

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

SISTEMA EN TIEMPO REAL DE DETECCIÓN Y SEGUIMIENTO DE OBJETOS ESFÉRICOS EN VIDEOS DIGITALES

Tesis para optar el Título de Ingeniero Electrónico, que presenta el bachiller:

José Elvis Tafur Tafur

ASESOR: Paul Rodriguez Valderrama

Lima, Diciembre del 2015

Dedicatoria

A mi madre María Adela Tafur Villa a quién admiro, respeto y amo.

A mis hermanos Kelly y Erick quienes siempre están a mi lado.

A mi padre Guzmán, quién desde el cielo siempre me ha acompañado.

A mis tíos, primos y familia. En especial a Vicente, Juana y Olga quienes sin ellos nada hubiese sido posible.

A mis amigos del colegio, universidad y trabajo quienes me acompañaron y animaron en este camino: Christiam, Christopher, Chacón, Marco, Cristian, Dino, Kenny, Karina, Zoila, Claudia, Silvia, Javier y tantas otras personas. Gracias.

Finalmente a Dios, él siempre está conmigo.

Resumen

En la actualidad, el desarrollo de sistemas que permitan la interacción ser humano - computador es un tema que es ampliamente abordado en diversas campos. Dispositivos tales como el Iphone, Ipad, la Samsung Galaxy Tab, entre otros, poseen aplicaciones de uso sencillo e intuitivo, brindando al usuario una experiencia de mayor realismo. Algunas de las aplicaciones más exitosas se dan en el sector del entretenimiento: El XBox 360, Nintendo Wii, Play Station 3 y 4 con sus diversos accesorios (cámaras, controladores, etc.) ofrecen juegos que interactúan con el ser humano en tiempo real.

En medio de este contexto y conociendo del interés que existe por la investigación y desarrollo de sistemas en visión por computador, en la presente tesis se muestra el desarrollo de un sistema que permite la interacción ser humano - computador. Este sistema permite la detección y seguimiento de un objeto esférico en tiempo real, esto a través de una interfaz en la cual el usuario puede percibir el seguimiento de una esfera y cuyos movimientos son captados por una cámara de video. El planteamiento de la solución está inspirado en el sistema Play Station Move, el cual permite el seguimiento de una esfera ubicada en un controlador de movimiento, brindando al usuario una mayor sensación de realismo y control de las acciones realizadas. No obstante, y a diferencia del sistema desarrollado por Play Station, el presente estudio no hace uso de acelerómetros para el seguimiento planteado.

El sistema a desarrollar opera bajo condiciones de precisión y tiempo de procesamiento que permite una interacción aceptable con el usuario, y se centra en el desarrollo e implementación de algoritmos matemáticos (en lenguaje C) en tiempo real que permiten la detección y seguimiento del objeto esférico. Para la interfaz computador - cámara de video se utiliza una librería ya existente, previamente desarrollada en el laboratorio de DSP, la cual a su vez está basada en la librería ffmpeg.

Índice general

1. VISIÓN POR COMPUTADOR: APLICACIONES EN VIDEO JUEGOS	2
1.1. Visión por computador	2
1.2. Video juegos: Análisis del contexto	3
1.2.1. Microsoft XBox 360 - Kinect	4
1.2.2. Nintendo Wii - Wii Motion Plus	4
1.2.3. Play Station 3 - Play Station Move	5
1.2.4. Competitividad en el mercado	6
1.3. Principio de funcionamiento del Play Station 3	7
1.3.1. Calibración	8
1.3.2. Ejecución de la aplicación	8
1.4. Síntesis de la situación actual	8
2. MÉTODOS DE SEGMENTACIÓN Y SEGUIMIENTO	10
2.1. Definiciones generales	11
2.1.1. Video	11
2.1.2. Imagen	12
2.2. Espacios de color	12
2.3. Segmentación de imágenes	13
2.4. Seguimiento de objetos	15
2.5. Síntesis de la investigación	16
3. DETECCIÓN Y SEGUIMIENTO DE UN OBJETO ESFÉRICO	17
3.1. Descripción del sistema	18
3.1.1. Modelo teórico del sistema	18
3.1.2. Lógica de la programación	20
3.2. Adquisición de la imagen	23
3.3. Calibración	23
3.3.1. Región local de análisis	25
3.3.2. Criterios de calibración	29
3.3.3. Situaciones presentes en la calibración del sistema	32
3.4. Segmentación automática	35
3.4.1. Identificación del plano a trabajar	35

