran parte de su potencia para implementar el sistema.

MIL ha sido diseñada para llevar a cabo un desarrollo rápido de aplicaciones y ser fácil de usar. Tiene un sistema de control completamente transparente que maneja objetos físicos (*systems, digitizers, displays* y *data buffers*) como objetos virtuales, permitiendo aplicaciones independientes de la plataforma. Esto significa que una aplicación MIL puede ejecutarse en cualquier placa VGA sobre diferentes sistemas operativos.

Aunque MIL se compone de dos versiones distintas, la versión utilizada en el proyecto es la completa, que contiene toda la funcionalidad para el procesado y análisis de imágenes. Permite grabar, mostrar, guardar, transferir imágenes de modo simple así como una extensa base de operaciones de procesamiento

y análisis.

Las operaciones de procesado incluyen algunas como punto a punto, estadísticas, filtrado espacial, morfológico, geométrico y transformaciones de Fourier, entre las más destacadas. Las operaciones de análisis incluyen, medición, análisis de regiones, calibración, reconocimiento de caracteres y correspondencia de patrones, entre las más importantes.

Se puede encontrar más información sobre las librerías en el Apéndice B. Por último destacar, que este paquete de librerías no es gratuito, por lo que se necesita la correspondiente licencia que habilite todas las operaciones que tiene implementadas.

5.2. Cámaras y lentes

Para este trabajo, se han necesitado dos cámaras a través de las que capturar y procesar la información. El modelo escogido de las cámaras es *Scout* de la marca *Basler*. En la figura 5.1 se muestra una captura de una de estas cámaras. Las cámaras poseen una entrada IEEE 1394b así como una toma de corriente por si se necesita energía adicional. Además dispone de un indicador *led* que muestra su estado. Verde continuo significa que todo está bien, verde parpadeando quiere decir que a la cámara no le llega alimentación suficiente (esto puede ser por el ancho de banda limitado a unas resoluciones de imágenes no tan grandes como cabría esperar) y naranja significa suspensión.



FIGURA 5.1

El resto de las características específicas de la cámara se encuentran de-

scritas en la siguiente tabla 5.2. Es importante fijarse en el tamaño de píxel que es de $6.45 \mu\text{m}$ puesto que ese valor será fundamental para el cálculo de la disparidad.

scA1400-17fm/fc	1392 × 1040	Sony ICX285	Progressive scan CCD	2/3"	6.45 × 6.45	17	3 W	160g
Mono/Color	Yes/Yes							
Video Output Type (Interface)	IEEE1394b (screw lock possible)							
Video Output Format	Mono 8: 8 bits/pixel Mono 16: 12 bits/pixel YUV 4:2:2: 16 bits/pixel average YUV 4:2:2: (YUYV): 16 bits/pixel average Raw 8: 8 bits/pixel (R,G, or B) Raw 16: 12 bits/pixel (R,G, or B) scA750-60fm/fc (8 bits/pixel only)							
Synchronization	Via external trigger, via the IEEE1394 bus, or free run							
Exposure Control	Programmable via the 1394 bus							
Power Requirements	8 – 36 VDC; provided via the IEEE1394 cable; < 1% ripple							
Lens Mount	C-mount; CS-mount (optional)							
Housing Size (L × W × H)	73.7 mm × 44 mm × 29 mm (without lens adapter)							
Housing Size 90° (L × W × H)	41.8 mm × 44 mm × 97 mm (incl. lens adapter)							
Conformity	CE, FCC, DCAM, RoHS, IP 30							
I/O Ports	2 opto-isolated input ports, 4 opto-isolated output ports							
Software Driver	Basler pylon, IEEE1394b driver, Basler BCAM driver							
GenICam Compatible	Yes (No, with BCAM 1.9)							

Specifications are subject to change without prior notice.

FIGURA 5.2

Por otra parte, cabe comentar las lentes utilizadas. Para este trabajo, se han utilizado dos lentes distintas, de 25mm y 100mm respectivamente. La primera de ellas, se ha utilizado para los experimentos en distancias cortas (hasta 60 metros) mientras que la segunda se ha usado en las pruebas a más distancia (a partir de 200 metros). Y como es obvio, se dispone de cuatro lentes, dos para cada cámara.

5.3. Estructura general

La aplicación diseñada está implementada en C++ bajo el entorno de desarrollo *Visual Studio .NET*. La aplicación está dividida en dos partes, desde el punto de vista del código:

- La interfaz (`Halcon.h` y `Halcon.cpp`). Este fichero engloba todos los métodos y funciones correspondientes a la vista del sistema, así como todas las llamadas y el uso de mensajes de la API de Windows. Los

comandos aquí especificados hacen uso del fichero de procesamiento y análisis.

- El procesamiento y análisis (`Procesamiento.h` y `Procesamiento.cpp`). En este archivo, se han implementado todas las funciones correspondientes al procesamiento y análisis de las imágenes digitales. Utilizan algunas de las funciones proporcionadas por las librerías MIL y también el resto de ellas han sido implementadas para su uso específico en esta aplicación.

Capítulo 6

Resultados experimentales

Una vez implementadas las dos técnicas de visión, llega el momento de obtener y evaluar los resultados. El objetivo del proyecto es obtener distancias y medidas de manera precisa, por lo que en este capítulo se analizarán factores importantes, como la repetitividad de los datos, los errores en la distancia, la influencia de los factores meteorológicos y la precisión, entre otros. En primer lugar se comentará la calidad de los resultados, para una vez constatada esta continuar con otros resultados.

6.1. Inicio y preparación de las pruebas

Tras haber acabado el desarrollo del sistema, y después de haber llevado a cabo distintas pruebas pequeñas de partes aisladas del proyecto llega el momento de realizar pruebas a mayor escala. Para ello, se buscaron distintas localizaciones donde poder probar y testear nuestro trabajo.

En un primer momento, se estuvo pensando en la posibilidad de realizar varias pruebas directamente en un pantano, pero esa idea fue dejada para más adelante para centrarnos en obtener medidas a corta y larga distancia. Para las pruebas a corta distancia, utilizamos las instalaciones de la Facultad de Físicas, puesto que eran más que suficientes. En cambio, para las pruebas a larga distancia, se decidió por realizar dos escapadas: una al Monte del Pardo

y otra a los campos de fútbol de Cantaranas, en el campus de la Universidad Complutense de Madrid.

Es por ello, que en las siguientes secciones se estudiarán de manera alterna las experiencias y resultados extraídos de cada actividad.

6.2. Calidad de los resultados

Para evaluar la calidad de los resultados obtenidos, se han de comentar primero un par de factores importantes. Centrándonos en la técnica de visión estereoscópica, se ha de decir primero, que el proceso de alineación de las cámaras no es algo sencillo, ya que conseguir que los ejes ópticos de las dos cámaras sean totalmente paralelos es bastante complicado. La plataforma que las sostiene, consta de unas cabezas de trípode micrométricas a través de las cuales poder girar a izquierda o derecha y/o arriba o abajo cada una de las cámaras por separado. El sistema Halcón dispone de una herramienta diseñada para realizar un proceso de alineación supervisado y es el siguiente:

1. Posicionar el objeto que se quiera a una distancia z conocida.
2. Con ese valor, el sistema analizará las imágenes y de acuerdo con la ecuación de la disparidad obtenida en el capítulo 2 $z = \frac{f \cdot b}{d}$, lo que se realiza es un cálculo de la disparidad teórica despejando d :

$$d = \frac{f \cdot b}{z}$$

3. Se compara la disparidad teórica con la disparidad experimental y en función de su valor se van ajustando las cámaras convenientemente. La alineación es por tanto, una tarea que se realiza de forma manual.

Así en teoría conseguiríamos una disparidad exactamente igual a la buscada, pero existe un pequeño problema que hace que la alineación no llegue a ser perfecta: la luz. Efectivamente, en el procesamiento y análisis de las imágenes puede ocurrir que dependiendo de la intensidad de luz de la escena, aunque

a primera vista pudiese parecer que ambas imágenes deben tener las mismas condiciones, esto no sea así y en algunas ocasiones el centro de gravedad del objeto identificado varíe un poco en cada captura. Sin embargo, este desfase será como mucho de un píxel.

Segundo, antes de proceder a realizar cualquier experimento es necesario ajustar una serie de parámetros, como son, las intensidades de rojo y verde adecuadas a las condiciones de luz para que la extracción e identificación del objeto sea la mejor; la focal utilizada en las lentes, la distancia entre las cámaras y la altura del sistema; la latitud, longitud y ángulo con respecto al norte del sistema de cámaras, necesario para poder realizar la conversión a coordenadas geográficas. Si se desea más información acerca del uso del sistema, refiérase al apéndice A.

Una vez comentados estos aspectos iniciales, se va a estudiar la validez de los resultados en distintos experimentos. Como ya se ha comentado en capítulos anteriores, se dispone de un GPS para referenciar el sistema de las cámaras y para tomar mediciones en cada uno de los puntos a medir. Por tanto, estudiar la calidad de los resultados significa que se va a realizar una comparación entre los datos calculados por el GPS, el sistema Halcón y los datos teóricos.

En esta sección, nos centraremos en el análisis de los puntos en Cantaranas, por ser unos de los más interesantes.

Se realizó un experimento en los campos de fútbol del campus. Como objeto a calcular su distancia, escogimos una bandera roja de tamaño 40 x 40 centímetros. Se colocó a tres distancias distintas, 100, 150 y 200 metros y las lentes utilizadas tenían por focal 100mm. En la tabla 6.1 se observan las medidas a cada uno de los puntos dadas por el GPS. Como se puede ver, se comete un error de aproximadamente 10 metros en cada posición. Sin embargo, en la tabla 6.2 están los datos medidos por nuestro sistema. Se ve claramente, que las medidas obtenidas son mucho más próximas a las ideales, aunque no llegan a serlo del todo.