3.4.2.	Segmentación automática de la esfera	35
3.4.3.	Segmentación de la imagen en la región global de análisis	36
3.5.	Seguimiento del objeto	41
3.5.1.	Extracción de bordes	41
3.5.2.	Estimación del centro y radio de la región segmentada	41
4.	EVALUACIÓN DEL SISTEMA DE DETECCIÓN Y SEGUIMIENTO	44
4.1.	Descripción de las aplicaciones desarrolladas	44
4.1.1.	Aplicación 1: Sobreposición de una circunferencia	44
4.1.2.	Aplicación 2: Trayectoria del centro de la circunferencia	45
4.1.3.	Aplicación 3: Sobreposición del logo de la PUCP	45
4.1.4.	Aplicación 4: Seguimiento de la trayectoria de la circunferencia	45
4.2.	Consideraciones de las pruebas realizadas	46
4.2.1.	Montaje de los dispositivos para realización de las pruebas	46
4.2.2.	Condiciones bajo las cuales se realizaron las pruebas	47
4.2.3.	Metodología de las pruebas realizadas	48
4.3.	Análisis de exactitud del sistema	49
4.3.1.	Profundidad constante - Movimiento en el eje “X” y eje “Y” constante	49
4.3.2.	Profundidad variable - Ejes “X” y “Y” constantes	55
4.4.	Análisis de la velocidad del movimiento	60
4.5.	Tiempo de procesamiento de las aplicaciones desarrolladas	63
4.5.1.	Aplicación 1 - Análisis de Tiempo	64
4.5.2.	Aplicación 2 - Análisis de Tiempo	65
4.5.3.	Aplicación 3 - Análisis de Tiempo	66
4.5.4.	Aplicación 4 - Análisis de Tiempo	67
4.6.	Análisis económico	68
4.7.	Discusión de resultados	68
4.7.1.	Acerca de la exactitud	69
4.7.2.	Acerca del tiempo de respuesta	69
4.7.3.	Acerca de las condiciones de funcionamiento	69

Índice de figuras

1.1. Aplicación industrial de un sistema de visión por computador	2
1.2. Kinect del XBox 360	4
1.3. Controlador de nintendo Wii	5
1.4. Move controller motion del Play Station 3	6
1.5. Cuadro comparativo de ventas de consolas de video juegos	7
2.1. Proceso de análisis de una imagen	11
2.2. Ejemplo de segmentación de una tapa circular amarilla	13
2.3. Histograma multimodal	14
3.1. Modelo teórico del sistema a desarrollar	18
3.2. Controlador manual a usar en el sistema	19
3.3. Diagrama de bloques: Obtención y procesamiento de un cuadro	20
3.4. Cámara de video	21
3.5. Computador a utilizar	22
3.6. Diagrama de bloques: Calibración	24
3.7. Zona de calibración: Región de análisis local	25
3.8. Caso ideal: Esfera ocupando la región de análisis local	25
3.9. Análisis en la región local para cada uno de los 3 planos del espacio de color RGB	28
3.10. Imagen binarizada a partir del análisis de histogramas	29
3.11. Zona que debe ocupar la esfera en la calibración	29
3.12. Regiones de trabajo para la detección de bordes	30
3.13. Posibles ubicaciones de la esfera en relación a la región de análisis local	33
3.14. Calibración: Esfera de colores naranja, morado y verde.	34
3.15. Diagrama de bloques: Segmentación automática	35
3.16. Detección de umbrales en la segmentación automática	36
3.17. Análisis en región global	37
3.18. Ejemplo 1: Seguimiento de la esfera a través regiones globales de análisis	38
3.19. Ejemplo 2: Seguimiento de la esfera a través regiones globales de análisis	39
3.20. Ejemplo 3: Seguimiento de la esfera a través regiones globales de análisis	39
3.21. Ejemplo 4: Seguimiento de la esfera a través regiones globales de análisis	40
3.22. Segmentación automática de la esfera de color naranja	40

3.23. Diagrama de Bloques: Seguimiento	41
3.24. Estimación de la circunferencia dentro de una región local de análisis	42
3.25. Seguimiento del objeto esférico analizando los bordes de la región segmentada	42
4.1. Aplicación 1: Sobre imposición de una circunferencia	45
4.2. Aplicación 2: Trayectoria del centro de la circunferencia	45
4.3. Aplicación 3: Sobreimposición del logo de la PUCP	45
4.4. Aplicación 4: Seguimiento y limpieza de la trayectoria de la circunferencia	46
4.5. Montaje de estructura para control de movimiento del controlador manual	46
4.6. Disposición física de dispositivos para realización de pruebas	47
4.7. Aplicaciones 1 y 3: Movimiento del objeto a 20 cm	50
4.8. Aplicación 2: Movimiento del objeto a 90 cm	50
4.9. Aplicación 3: Movimiento del objeto a 115 cm	50
4.10. Aplicaciones 3 y 4: Movimiento del objeto a 45 cm en ambiente oscuro	51
4.11. Aplicación 1: Movimiento del objeto a 85 cm en ambiente oscuro	51
4.12. Comportamiento del sistema en el eje “X”	52
4.13. Comportamiento del sistema en el eje “Y”	53
4.14. Comportamiento del sistema en el eje “Z” - Profundidad	54
4.15. Análisis de error en el eje “Z”	54
4.16. Aplicación 3: Acercamiento de la esfera	55
4.17. Aplicaciones de acercamiento y alejamiento	56
4.18. Aplicación 4: La esfera se aleja de la cámara	56
4.19. Aplicaciones 1 y 3: La esfera se acerca a la cámara en un ambiente oscuro	57
4.20. Comportamiento del sistema en el eje “X”	57
4.21. Análisis de error en el eje “X”	58
4.22. Comportamiento del sistema en el eje “Y”	58
4.23. Análisis de error en el eje “Y”	59
4.24. Comportamiento del sistema en el eje “Z” - Profundidad	59
4.25. Análisis de error en el eje “Z” - Profundidad	60
4.26. Velocidad máxima del controlador: Caso de análisis	61
4.27. Velocidad de movimiento de las pruebas realizadas I	62
4.28. Velocidad de movimiento de las pruebas realizadas II	63
4.29. Análisis de tiempo de ejecución - Aplicación 1	64
4.30. Análisis de tiempo de ejecución - Aplicación 2	65
4.31. Análisis de tiempo de ejecución - Aplicación 3	66
4.32. Análisis de tiempo de ejecución - Aplicación 4	67

Índice de cuadros

4.1. Tabla - Tiempos en Aplicación 1	65
4.2. Tabla - Tiempos en Aplicación 2	66
4.3. Tabla - Tiempos en Aplicación 3	67
4.4. Tabla - Tiempos en Aplicación 4	68
4.5. Presupuesto utilizado para el desarrollo del proyecto	68



Introducción

En las últimas décadas, los sistemas y dispositivos que permiten una interacción entre el ser humano y computador han logrado desarrollarse ampliamente, lo cual ha permitido que se mejore la interacción con sistemas computacionales. El sector del entretenimiento ha manifestado un gran crecimiento en las últimas décadas, lo cual se ha visto plasmado en la gran cantidad de consolas de video juegos vendidas [1], así como en la intensa campaña publicitaria que gira alrededor de estas o incluso en la cantidad de horas semanales que algunos usuarios les dedican [2]. Por ello, es importante el estudio de estos sistemas y las mejoras que se les puedan realizar con el fin de tener un producto más competitivo. Es así que en el presente documento se explican las consideraciones realizadas para la implementación de un sistema en tiempo real que permita el seguimiento y detección de un objeto esférico.

Para comenzar, en el Capítulo 1 se analiza el impacto que tiene en la actualidad el desarrollo de sistemas de visión por computador, haciendo especial énfasis en el sector del entretenimiento a través de la descripción de las diversas consolas de video juegos existentes y la tecnología que usan. Además, se explica cuál es el principio de funcionamiento básico que emplea el Play Station 3 en su sistema, esto servirá para poner en contexto la utilidad del sistema a implementar y los diversos conceptos a tener en cuenta en su desarrollo.

En el Capítulo 2 se describen los aspectos y conceptos más importantes en la segmentación y seguimiento de objetos orientados al procesamiento de una señal de video, lo que permitirá un mejor entendimiento de la metodología usada en el presente estudio. Finalmente, se mencionan algunos de los métodos alternativos que existen para el desarrollo de estos sistemas.

Posteriormente, en el Capítulo 3 se explica la metodología utilizada en la implementación de los algoritmos que permiten desarrollar este sistema, y se anotan las consideraciones que se tuvieron en cuenta en el desarrollo de cada etapa de la misma.

Finalmente en el Capítulo 4, se analiza la exactitud con la que se realiza el seguimiento de la esfera y las condiciones de funcionamiento del mismo; además, se analizan los tiempos de ejecución obtenidos para las aplicaciones desarrolladas tal que estén dentro de los valores esperados.

Por último, se presentan las conclusiones y recomendaciones pertinentes.

Capítulo 1

VISIÓN POR COMPUTADOR: APLICACIONES EN VIDEO JUEGOS

1.1. Visión por computador

La visión por computador es aquel campo de la ciencia y tecnología que tiene como objetivo el desarrollo de sistemas artificiales (*hardware* y *software*) que optimicen la interacción entre el ser humano y el computador utilizando para ello algún dispositivo (como una cámara) que permita captar las señales del entorno a analizar [3]. Existen varios beneficios derivados del uso de sistemas de visión por computador, uno de ellos es que este tipo de sistemas permiten ejecutar tareas con mayor velocidad que el ser humano. Además, se presenta como una solución en términos económicos, ya que desarrollar sistemas a través de un computador permite reducir costos en capacitación y manutención de personal. Finalmente, se presenta como una eficiente alternativa para la automatización y optimización de diversos procesos; ofreciendo además, la posibilidad de ubicarse y trabajar en lugares de difícil acceso. La Figura 1.1 muestra la estructura típica de un sistema de visión por computador aplicado al sector industrial.