	P.0m	P.100m	P.150m	P.200m	P.200m ⇒	P.200m ⇐
P.0m	xxxxxx	90.61	140.69	190.38	xxxxxx	xxxxxx
P.100m	xxxxxx	xxxxxx	51.09	101.12	xxxxxx	xxxxxx
P.150m	xxxxxx	xxxxxx	xxxxxx	50.04	xxxxxx	xxxxxx
P.200m	xxxxxx	xxxxxx	xxxxxx	xxxxxx	2	2
P.200m derecha	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
P.200m izquierda	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx

TABLA 6.1: Distancias entre los puntos elegidos en Cantarranas con GPS.

	P.0m	P.100m	P.150m	P.200m	P.200m ⇒	P.200m ⇐
P.0m	xxxxxx	99.62	146.032	198.35	xxxxxx	xxxxxx
P.100m	xxxxxx	xxxxxx	46.412	98.73	xxxxxx	xxxxxx
P.150m	xxxxxx	xxxxxx	xxxxxx	52.318	xxxxxx	xxxxxx
P.200m	xxxxxx	xxxxxx	xxxxxx	xxxxxx	2	2
P.200m derecha	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
P.200m izquierda	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx

TABLA 6.2: Distancias entre los puntos elegidos en Cantarranas con Halcón.

Con el motivo de poder analizar un poco más esta situación, en la figura 6.1 se muestran gráficamente los valores obtenidos para estas tres posiciones.

Por otro lado, las medidas anteriores convertidas a coordenadas geográficas están descritas en la figura 6.2. Las dos primeras columnas de cada tabla son los valores de latitud y longitud hallados para cada uno de los puntos de Cantarranas y El Pardo. En la primera tabla se muestra una conversión de coordenadas geográficas y UTM datum WGS84 a datum ED50e, mientras que en la inmediatamente inferior se hace lo inverso. En ambos experimentos, se utilizó la focal de 100mm para el cálculo de medidas. Los puntos Pardo 1, Pardo 2 y Pardo 3, corresponden a las medidas a 0, 100 y 200 metros respectivamente. Los de Cantarranas expuestos a continuación corresponden a los explicados en esta sección del capítulo.

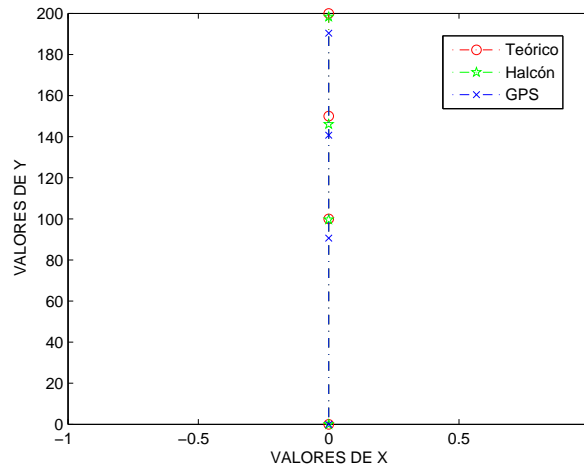


FIGURA 6.1: Gráfica que muestra la media de resultados para los tres puntos.

6.3. Repetitividad

La repetitividad de los datos es otro factor importante a considerar. Determina el número de valores posibles que pueden tomar x e y a lo largo del tiempo, es decir, tras tomar una gran cantidad de datos en la misma posición.

Por ello, en primer lugar se va a estudiar la repetitividad en distancias grandes, para 100, 150 y 200 metros como en el apartado anterior. Las figuras 6.3, 6.4 y 6.5 reflejan resultados interesantes. En cada una de las gráficas se muestran en azul los valores obtenidos durante un muestreo de 200 datos. Se puede apreciar en cada una de ellas, que los valores de distancia y siempre se encuentran oscilando en torno a dos valores, dependiendo de la situación. Y ya conocemos el por qué de este resultado: las cámaras están alineadas con una disparidad que difiere de la óptima como mucho en 1 píxel. A su vez, también se muestra en cada una de las figuras, el valor medio de todos los datos, expresado en las tablas 6.1 y 6.2 y pintada en color rojo.

En segundo lugar, tal y como hemos dicho vamos a analizar la repetitividad de los datos en distancias cortas. En esta ocasión se hicieron pruebas bastante cerca, a 20 y 30 metros. Las figuras expuestas a continuación 6.6 y 6.7 mues-

Conversión coordenadas geográficas y UTM datum WGS84 a datum ED50																									
n° pun to	ENTRADAS: Latitud				ENTRADAS: Longitud				Salida UTM (WGS84)		SALIDA: Coordenadas Geográficas (ED50)		SALIDA: UTM (ED50)												
	g°	min'	s"	N/S	g°	min'	s"	E/W	Huso	Este X	Norte Y	g°	min'	s"	N/S	g°	min'	s"	E/W	Huso	Este X	Norte Y			
1°	40°	19'	0,5"	N	3°	15'	21,5"	W	80T 478251	: 446295340°	19' 4.78"N	3° 15' 16.89"W	N40 19 4.78	W3 15 16.89	30T 478360	:: 4463160					30T 478360	:: 4463160			Pardo 1
2°	40°	19'	0,2"	N	3°	15'	17,6"	W	80T 478343	: 446294440°	19' 4.49"N	3° 15' 12.99"W	N40 19 4.49	W3 15 12.99	30T 478452	:: 4463151					30T 478452	:: 4463151			Pardo 2
3°	40°	19'	0,1"	N	3°	15'	13,7"	W	80T 478435	: 446293940°	19' 4.35"N	3° 15' 9.06"W	N40 19 4.35	W3 15 9.06	30T 478545	:: 4463147					30T 478545	:: 4463147			Pardo 3
4°	40°	26'	39,3"	N	3°	43'	44,8"	W	80T 438168	: 44773240°	26' 43.57"N	3° 43' 40.09"W	N40 26 43.57	W3 43 40.09	30T 438277	:: 4477530					30T 438277	:: 4477530			Cantarranas 0m
5°	40°	26'	39,8"	N	3°	43'	48,6"	W	80T 438079	: 447733940°	26' 44.07"N	3° 43' 43.91"W	N40 26 44.07	W3 43 43.91	30T 438188	:: 4477546					30T 438188	:: 4477546			Cantarranas 100m
6°	40°	26'	39,7"	N	3°	43'	50,8"	W	80T 438028	:: 44773340°	26' 43.97"N	3° 43' 46.06"W	N40 26 43.97	W3 43 46.06	30T 438137	:: 4477543					30T 438137	:: 4477543			Cantarranas 150m
7°	40°	26'	39,7"	N	3°	43'	52,9"	W	80T 437978	: 44773340°	26' 43.90"N	3° 43' 48.17"W	N40 26 43.90	W3 43 48.17	30T 438087	:: 4477541					30T 438087	:: 4477541			Cantarranas 200m
8°	40°	26'	39,7"	N	3°	43'	52,9"	W	80T 437978	: 447733340°	26' 43.92"N	3° 43' 48.19"W	N40 26 43.92	W3 43 48.19	30T 438087	:: 4477542					30T 438087	:: 4477542			Cantarranas 200m 2m D
9°	40°	26'	39,5"	N	3°	43'	52,9"	W	80T 437978	: 44773340°	26' 43.79"N	3° 43' 48.17"W	N40 26 43.79	W3 43 48.17	30T 438087	:: 4477538					30T 438087	:: 4477538			Cantarranas 200m 2m I

Conversión coordenadas geográficas y UTM datum ED50 a datum WGS84																									
n° pun to	ENTRADAS: Latitud				ENTRADAS: Longitud				Salida UTM (ED50)		SALIDA: Coordenadas Geográficas (WGS84)		SALIDA: UTM (WGS84)												
	g°	min'	s"	N/S	g°	min'	s"	E/W	Huso	Este X	Norte Y	g°	min'	s"	N/S	g°	min'	s"	E/W	Huso	Este X	Norte Y			
1°	40°	19'	0,5"	N	3°	15'	21,5"	W	80T 478250	: 446303040°	18' 56.28"N	3° 15' 26.19"W	N40 18 56.28	W3 15 26.19	30T 478140	:: 4462822					30T 478140	:: 4462822			Pardo 1
2°	40°	19'	0,2"	N	3°	15'	17,6"	W	80T 478342	: 44630240°	18' 55.99"N	3° 15' 22.29"W	N40 18 55.99	W3 15 22.29	30T 478233	:: 4462813					30T 478233	:: 4462813			Pardo 2
3°	40°	19'	0,1"	N	3°	15'	13,7"	W	80T 478434	: 44630160°	18' 55.85"N	3° 15' 18.36"W	N40 18 55.85	W3 15 18.36	30T 478325	:: 4462808					30T 478325	:: 4462808			Pardo 2
4°	40°	26'	39,3"	N	3°	43'	44,8"	W	80T 438166	: 447740040°	26' 35.07"N	3° 43' 49.49"W	N40 26 35.07	W3 43 49.49	30T 438057	:: 4477192					30T 438057	:: 4477192			Cantarranas 0m
5°	40°	26'	39,8"	N	3°	43'	48,6"	W	80T 438076	: 44774160°	26' 35.57"N	3° 43' 53.31"W	N40 26 35.57	W3 43 53.31	30T 437967	:: 4477208					30T 437967	:: 4477208			Cantarranas 100m
6°	40°	26'	39,7"	N	3°	43'	50,8"	W	80T 438025	: 447741340°	26' 35.47"N	3° 43' 55.46"W	N40 26 35.47	W3 43 55.46	30T 437916	:: 4477206					30T 437916	:: 4477206			Cantarranas 150m
7°	40°	26'	39,7"	N	3°	43'	52,9"	W	80T 437975	: 44774140°	26' 35.40"N	3° 43' 57.57"W	N40 26 35.40	W3 43 57.57	30T 437866	:: 4477204					30T 437866	:: 4477204			Cantarranas 200m
8°	40°	26'	39,7"	N	3°	43'	52,9"	W	80T 437975	: 447741340°	26' 35.42"N	3° 43' 57.59"W	N40 26 35.42	W3 43 57.59	30T 437866	:: 4477205					30T 437866	:: 4477205			Cantarranas 200m 2m D
9°	40°	26'	39,5"	N	3°	43'	52,9"	W	80T 437975	: 447740840°	26' 35.29"N	3° 43' 57.57"W	N40 26 35.29	W3 43 57.57	30T 437866	:: 4477201					30T 437866	:: 4477201			Cantarranas 200m 2m I

FIGURA 6.2: Tablas que muestran la conversión de coordenadas geográficas y UTM datum WGS84 a datum ED50 (y viceversa)

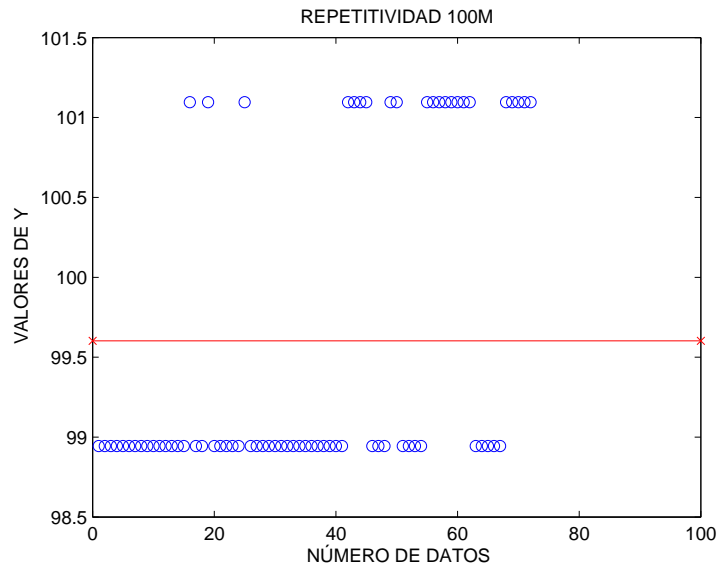


FIGURA 6.3: Repetitividad en los datos obtenidos para 100 metros. En azul y con círculos se muestran los valores obtenidos por el sistema. En rojo se muestra la media de los datos.

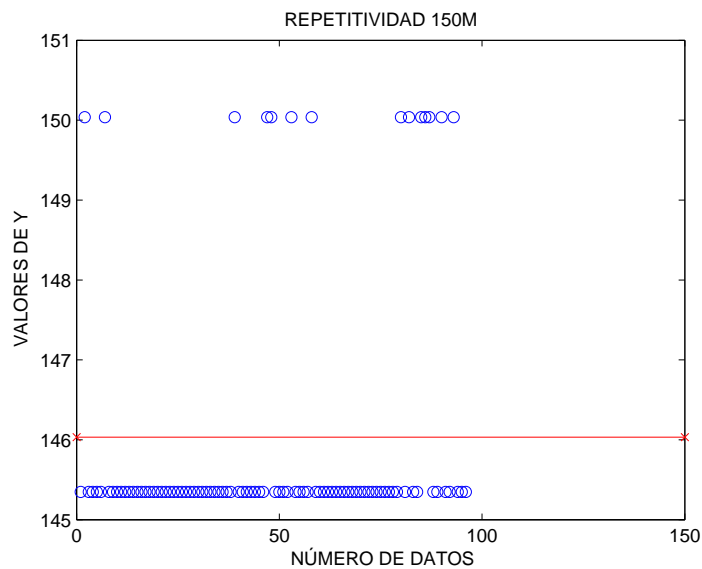


FIGURA 6.4: Repetitividad en los datos obtenidos para 150 metros. En azul y con círculos se muestran los valores obtenidos por el sistema. En rojo se muestra la media de los datos.

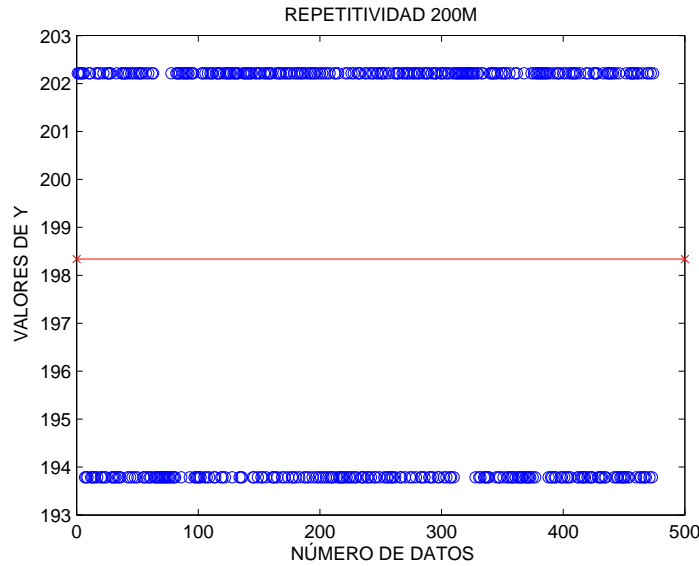


FIGURA 6.5: Repetitividad en los datos obtenidos para 200 metros. En azul y con círculos se muestran los valores obtenidos por el sistema. En rojo se muestra la media de los datos.

tran la situación obtenida. Como se puede apreciar, los resultados en términos de valores calculados y repetitividad son prácticamente idénticos dando lugar a la alternancia de dos valores en toda la muestra.

Por tanto, se puede concluir que la variación en los datos, ya sea a corta o larga distancia depende mucho de la alineación de las cámaras, dando lugar a que el cálculo de las medidas en y se vea afectado de tal manera que siempre existan dos valores distintos pero cercanos entre sí y cuya media es siempre muy fiable a la ideal.

6.4. Error relacionado con la distancia

Otro aspecto fundamental del análisis fue el error que se comete cuanto más nos alejamos del sistema de cámaras, es decir, que error se comete con la distancia. En la figura 6.8, se muestra el efecto de la distancia en la toma de medidas. Como podemos observar, el error aumenta a medida que nos vamos alejando, como es de esperar, de manera que a 20 metros el error es de 0.0784

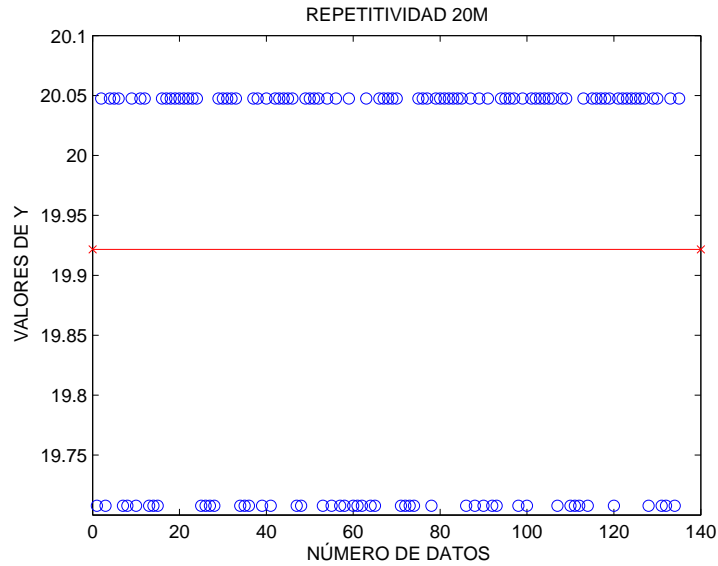


FIGURA 6.6: Repetitividad en los datos obtenidos para 20 metros. En azul y con círculos se muestran los valores obtenidos por el sistema. En rojo se muestra la media de los datos.

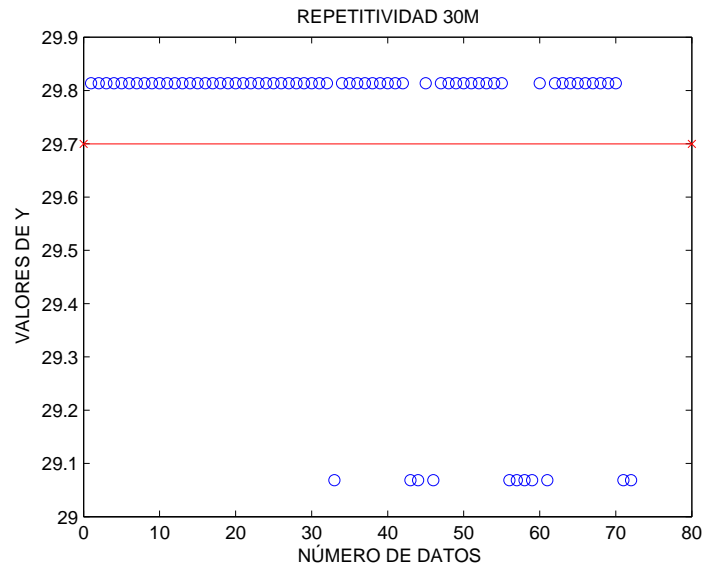


FIGURA 6.7: Repetitividad en los datos obtenidos para 30 metros. En azul y con círculos se muestran los valores obtenidos por el sistema. En rojo se muestra la media de los datos.

y termina con un valor de 4.2039 al cabo de los 200 metros.