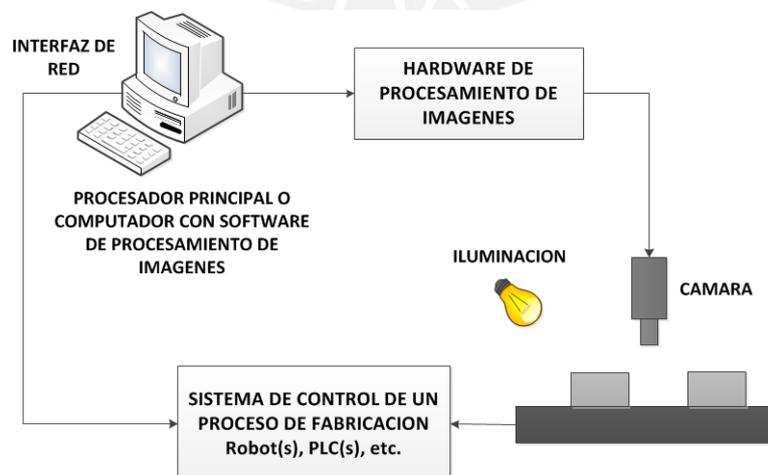


Figura 1.1: Aplicación industrial de un sistema de visión por computador [4]

Del esquema mostrado en la Figura 1.1 se destaca la intervención de factores tales como la cámara para la adquisición de imágenes, procesamiento de la imagen o video adquirido, posición y ubicación de la cámara, condiciones de iluminación del ambiente de análisis, computador donde se desarrollara la aplicación, entre otros [4]. Estos sistemas de visión por computador hacen uso de componentes en hardware y software que posibilitan su desarrollo y uso. Accesorios como cámaras de video, controladores lógicos programables, microcontroladores, DSPs, computadores personales, actuadores y los diversos algoritmos que deben ser implementados en los computadores son elementos cuya correcta interacción y desarrollo permitirán cumplir con los requerimientos planteados para el sistema.

De forma genérica, la implementación de un sistema de visión por computador debe seguir los siguientes pasos [4]:

1. Adquisición de la imagen.- En este paso se obtendrá la imagen que contiene la información a analizar.
2. Procesamiento de la imagen.- En este paso se aplicarán aquellas técnicas que permitan un mejor análisis de la imagen, además de mejorar su calidad. Esto permitirá la corrección de errores que hayan podido aparecer al momento de adquirir la imagen.
3. Extracción de características.- En este paso se segmentará la imagen resaltando el área u objeto de interés, para luego proceder a extraer características tales como tamaño, posición, bordes y centro, entre otras.
4. Toma de decisión.- Una vez que estén analizadas las características, se procederá a ejecutar alguna acción o tomar una decisión según sea la aplicación que se esté desarrollando.

En ese sentido, se busca desarrollar sistemas que permitan desarrollar habilidades computacionales que emulen algunas características de la visión humana. De esta manera, con las facilidades que brinda hoy en día el uso de computadoras rápidas, será posible desarrollar aplicaciones eficaces y eficientes según los requerimientos que se planteen para determinados proyectos, desarrollando para ello aplicaciones en campos como la vigilancia[5], monitoreo remoto aplicado en navegación, robótica, video juegos[6] [7], seguridad ante desastres naturales [8]; entre otros.

1.2. Video juegos: Análisis del contexto

Un campo de análisis de amplio desarrollo a nivel mundial es el de los video juegos, cuyo impacto a nivel social es sobresaliente dada la popularidad que tienen estas consolas hoy en día. La peculiaridad de estos juegos radica en el hecho de simular entornos reales, tal que con los datos proporcionados por el uso de sensores permitan al usuario disfrutar de aplicaciones que simulen movimientos y acciones en tiempo real. Desde hace varias décadas se vienen desarrollando diversos juegos en diversas plataformas [9], desde el popular Atari, pasando por el tradicional Nintendo,

Nintendo 64, Play Station 1, X Box, hasta lo que hoy en día se tiene con las consolas del Sony Play Station 3 y 4 [6], Microsoft Xbox 360 [10] con el Kinect y el Nintendo Wii Motion Plus [11]. Todos han revolucionado en su momento al ofrecer al usuario un concepto nuevo de realidad virtual. A continuación, se procede a describir estos 3 equipos de amplio uso en la actualidad.

1.2.1. Microsoft Xbox 360 - Kinect

Las últimas consolas del Xbox 360 con Kinect resaltan principalmente por no hacer uso de algún controlador o mando que el usuario tenga que usar para poder interactuar con el sistema. Su principio de funcionamiento se basa en la facultad que tiene de poder calibrar algunas partes del cuerpo humano (mano, brazos, piernas, etc.) y, en base a ello, poder generar a través de la interfaz gráfica un avatar que pueda simular los principales movimientos que el usuario realice con sus extremidades superiores e inferiores.