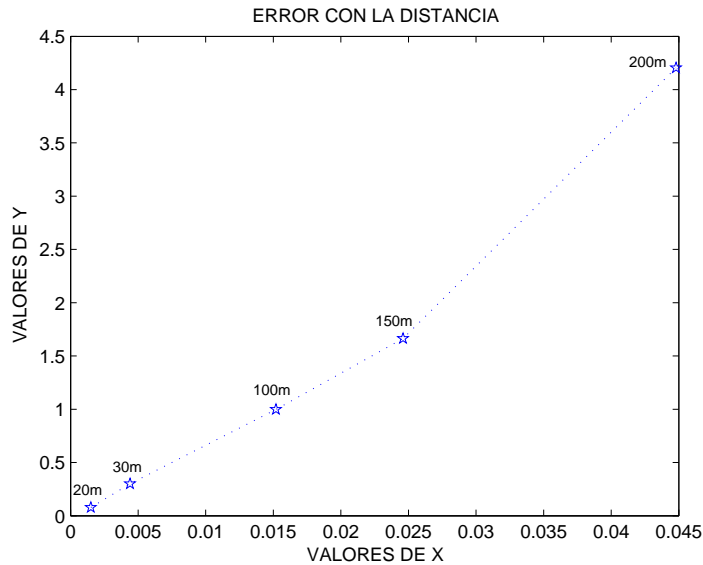


FIGURA 6.8: Efecto de la distancia con respecto a los resultados obtenidos.

6.5. Precisión

Para llevar a cabo unas pruebas de precisión, se hizo un experimento en la facultad a una distancia relativamente pequeña. Para ello, colocamos el objeto a ser identificado de color rojo a una distancia de 30 metros del sistema de cámaras y realizamos varios análisis en determinadas posiciones para intentar medir la resolución lateral del sistema.

Antes de comenzar con el proceso de captura de resultados, tuvimos que alinear las cámaras convenientemente con ayuda de la aplicación creada para este fin. El procedimiento consistió en los siguientes pasos:

1. Tomar 100 medidas al objeto situado a 30 metros y en el centro del sistema estéreo.
2. Desplazar el objeto 40 centímetros a cada lado y medir de nuevo.

3. Repetir sucesivamente el paso anterior acercandonos cada vez más a la posición inicial y estudiar como se comporta el sistema en movimientos relativos en el plano perpendicular al eje de las cámaras.

Los resultados que obtuvimos se muestran en la tabla 6.3 en media. Podemos observar que la distancia y desde el sistema hasta dónde se encuentra el objeto es siempre la misma (aunque difiere en 0.8 metros debido al problema de la alineación comentado en la sección 6.2). Por otro lado, la variación en x es la deseada, puesto que en cada cambio de localización su valor varía exactamente lo necesario.

Distancia	Coordenada x	Coordenada y
30m - 40cm Izquierda	-0,115288	29,8149
30m - 20cm Izquierda	0,084544	29,815
30m - 10cm Izquierda	0,179845	29,0692
30m - 5cm Izquierda	0,230576	29,8143
30m - 3cm Izquierda	0,256376	29,8143
30m - 2cm Izquierda	0,264209	29,8143
30m - 1cm Izquierda	0,274576	29,8143
30m - Centro	0,284377	29,8138
30m - 40cm Derecha	0,684041	29,8073
30m - 20cm Derecha	0,484209	29,812
30m - 10cm Derecha	0,376607	29,8128
30m - 5cm Derecha	0,329716	29,0679
30m - 3cm Derecha	0,31512	29,8135
30m - 2cm Derecha	0,307434	29,8136
30m - 1cm Derecha	0,294748	29,8136

TABLA 6.3: Precisión del sistema en distancias cortas.

En las figuras 6.10 y 6.9 podemos observar gráficamente los resultados obtenidos. Como se ha comentado, la precisión conseguida es de centímetros, por lo que el sistema es capaz de detectar cambios en x e y en el plano ortogonal a las cámaras. Cabe comentar también el hecho de que en dos puntos de las mediciones la distancia en y varíe 80cm con respecto a las otras. Esto es debido fundamentalmente a la alineación de las cámaras que unido al hecho de que

algún tipo de sombra influya en la detección automática del objeto, provoca que el cálculo del centro de gravedad de ese objeto en la imagen difiera en 1 píxel.

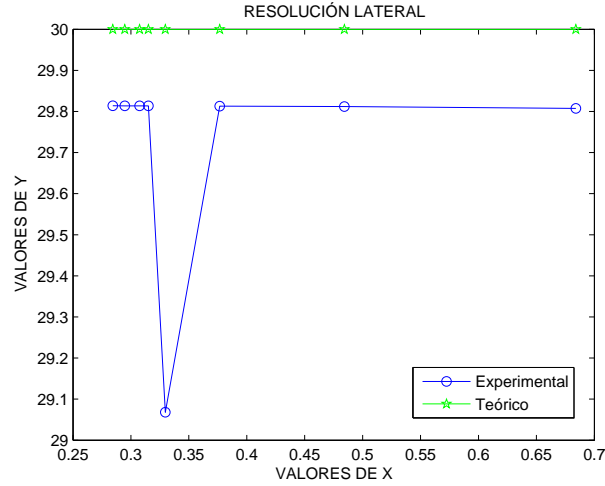


FIGURA 6.9: Gráfica que muestra la precisión del desplazamiento lateral hacia la derecha del objeto en la escena a 30 metros.

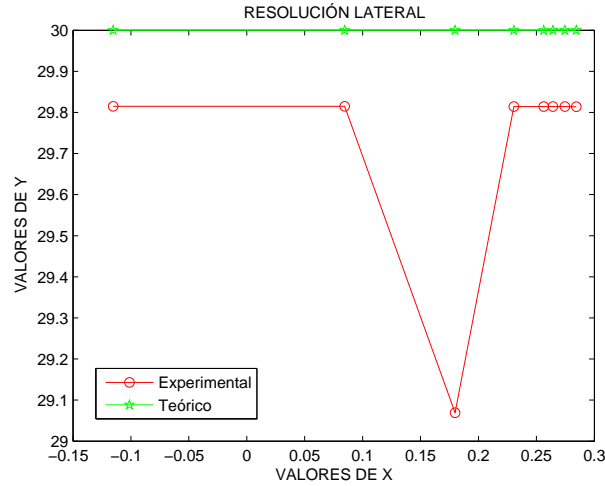


FIGURA 6.10: Gráfica que muestra la precisión del desplazamiento lateral hacia la derecha del objeto en la escena a 30 metros.

Capítulo 7

Conclusiones y trabajo futuro

7.1. Conclusiones

Como hemos estado comentando a lo largo de todo documento, la Visión Artificial es una subcampo de la inteligencia artificial que tiene el propósito de capturar la información visual del entorno físico para extraer características relevantes visuales, utilizando procedimientos automáticos. En definitiva el proceso consiste en capturar una o más imágenes, digitalizarlas e interpretarlas de acuerdo a las especificaciones particulares del dominio. Han sido muchas las aplicaciones relacionadas con esta disciplina a lo largo de estos años, cubriendo gran cantidad de aspectos relacionados con reconstrucciones tridimensionales de las escenas, para que las mediciones y la obtención de distancias sean mucho más precisas, eficientes y flexibles.

Y no es de ninguna manera difícil de extrañar que esto haya sido y siga siendo así, pues las ventajas de estos sistemas son muchas y muy buenas. En primer lugar, la adquisición de las imágenes es rápida, simple y mínimamente invasiva, ya que el único aspecto involucrado es la cámara que captura las imágenes del entorno. No hay que olvidarse tampoco, que en la actualidad, los costes de los componentes (cámaras, procesadores, lentes) se han reducido de forma drástica, mientras que las prestaciones han ido aumentando de manera espectacular. En segundo lugar, destacar que se en los últimos años, se han

ido desarrollado nuevas tecnologías como por ejemplo, la visión multispectral, que permiten ver más allá de lo que es capaz de captar el sistema visual humano. De este modo, ampliando la respuesta de las cámaras a los espectros ultravioleta e infrarrojo, es posible detectar defectos y problemas que, de otro modo, pasarían inadvertidos.

Este trabajo ha intentado cubrir en la manera de lo posible dos de las grandes técnicas de Visión por Computador: la visión *estereoscópica* o *binocular*, y la visión *monocular* o *metrología*. Además, se han utilizando diferentes métodos de procesamiento y análisis de imágenes digitales, que han llegado a ser fundamentales para que la detección automática de los objetos pueda llevarse a cabo de manera no supervisada satisfactoriamente.

Es por ello, que una parte del trabajo se ha centrado en investigar qué métodos, funciones y técnicas eran las más apropiadas para nuestro problema en particular. Los algoritmos tenían que ser válidos para cualquier tipo de condiciones, ser capaces de adaptarse fácilmente a cada tipo de situación, flexibles y eficientes. Y podemos decir que así ha ocurrido, ya que nuestro sistema trabaja realmente bien en todo tipo de condiciones de luz, excepto cuando las imágenes presentan un nivel muy bajo de luminosidad, cercano a la oscuridad absoluta.

En cuanto a la visión *estereoscópica*, hay que destacar los grandes resultados obtenidos. El cálculo de medidas en líneas generales, ha sido bastante satisfactorio tanto en distancias pequeñas como grandes, comparado con el otro sistema de medición utilizado en este trabajo, el sistema de posicionamiento global, que a pesar de tener unas excelentes características no ha llegado a ser tan preciso como nuestra aplicación, en las pruebas realizadas. La única desventaja que se puede comentar acerca de esta técnica, es que es absolutamente necesario un excelente proceso de calibración y alineación de las dos cámaras, para conseguir que los ejes ópticos de las mismas sean totalmente paralelos. Y con la plataforma creada en la actualidad, algo inestable en ocasiones, puede llegar a ser difícil de conseguir.

Por otro lado, la visión *monocular* tiene la gran ventaja de que se puede trabajar con imágenes que no necesitan ningún tipo de calibración. Esto es bastante interesante, puesto que la ardua tarea de alinear y calibrar presente en la visión *binocular* aquí es inexistente. Sin embargo, la búsqueda de unos puntos de referencia adecuados puede llegar a ser complicado pues depende de la ubicación de la escena en el mundo real. Y además, conocer la localización geográfica de esos puntos lleva asociado un error proveniente del uso de cinta métricas o sistemas de posicionamiento global. Aún así, los resultados extraídos han sido muy acertados.

Por último mencionar algunos aspectos extraídos de los resultados experimentales llevados a cabo. En primer lugar, la precisión de centímetros obtenida es más que suficiente para los objetivos de este proyecto. El sistema es por tanto, capaz de detectar variaciones en la componente horizontal de manera precisa sin ningún tipo de problema. En segundo lugar, el error cometido con la distancia aumentará como es lógico al alejarnos del sistema de referencia inicial y es una tarea que habría que intentar solucionar, aunque no se trata de un problema crítico. Este proyecto seguirá ampliándose con nuevas funcionalidades y mejoras de cara a los próximos años por lo que su continuidad está asegurada.

7.2. Trabajo futuro

El trabajo que se puede realizar tomando como base este proyecto es muy interesante y de gran utilidad para el campo de la Visión por Computador.

En primer lugar, una posibilidad es añadir una nueva funcionalidad que permita detectar en la imagen un segundo objeto. Este objeto podría ser del mismo o diferente color al que siempre hemos utilizado. Estos nuevos métodos no supondrían un gran trabajo, puesto que todo está implementado de manera que la incorporación de nuevas funcionalidades no sea una tarea complicada. Y este hecho, supondría en realidad una gran ventaja, pues se podrían calcular

medidas a dos objetos móviles entre sí y por tanto, la distancia existente entre ellos mismos. De este modo, se puede llevar un control mucho más preciso y eficaz de cara a futuros usos en el control y guiado de vehículos marinos y aéreos.

En segundo lugar, y tal como se ha comentado en las conclusiones, se podría mejorar de alguna manera el proceso de alineación de las cámaras, para que no sea una tarea tan supervisada por el usuario. Por dar algún tipo de solución, podría considerarse el uso de una plataforma automatizada que permitiese la rotación de las cámaras de manera fácil. Conectada al computador, éste le enviaría mensajes para interpretar los movimientos adecuados. Probablemente, esta futura aplicación consiga eliminar el problema existente en la actualidad, la diferencia de como mucho un píxel entre la disparidad teórica y la experimental, con el consiguiente aumento de precisión en los cálculos de las distancias.

En tercer lugar, otra posible mejora del sistema incluiría una optimización de los algoritmos de procesamiento y análisis. Actualmente se pueden capturar del orden de dos imágenes por segundo, cifra aceptable pero mejorable. Para ello, podría portarse el programa a GPU en lugar de CPU, y tratar de paralelizar lo más posible para conseguir unos buenos resultados y optimizaciones. Esa es una buena posibilidad que habría que estudiar detenidamente. Otra solución válida sería intentar optimizar aún más los algoritmos utilizados para el filtrado y binarización, así como para la extracción de regiones.

Por último, la integración de este proyecto con sistemas de control y cooperación supondría notables mejoras. Bastaría con estudiar, las incompatibilidades del código, y homogeneizar los algoritmos para que se engloben por completo en estos sistemas.

Apéndice A

Instalación y uso

Este apéndice trata acerca de la instalación y uso de la aplicación. Se ha intentado que dichas tareas sean lo más intuitivas posible para que el usuario no tenga que referirse a este apéndice. De todas formas se incluye para ayudar a quien encuentre dificultades.

A.1. Instalación

La instalación del programa es una tarea muy sencilla: únicamente basta con descomprimir el archivo comprimido con las fuentes, y ejecutar el fichero `halcon.exe`. Antes de este paso es necesario disponer en el sistema de los siguientes componentes:

- Ordenador con Windows XP
- Dos cámaras con interfaz IEEE 1394b (en nuestro caso marca Basler).
- Dos puertos IEEE 1394b.
- Matrox Imaging Library. Actualmente la última versión es la 8.0, con licencia completa. Este paquete, ya incluye todas las herramientas necesarias, así como los drivers necesarios para las cámaras utilizadas.
- Visual Studio 2003 o superior. Necesario para compilación y depuración en caso de utilizarlo.

Una vez que toda la aplicación ha sido descomprimida en la ubicación elegida por el usuario, se conecta el sistema de cámaras, teniendo especial atención en controlar que las dos cámaras no parapadean y por tanto, asegurarse que han sido reconocidas perfectamente.

A.2. Uso

El uso de la aplicación también es bastante sencillo. En la figura A.1 se muestra una captura de la pantalla inicial de la aplicación. Se puede ver, que el menú situado en la parte superior dispone de diferentes opciones:

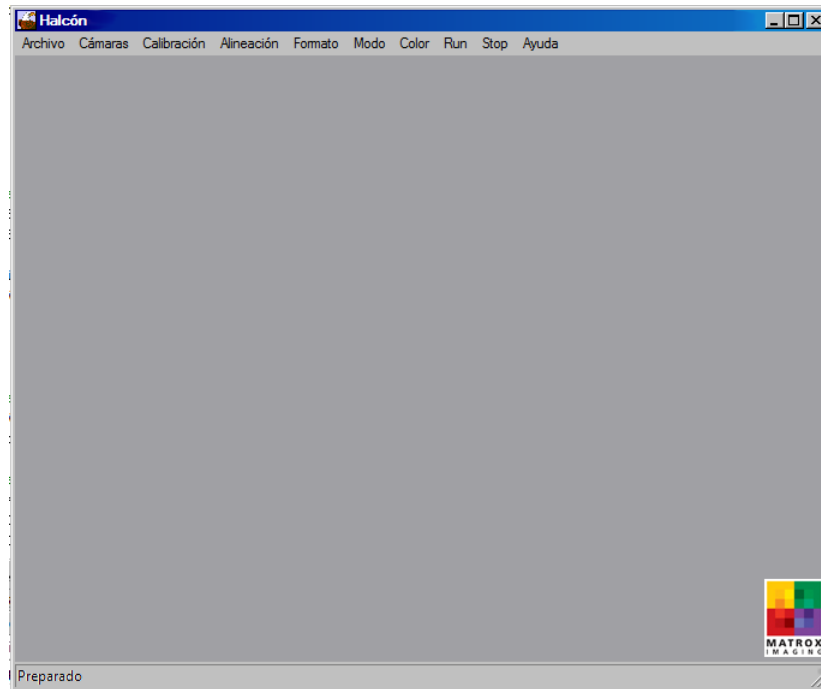


FIGURA A.1: Ventana principal de Halcón.

- **Archivo.** En este menú se dan las opciones de ajustar los parámetros del programa, así como guardar los resultados. También dispone de una opción que permite salir de la aplicación.
- **Cámaras.** El submenú permite visualizar cada una de las cámaras por separado, para poder enfocarlas y ajustar el diafragma lo mejor posible.

Tras esta acción se deben detener pulsando **Stop**.

- **Calibración**. Permite asignar los puntos de referencia necesarios para calibrar la imagen en el método monocular.
- **Alineación**. La alineación es necesaria para que el cálculo de medidas en estéreo sea óptimo. Solicita la distancia a la que se encuentra el objeto y tras aceptar el cuadro de diálogo, en la barra de estado inferior, se indica la acción a realizar para ajustar la cámara izquierda.
- **Formato**. Permite seleccionar entre los dos tipos de formato de los datos: LLA (longitud y latitud) y PRC (x e y).
- **Modo**. Se dan dos opciones en este submenú. **Depuración** muestra el análisis de cada una de las imágenes divididas en pasos, para poder monitorizar la localización del objeto en la imagen. **Medida** elimina la parte gráfica y solo se centra en los resultados, por lo que es más rápido que la opción anterior.
- **Color**. Ajusta la intensidad de rojo y de verde para que la extracción del objeto en la imagen sea la mejor posible.
- **Run**. Ejecuta el cálculo de medidas. Se divide en dos submenús: **Run Estéreo** computa resultados para el sistema en estéreo. Por el contrario **Run Mono**, hace lo propio para el sistema monocular.
- **Stop**. Para el cálculo de resultados. Esta acción puede realizarse en cualquier momento.
- **Ayuda**. Ayuda disponible sobre la aplicación y sus creadores.

Ahora se va a explicar como se lleva a cabo el procedimiento para capturar imágenes y el cálculo de distancias en **estéreo** (el primer requisito es que las dos cámaras estén conectadas adecuadamente).

1. Ajustar los parámetros generales del sistema (en caso de ser necesario), pulsando **Archivo** → **Parámetros**. En la figura A.2 se muestra un ejemplo de este proceso.

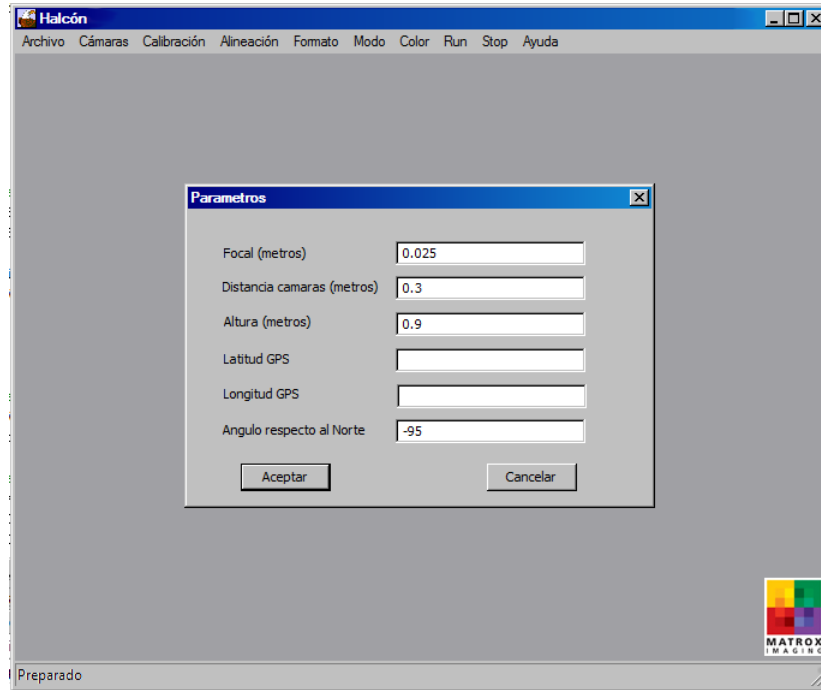


FIGURA A.2: Diálogo que permite modificar los parámetros de Halcón.