Para cumplir con lo descrito, el Xbox hace uso de un dispositivo llamado Kinect (ver Figura 1.2) que, a modo de sensor, detecta el movimiento del usuario. Por ello, ofrece la posibilidad de jugar inalámbricamente. El Kinect del Xbox cuenta con una cámara que registra el movimiento del cuerpo humano. Esta cámara, además, permite medir la distancia a la que se encuentra el usuario. Asimismo, posee un micrófono que permite la entrada de señales de voz.



Figura 1.2: Kinect del Xbox 360 [10]

Desde su lanzamiento en el 2005, el Xbox360 ha ido incrementando sus ventas, superando las 6 millones de unidades durante el año 2010 (Figura 1.5). Además, al 9 de marzo del 2011, Microsoft ya había vendido alrededor de 10 millones unidades del controlador Kinect [12], llegando a estar en los récord Guinness vendiendo alrededor de 8 millones de unidades a solo 60 días de su lanzamiento (noviembre del 2010), teniendo un promedio de ventas de 133333 unidades por día.

1.2.2. Nintendo Wii - Wii Motion Plus

Lanzado en noviembre del 2006, el Nintendo Wii se ha convertido en un claro ejemplo de éxito en el sector de los video juegos. La aparición de esta consola representó un cambio en la experiencia de juego del usuario, ya que reemplazó el uso de los típicos controles o mandos tradicionales -como los usados en el Super Nintendo- por el uso de un controlador llamado Wii Motion Plus (ver en la Figura 1.3); el cual, sujetado a la muñeca del usuario brinda una interacción más real con el computador.

El controlador Wii Remote permite la detección del movimiento en 3 dimensiones. Para esto, en primer lugar se pasa por una etapa de calibración que permitirá al sistema detectar los movimientos del ser humano; luego, el Wii Remote dispone de acelerómetros y un sistema de detección infrarrojo que permiten captar y registrar los movimientos y la posición del controlador operado por el usuario, y poder comunicarlos a la consola a través de Bluetooth. Una característica importante de este controlador es la posibilidad que tiene de ser conectado a dispositivos extras como el Nunchuk o el Wii Motion Plus. Este último incorpora sensores adicionales al Wii Remote que permite captar mejor los movimientos del usuario.



Figura 1.3: Controlador de nintendo Wii [11]

En un estudio realizado por NPD Group, entidad norteamericana especializada en la realización de estudios de mercado, en el 2009 se puede observar la gran popularidad que poseen sus juegos tales como: *Wii Sports Resorts*, *Super Mario Bros Wii*, *Wii Fit Plus*, entre otros [13]. La cantidad de unidades vendidas en aquel entonces oscilaban entre las 5000 y 8000 unidades. Esto refleja su importancia y potencial desde sus primeros años en el mercado.

1.2.3. Play Station 3 - Play Station Move

El Play Station 3 diseñado por IBM hace uso de los microprocesadores CELL [14], y con el uso de las unidades PPE (Power Processing Element) y SPE (Synergistic Processing Element) se tiene una arquitectura tal que permite el procesamiento rápido de determinados algoritmos y la obtención de imágenes de mejor calidad. Dentro de los accesorios que posee el Play Station 3 podemos observar la presencia del tradicional Dualshock 3, del Play Station Eye (que es una cámara digital que permite el registro de imágenes con una resolución de 640x480 píxeles a 60 Hz o 320 x 240 píxeles a 120 Hz del medio a analizar). Además, posee un micrófono que le permite registrar el sonido que hay alrededor, así como realizar la cancelación de ruido para contribuir a su mejor recuperación.

El Play Station 3 dispone de un controlador de movimiento (ver Figura 1.4) que, en conjunto con el Play Station Eye Camera, permite el seguimiento de la esfera que posee este controlador (esta esfera emite luces de fuerte intensidad que permiten distinguirla del fondo y así simplificar su identificación y posterior procesamiento.). Finalmente, con la ayuda de sensores como ace-

lerómetros se podrá detectar y simular determinados efectos del movimiento, estableciendo la comunicación con la consola a través del uso del Bluetooth 2.0.



Figura 1.4: Move controller motion del Play Station 3 [6]

Según un artículo publicado por Sony [1], al 29 de marzo del 2011 se habían vendido alrededor de 50 millones de unidades de Play Station 3 y alrededor de 8 millones de unidades del controlador Play Station Move, denotando así su rápido crecimiento en el mercado y perfilándose como uno de los más fuertes competidores del Xbox y el Nintendo Wii.

1.2.4. Competitividad en el mercado

En una sociedad en la que cada vez los avances tecnológicos causan un mayor impacto, son los videos juegos claros representantes de esta tendencia (esto puede ser constatado en las millonarias ventas que tienen cada una de estas consolas, cuyo uso mayoritario es de un público juvenil [15]). El Nintendo Wii, el Xbox 360 y el Play Station 3, llamados juegos de la séptima generación, compiten para conquistar el mercado y desarrollar lo último en la tecnología del entretenimiento.