2. Comprobar el estado de las cámaras. En el menú **Cámaras** → **Cámara derecha** o **Cámaras** → **Cámara izquierda** se controla el enfoque y la luminosidad de cada una de ellas. Tras este paso, se debe detener la captura mediante **Cámaras** → **Stop**.
3. Si las cámaras no están alineadas, seleccionar la opción **Alineación** . El sistema solicitará la distancia a la que se encuentra el objeto como se muestra en la figura A.3.
4. Mientras se está ejecutando la tarea de alineación se recomienda cambiar el modo de representación a depuración haciendo clic en **Modo** → **Depuración** . De esta manera, podemos averiguar que está sucediendo en cada momento, comprobar si se está aislando bien el objeto o no, y si existe mucho ruido en la imagen.
5. En estos momentos, pueden ocurrir dos cosas, que se esté localizando el objeto correctamente o que no. Si estamos en la segunda situación,

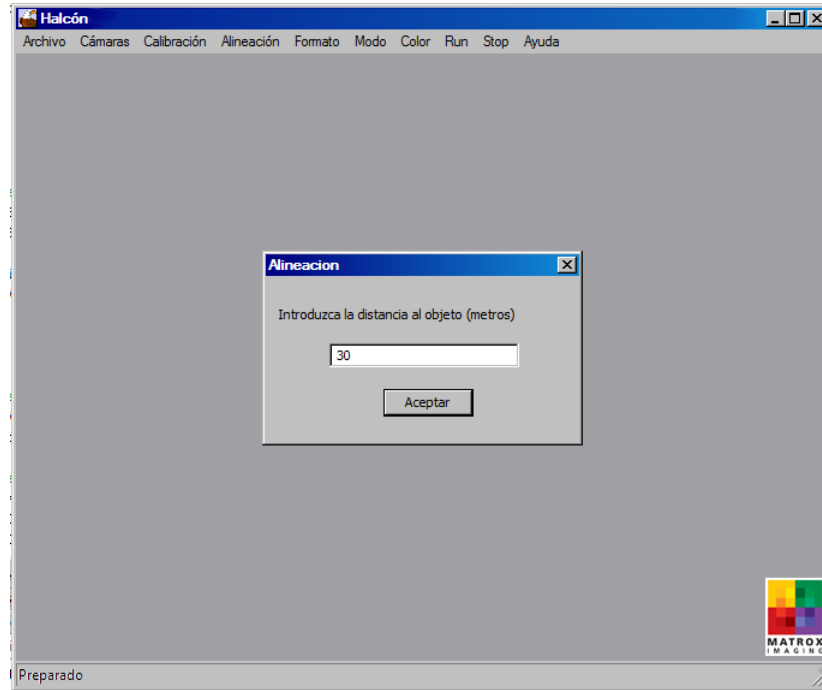


FIGURA A.3: Ventana que muestra el proceso de introducción de la distancia en la alineación.

saltar este paso. Si por el contrario, nos encontramos en la primera, lo que se debe hacer es ajustar las intensidades de rojo y verde mediante la opción **Color** , como se observa en la figura A.4.

Para ello, se debe bajar o subir la intensidad del rojo en función de si capta o no el objeto. Lo mismo se debe hacer con la intensidad del verde, teniendo en cuenta que cuanto más bajo sea su valor más impurezas de la imagen se eliminarán. Se recomienda realizar este paso, hasta que la identificación sea perfecta.

6. Tras haber conseguido aislar el objeto perfectamente, volvemos a seleccionar el modo medida mediante **Modo** → **Medida** . Si nos fijamos ahora, en la parte inferior de la ventana (en la barra de estado), aparecen una serie de notificaciones que indican la disparidad actual y la acción a realizar sobre las cabezas micrométricas de las cámaras. El mensaje será del estilo **Disparidad: 2.48 píxeles - Gire hacia la izquierda** (en el sentido contrario a las agujas del reloj). De esta manera, habrá

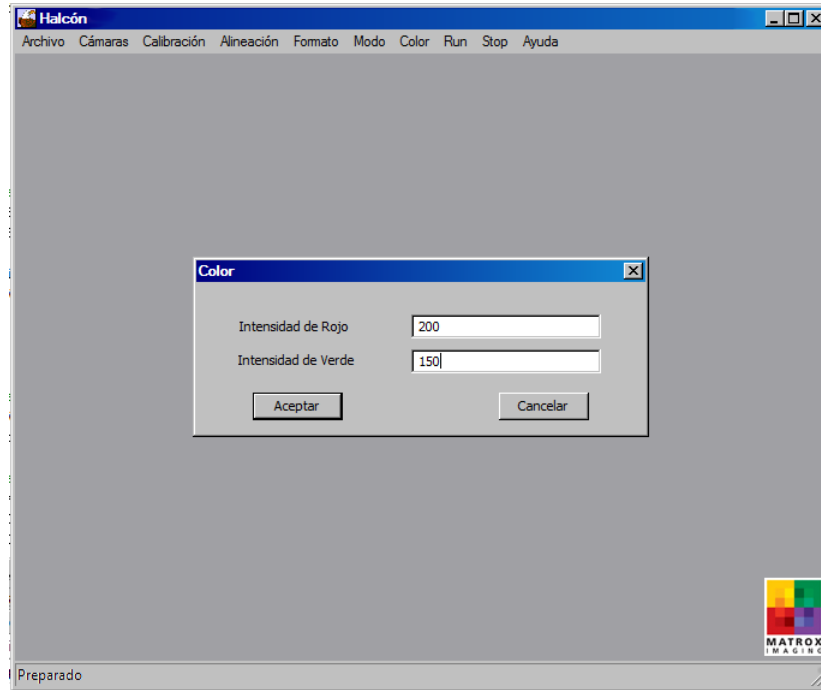


FIGURA A.4: Diálogo que muestra la utilización de las intensidades de color.

que ir ajustando hasta que la disparidad mostrada sea 0 o lo más próximo a 0.

7. Una vez realizados los pasos más difíciles, lo siguiente no es nada complicado. Basta con ejecutar `Run` → `Run Estéreo` y las medidas irán apareciendo en pantalla como se muestra en la figura A.5.
8. Por último, tras calcular varias medidas, se para el cómputo mediante `Stop` o `Aceptar`, y se guardan los resultados si así se desea haciendo clic en `Archivo` → `Guardar`.

El procedimiento a seguir para el cálculo en visión **monocular** es bastante más sencillo. Los pasos de depuración y ajuste de color son idénticos a la visión binocular. Los pasos a seguir a continuación son los siguientes:

1. Calibrar la imagen. Pulsando `Calibración` → `Establecer puntos de referencia`, se cargará en pantalla la imagen correspondiente a la cá-

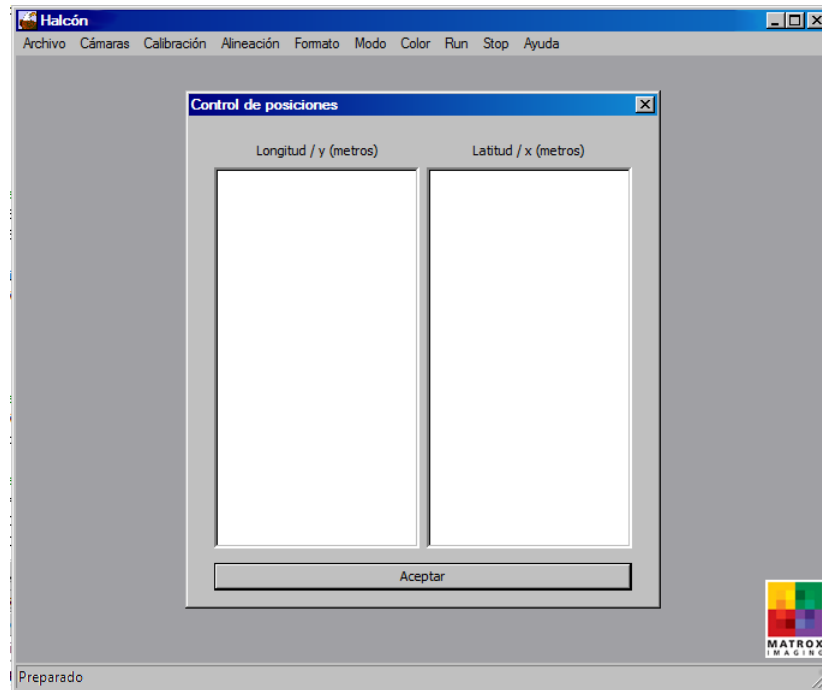


FIGURA A.5: Diálogo que muestra los resultados y medidas obtenidas.

para izquierda para poder establecer los puntos de alineación. Un ejemplo puede verse en la figura A.6.

Tras la carga de la imagen, hay que seleccionar cuatro puntos de la imagen. Para ello, basta con hacer clic en cualquier píxel de la imagen y saldrá un pequeño formulario donde introducir los valores de x e y reales, tal y como se muestra en la imagen A.7.

2. Una vez seleccionados los puntos, solamente queda ir a `Run` → `Run Mono` y las medidas irán apareciendo en pantalla como en la situación anterior. El resto de pasos son exactamente los mismos tanto para parar el cómputo como para guardar los resultados obtenidos.

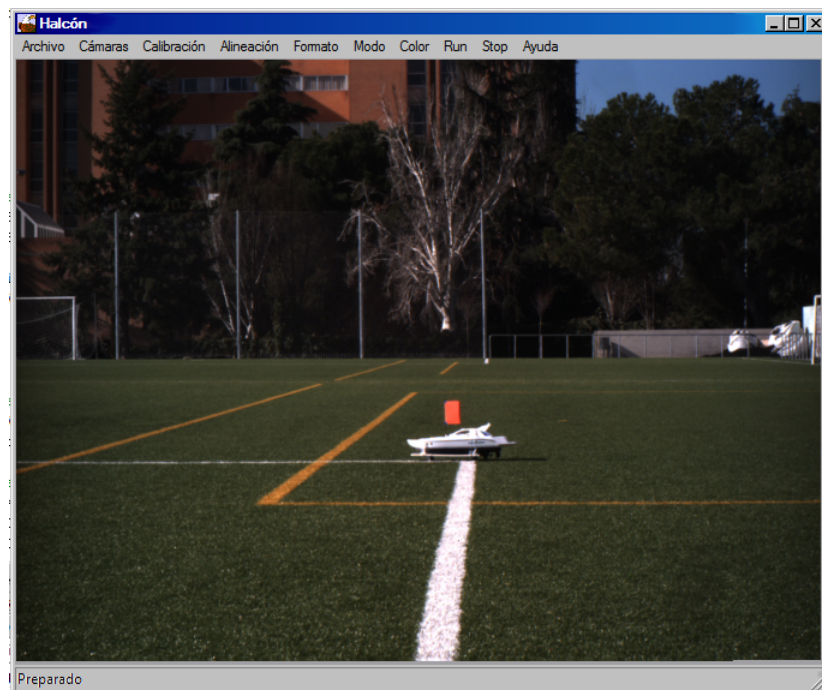


FIGURA A.6: Calibración de las cámaras, selección de puntos.

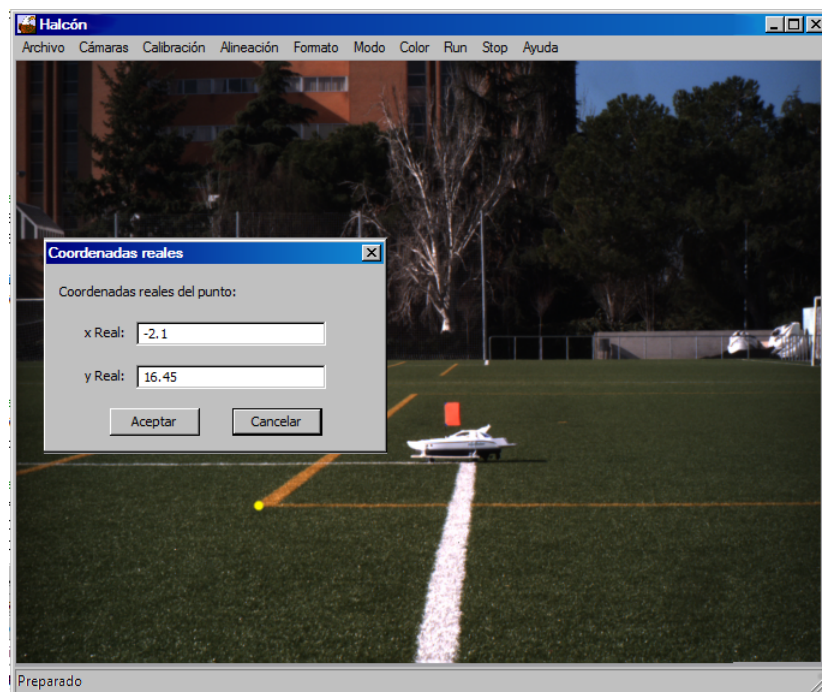


FIGURA A.7: Calibración de las cámaras, introducción de coordenadas.

Apéndice B

Matrox Imaging Library

Como se ha comentado en capítulos anteriores, *Matrox Imaging Library* (MIL) es una librería de 32bits modular e independiente del hardware. Es por esa razón que nosotros hemos utilizado gran parte de su potencia para implementar el sistema.

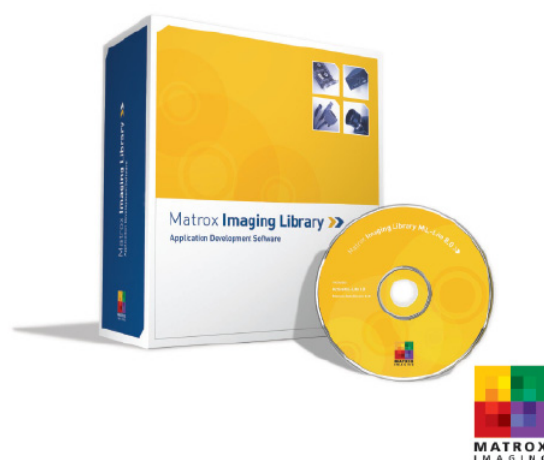


FIGURA B.1: Paquete de librerías *Matrox Imaging Library*.

Las librerías MIL son un conjunto de funciones de alto nivel para captura y procesamiento de imágenes (Ej- operaciones punto a punto, estadísticas, filtrados, morfología, transformaciones geométricas, FFT y segmentación), reconocimiento de patrones, análisis de blobs, extracción y análisis de bordes,

lectura de caracteres, lectura de códigos 1D y 2D, calibración, gráficos, compresión, display y almacenamiento de imágenes.

Diseñadas para facilitar el desarrollo y aumentar la productividad, MIL ofrecen un típico API en C que soporta toda la oferta hardware de Matrox, y un conjunto de funciones intuitivas y fáciles de usar. MIL también incluyen interfaces para manejar I/O, ajuste de parámetros y gestión de resultados, todo enfocado a simplificar el desarrollo de la aplicación.

Para agilizar el desarrollo de la aplicación en Windows, MIL incluye ActiveMIL, un conjunto de controles ActiveX (OCX) para captura de imagen, procesamiento, análisis, display y almacenamiento. ActiveMIL permite combinar fácilmente la aplicación de visión artificial con un interfaz de usuario Windows. En este caso el desarrollo de la aplicación consiste en configurar las herramientas necesarias con el uso del ratón, teniendo que escribir mucho menos código. Con ActiveMIL los OEMs e integradores de visión reducen considerablemente el tiempo de desarrollo, centrándose en la aplicación de visión en vez de en el interfaz de usuario.

Algunas de sus características son:

- Compresión /descompresión. Comprime y descomprime imágenes color y monocromo usando el estándar JPEG o JPEG2000.
- Medida. Localiza marcas de referencia de forma automática, y encuentra múltiples marcas mediante una única función de llamada.
- Análisis de blobs. Analiza blobs (grupos de pixels relacionados por una característica) en una imagen.
- Edge Finder. Poderosa herramienta de extracción de bordes para detección de defectos, análisis y reconocimiento de formas.
- Geometric Model Finder. Usa características geométricas (contornos) para guiado de robots, inspección de obleas y muchas aplicaciones más.

Utiliza un algoritmo único que proporciona una gran flexibilidad con niveles sin precedentes de robustez y precisión.

- **Pattern Matching.** Utiliza correlación normalizada en escala de grises (NGC) para resolver aplicaciones de alineamiento, medida e inspección. El algoritmo usado maneja imágenes degradadas, ruidosas o desenfocadas, y utiliza una estrategia de búsqueda inteligente para reducir el tiempo de proceso.
- **Metrología.** Mide y construye distancias geométricas en función de unas tolerancias definidas. Se pueden medir arcos, círculos, segmentos, puntos, etc... Realiza mediciones de alta precisión con gran flexibilidad a la hora de definir las distancias y sus tolerancias.
- **Calibración.** Compensa imágenes, posiciones y medidas para la representación espacial del campo de visión de una cámara o de objetos en una imagen usando interpolación lineal.
- **String reader.** Sofisticado reconocimiento de cadenas de caracteres basado en características geométricas. Especialmente pensado para aplicaciones de Lectura de Matrículas. Dispone de un interfaz gráfico de usuario (GUI) para crear el contexto más adecuado en cada caso de manera muy sencilla.
- **OCR.** Reconocimiento de caracteres ideal para leer o verificar caracteres mecánicamente impresos. Está optimizado para leer caracteres predefinidos, pudiendo leer también fuentes definidas por el usuario.

Una visión general de la arquitectura puede verse en la figura B.2. Para cualquier referencia, al uso de las funciones y comandos de las funciones utilizadas en el proyecto, refiérase al fichero `Mil User Guide.chm` entregado en el CD junto con el resto de la documentación.

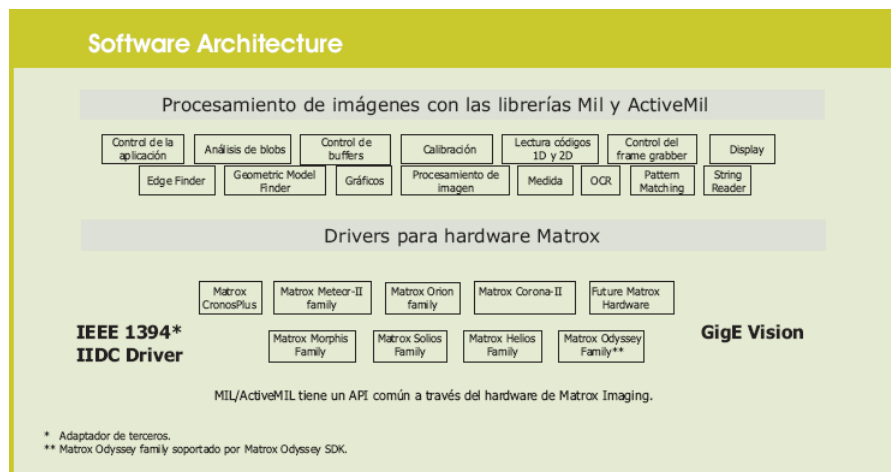


FIGURA B.2: Arquitectura software de MIL.