En el caso del Nintendo Wii se reportan ventas que superan las 40 millones de consolas en el continente americano desde su lanzamiento (Noviembre del 2006) hasta el primer semestre del 2011 [16], siendo el continente de América un principal foco de potenciales clientes [17]. Por otro lado, como ya se ha mencionado, el Kinect del Xbox 360 ha reportado ventas que ascienden a 8 millones de unidades a solo 2 meses de su lanzamiento [12]; y, finalmente, el Play Station 3, reportando ventas en todo el mundo que ascienden a 50 millones de unidades y 8 millones de Play Station Move [1] al 29 de marzo del 2011.

Estas cifras demuestran mayor presencia de estos equipos en la sociedad actual. Así, hoy en día las personas hablan con total normalidad acerca de los diversos juegos o novedades que podrían traer dichos sistemas. Es así que, en la actualidad la adquisición de estos equipos es algo sumamente cotidiano. Un estudio realizado también por el NDP Group refleja que desde el 2009 se ha incrementado el 10 %, aproximadamente, de el número de horas semanales que las personas le dedican a este tipo de juegos. Además, se ha reportado que un usuario juega alrededor de 7.3 horas semanales y 5.8 horas semanales de Xbox 360 y Play Station 3, respectivamente [2].

El marketing que se realiza alrededor de estos productos es bastante intenso. Los juegos que promocionan cada una de estas marcas son muy conocidos a tal punto de que juegos como el Wii

Sports Resort (Nintendo Wii) o Halo 3 (XBox 360) figuran dentro de los 5 juegos más populares de un estudio realizado en el 2009 [18]. Así, la popularidad de la que gozan estos juegos influye en forma directa al número de consolas que se venden. Finalmente, dichas empresas están en una constante búsqueda de mejorar sus productos. El Nintendo Wii, por ejemplo, lanzó hace unos años el nintendo Wii U, en un deseo de competir e incluso superar al Kinect del XBox 360 o al Play Station Move (Play Station 3); por otro lado, Play Station ha lanzado y sigue mejorando su última consola, el Play Station 4. Estas marcas compiten hoy en día por conquistar el mercado como se puede apreciar en la Figura 1.5, donde se muestran los altos índices de venta de estas consolas en el continente americano.

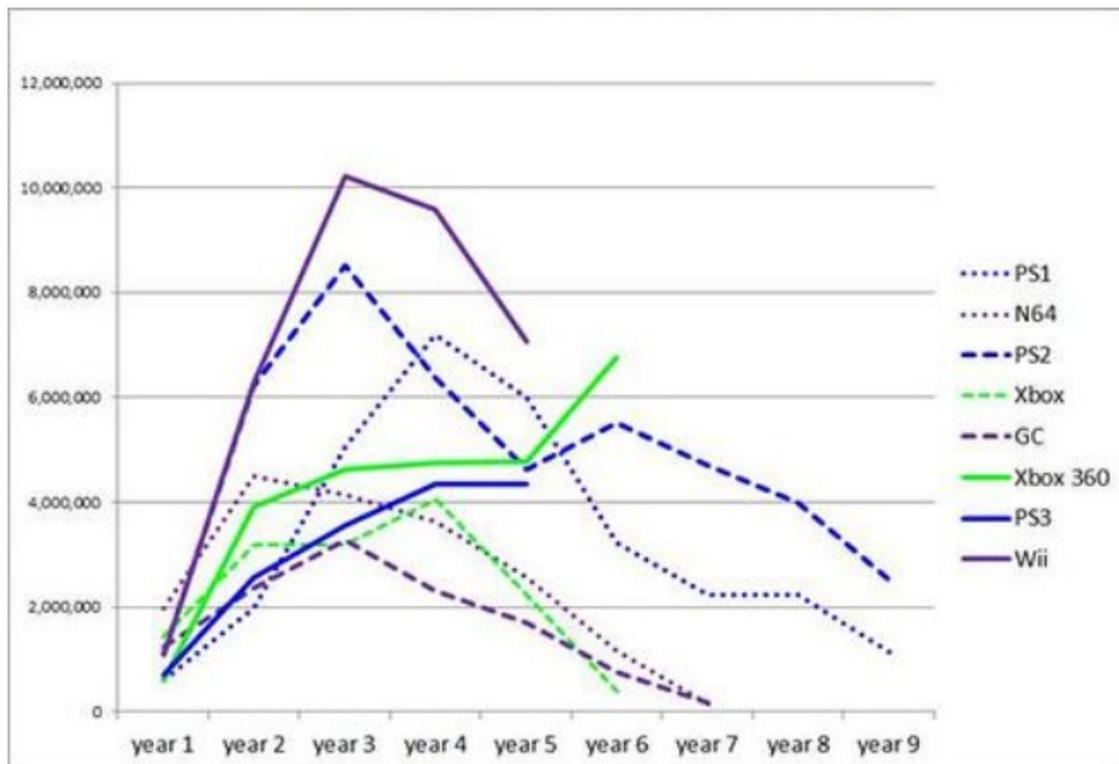


Figura 1.5: Cuadro comparativo de ventas de consolas de video juegos en América [13]. En donde “year1” hace referencia al año relativo en el que salió al mercado cada consola; 2006 para el Play Station 3 y el Nintendo Wii, 2005 para el XBox 360, dentro de los que más destacan.