Apéndice C

MIDG II Display

El MIDG II Display presenta el aspecto que puede verse en la figura C.1. Esta es la pantalla llamada *Nav/GPS*, la cual está dividida en siete áreas, consistentes en un Horizonte Artificial y seis Ventanas de Mensajes. Cada ventana de mensaje presentan valores con la intención de que sean familiares para el usuario.

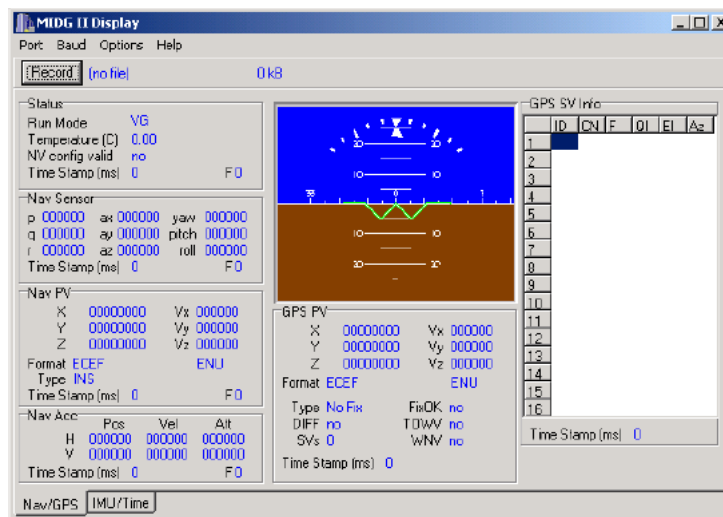


FIGURA C.1: Ventana principal MIDG II.

C.1. Horizonte Artificial

El Horizonte Artificial es una representación gráfica del posicionamiento del MIDG. Muestra lo siguiente y puede verse en la figura C.2:

- Pendiente. El área del cielo se incrementa si el MIDG es dirigido hacia arriba mientras que el suelo se incrementa si el MIDG es dirigido hacia abajo.
- Balanceo. Se muestra en la escala curva en lo alto de la ventana, en incrementos de 10° .
- Viraje. Está enumerado en el horizonte. Cada barra grande representa un incremento de 10° .

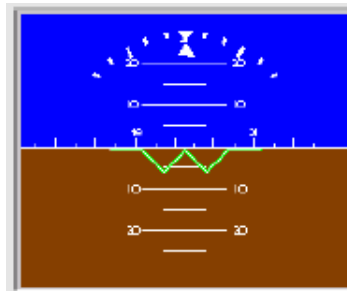


FIGURA C.2: Cuadro del horizonte artificial.

C.2. Ventana de Estado

Esta ventana muestra datos del estado del sistema, cosas como el modo de operación, la temperatura interna en $^\circ\text{C}$, etc.

C.3. Ventana de Datos del Sensor de Navegación

Muestra los índices angulares (p , q y r), los cuales son mostrados en grados por segundo. También muestra las aceleraciones (ax , ay , az) en mili-g. Finalmente muestra el viraje, la pendiente y el balanceo (yaw , $pitch$ y $roll$) en ángulos.

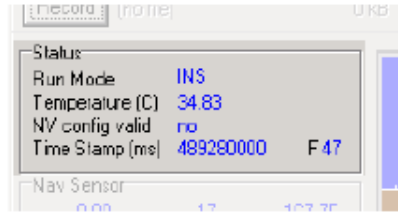


FIGURA C.3: Ventana de estado.



FIGURA C.4: Ventana de datos del sensor de navegación.

C.4. Ventana de la Posición de Navegación y Velocidad

La posición se muestra en metros (coordenadas ECEF o ENU) o en grados (coordenadas geográficas), mientras que la velocidad se muestra en metros por segundo.



FIGURA C.5: Ventana de la posición de navegación.

C.5. Ventana de Estimación de la Precisión de Navegación

Muestra la precisión estimada para la posición, velocidad y posicionamiento del MIDG en las direcciones vertical y horizontal. Representan la desviación estándar del error del parámetro asociado. La precisión de la posición se muestra en metros, la de la velocidad en metros por segundo y el del posicionamiento en grados.

Nav Acc	Pos	Vel	Att
H	0.44	0.10	0.51
V	0.44	0.04	8.19

Time Stamp (ms) 489280200 FC3

Time Stamp (ms) 489280000 FC0

Nav/GPS IMU/Time

FIGURA C.6: Ventana de estimación de la precisión.

C.6. Ventana de posición y velocidad dada por el GPS

Son datos que provienen directamente del GPS. Incluye posición, velocidad e información relevante. La posición se puede mostrar en metros (ECEF o ENU) o en grados (LLA), y la velocidad en metros por segundo.

X	1198573.82	Ve	0.00
Y	-4950364.52	Vn	0.00
Z	3825319.63	Vu	0.02
Format	ECEF	ENU	
Type	3D Fix	FixOK	yes
DIT	no	DUMV	yes
SVs	8	WNV	yes
Time Stamp (ms)	489280001		

10
11
12
13
14
15
16
Tim

FIGURA C.7: Ventana de posición y velocidad.

C.7. Ventana de información de los satélites del GPS

Muestra información del estado de cada satélite, incluyendo el número del satélite que está siendo recibido por cada canal, el cociente portadora/ruido, un indicador de la calidad de los datos y la elevación y el azimut de cada satélite en grados. El sistema proporciona otros datos, como medidas de sensores inerciales, medidas del valor del campo magnético, medidas del tiempo, etc... pero esa información no es relevante para nosotros.

	ID	C/N	F	D1	E1	Az
1	1	48	00	07	76	121
2	3	36	00	07	11	175
3	5	39	00	07	20	56
4	122	41	00	07	41	146
5	120	0	00	01	-9	0
6	27	0	00	01	-9	0
7	14	36	00	07	15	134
8	16	46	00	07	78	206
9	20	44	00	07	38	264
10	28	0	00	01	-9	0
11	121	0	00	01	-9	0
12	25	47	00	07	57	51
13	4	0	00	01	-9	0
14	13	0	00	01	-9	0
15	23	45	00	07	35	314
16	17	0	00	01	-9	0

Time Stamp (ms) 489280001

FIGURA C.8: Ventana de información de los satélites.

Bibliografía

- [ACZ99] I. Reid A. Zisserman A. Criminisi. Single view metrology. *International Conference on Computer Vision*, pages 434,442, September 1999.
- [ACZ99] I. Reid A. Criminisi and A. Zisserman. A plane measuring device. *Image and Vision Computing*, 17:625,634, 1999.
- [Bas] Basler home page. <http://www.baslerweb.com/>.
- [Cri99] Antonio Criminisi. *Accurate Visual Metrology from Single and Multiple Uncalibrated Images*. PhD thesis, Robotics Research Group. Department of Engineering Science. University of Oxford, 1999.
- [DL98] A. Zisserman D. Liebowitz. A metric rectification for perspective images of planes. *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 482,488, June 1998.
- [GPM] Jesús Manuel de la Cruz Gonzalo Pajares Martinsanz. On combining support vector machines and simulated annealing in stereovision matching. *IEEE Transactions on systems, man, and cybernetics*, 34, August.
- [GPM06] Jesús Manuel de la Cruz Gonzalo Pajares Martinsanz. Fuzzy cognitive maps for stereovision matching. *Pattern Recognition*, 39:2101,2114, April 2006.

- [GW05] Z.Y. Hu G.H. Wang, H.T. Tsui. Reconstruction of structured scenes from two uncalibrated images. *Pattern Recognition Letters*, 26:207,220, 2005.
- [GWW04] Z.Y. Hu G.H. Wang and F.C. Wu. Single view based measurement on space planes. *Journal of Computer Science and Technology*, 19:374,382, 2004.
- [GWW05] Z.Y. Hu G.H. Wang, H.T. Tsui and F.C. Wu. Camera calibration and 3d reconstruction from a single view based on scene constraints. *Image and Vision Computing*, 23:311,323, 2005.
- [IR96] A. Zisserman I. Reid. Goal-directed video metrology. *Proceedings of European Conference of Computer Vision, Cambridge*, pages 647,658, April 1996.
- [Mar82] David Marr. Vision: a computational investigation into the human representation and processing of visual information. *W.H. Freeman and Company*, New York 1982.
- [Mat] Matrox imaging home page. <http://www.cgl.ucsf.edu/chimera/>.
- [MdlC08] Gonzalo Pajares Martinsanz and Jesús Manuel de la Cruz. *Visión por Computador: Imágenes digitales y aplicaciones*. Ra-Ma, 2008.
- [Mor77] H. P. Moravec. Visual mapping by a robot rover. *Proc. 5th Joint International Conference of Artificial Intelligence(Tokyo,Japan)*, pages 598,600, August 1977.
- [MW01] P. Sturm M. Wilczkowiak, E. Boyer. Camera calibration and 3d reconstruction from single images using parallelepipeds. *Proceedings of International Conference on Computer Vision, Vancouver, Canada*, pages 142,148, July 2001.
- [PG00] R. Payrissat P. Gurdjos. About conditions for recovering the metric structure of perpendicular planes from the single ground plane to image homography. *Proceedings of International Conference on*

- Pattern Recognition, Barcelona, Spain*, pages 358,361, September 2000.
- [RH00] A. Zisserman R. Hartley. Multiple view geometry in computer vision. *Cambridge University Press, Cambridge*, 2000.
- [TK98] K. Hong T. Kim, Y. Seo. Physics-based 3d position analysis of a soccer ball from monocular image sequences. *Proceedings of International Conference on Computer Vision, Bombay, India*, pages 721,726, January 1998.
- [vdH98] F.A. van den Heuvel. 3d reconstruction from a single image using geometric constraints. *ISPRS Journal of Photogrammetry & Remote Sensing*, 53:354,368, 1998.