1.3. Principio de funcionamiento del Play Station 3

Uno de los accesorios más resaltantes del Play Station 3 es el Play Station Move, cuyo controlador posee una esfera que emite una luz que le permite estar en contraste con el medio que lo rodea, lo cual facilita su reconocimiento. De esta forma, cuando el controlador es captado por el Play Station Eye, se efectuará internamente todo el procesamiento que dará como resultado los juegos y aplicaciones que posee el dispositivo. Los datos extraídos a través del análisis de la esfera permiten determinar la posición del objeto y con las señales que proporcionen sensores como los

acelerómetros [14] se puede extrapolar datos como la orientación o la rotación que experimenta el controlador, lo cual permite la simulación de efectos que la persona podría realizar.

Es relevante para el presente estudio resaltar el análisis inicial que el sistema realiza con la imagen que proporciona la esfera. Este aspecto será explicado a continuación:

1.3.1. Calibración

El proceso de calibración es la fase inicial del sistema, a partir del cual se definen condiciones de inicio como la ubicación del usuario (tomando como referencia el lugar donde está colocado la cámara), el tamaño y color de la esfera, entre otros; además de poder distinguir la esfera del espacio que la rodea.

El sistema inicia indicándole al usuario, a través de su interfaz gráfica, la posición y distancia desde la cual deberá ubicarse con respecto a la consola. Luego, el usuario deberá colocar la esfera en tres posiciones claramente definidas por el sistema: uno alrededor del muslo, otro alrededor del ombligo y, finalmente, uno a la altura del hombro. Terminado este proceso, el sistema tendrá datos referenciales de la ubicación del usuario y los puntos sobre los cuales se moverá; asimismo, se habrá identificado el objeto a seguir (en este caso la esfera).

1.3.2. Ejecución de la aplicación

Como siguiente paso, el usuario procede a ejecutar el juego o aplicación que desee. Es importante recalcar que si el usuario se aleja determinada distancia de la posición inicial en que se calibró el sistema, este notificará de ello al usuario y le indicará que debe acercarse a la posición inicial del juego, para luego seguir con su ejecución normal.

1.4. Síntesis de la situación actual

Existe una búsqueda permanente de medios que permitan la simulación y el seguimiento de determinados objetos, a través del uso de datos obtenidos mediante una cámara web, por ejemplo. Asimismo, se está en la constante innovación de aplicaciones en diversos campos como el de la seguridad, la salud, la inteligencia artificial, la robótica y el entretenimiento, entre otros. Este último, y a través de consolas como el Play Station 3, el XBox 360 o el Nintendo Wii, goza de un gran interés y popularidad [2], evidenciado en las cuantiosas ventas de estos videojuegos [1], [9], [12], [13]. Esto demuestra que este es un campo de estudio con muchas oportunidades aún de desarrollo.

Uno de los principales atractivos de estos sistemas es la velocidad con la que realizan el procesamiento y la ejecución de determinadas acciones. Por esta razón, la realización de una aplicación en tiempo real adquiere un papel muy importante, ya que se busca brindar una sensación de mayor realismo que satisfaga al usuario.

1.4. SÍNTESIS DE LA SITUACIÓN ACTUAL

9

Para lograr estos objetivos se hace importante el análisis de la señal que proporciona un video digital (como lo es en el Xbox 360 y en el Play Station 3, a través de sus Kinect y Play Station Eye respectivamente). Al analizar esta señal de video como un conjunto de imágenes, se establecen relaciones entre las mismas, lo que permitirá la toma de decisiones ante determinadas acciones dependiendo de la aplicación. Para ello, en cada imagen se tiene que definir un objeto que será la referencia del sistema, y cuyos atributos se identificarán. La medida constante de estos atributos permitirá identificar cambios y sobre estos se tomarán determinadas acciones.

No obstante, el reconocimiento de un objeto es un campo bastante amplio a analizar, ya que implica tener en consideración diversos aspectos, tales como la forma, el tamaño del objeto, las condiciones ambientales y los movimientos que puede experimentar el objeto. Para solucionar este punto se plantea la existencia de una gran variedad de métodos, como el uso de descriptores invariantes en el tiempo, detección de características del objeto en función a la intensidad de color que pueda presentar, entre otros. Se opta, sobre todo, por aquella que esté acorde con los objetivos y el grado de dificultad que presente la aplicación, teniendo como objetivo principal el cumplimiento de los requerimientos del usuario.

Finalmente, es por todo lo anterior que el presente estudio busca implementar un sistema que permita la interacción entre ser humano y computador mediante el seguimiento de un objeto esférico en tiempo real a través de una cámara web. Para ello, se plantea una solución que siguiendo el principio de funcionamiento del Play Station 3, aplica los conocimientos del procesamiento de imágenes en el análisis de una señal de video. De esta manera, se busca cumplir con los requerimientos de tiempo y exactitud que permita una interacción aceptable con el usuario.

Capítulo 2

MÉTODOS DE SEGMENTACIÓN Y SEGUIMIENTO DE OBJETOS ORIENTADOS A VIDEO DIGITAL

Diversos estudios se han desarrollado en torno al procesamiento de una señal de video [5], [19], [20]; en muchos casos, a través de la obtención y procesamiento de las imágenes que la componen y, manifestadas en aplicaciones como el seguimiento de vehículos automotores para control de tráfico, entre otros. De estos ejemplos se pueden establecer características y fases comunes en el desarrollo de este tipo de sistemas:

1. Adquisición de imágenes correspondientes a una señal de video.
2. Identificación del objeto de interés en cada imagen.
3. Identificación y cálculo de parámetros de interés del objeto a analizar de cada imagen (centroide, varianza, desviación estándar, etc).
4. Establecimiento de relaciones y comparaciones entre los parámetros obtenidos de cada imagen de la señal de video a lo largo del tiempo.
5. A partir de las relaciones y comparaciones hechas, se realizarán determinadas acciones según la aplicación desarrollada.
6. Cierre de la aplicación y liberación de los recursos que se hayan utilizado para su uso en futuras ocasiones.

La mayoría de aplicaciones desarrolladas siguen estas fases de manera general, diferenciándose entre ellas en los métodos que usan.

Uno de los principales retos que se pueden encontrar en el procesamiento de señales de video se encuentra en la velocidad con la que se procesa dicha señal; sobre todo si se debe procesar correctamente esa cantidad de datos en un tiempo menor al tiempo que exista entre la obtención

de dos cuadros consecutivos. Además, debe tenerse en cuenta que un video digital posee una gran cantidad de datos por los diversos cuadros que lo conforman y estos a su vez están definidos por el tamaño y la calidad de la comprensión.

La velocidad de procesamiento de una señal de video de una cámara digital (fps: frames per second. En señales de video = 30 fps), es un factor principal para desarrollar operaciones o aplicar técnicas que permitan un adecuado procesamiento de la imagen y una posterior extracción de características de la región de interés.

En la Figura 2.1 se muestra un diagrama de bloques de las fases que se suelen seguir para el procesamiento de una imagen.

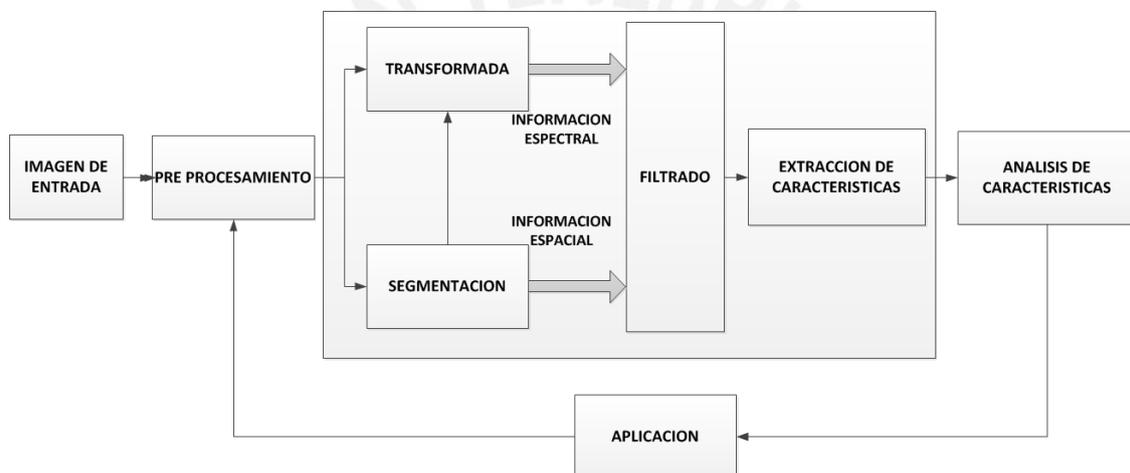


Figura 2.1: Proceso de análisis de una imagen [21]

Esta imagen ilustra lo explicado en el primer capítulo sobre las diversas fases que corresponden a un sistema de visión por computador, de las cuales las etapas de adquisición de la imagen, segmentación y seguimiento resaltan porque consumirán la mayor parte de tiempo disponible que haya entre cuadro y cuadro para poder procesar la imagen.

Para un correcto entendimiento del presente estudio se definirán algunos conceptos básicos y las técnicas de procesamiento más usuales para procesamiento de señales de video.

2.1. Definiciones generales

2.1.1. Video

El video es la tecnología de la captación, grabación, procesamiento, almacenamiento, transmisión y reconstrucción por medios electrónicos digitales o analógicos de una secuencia de imágenes que representan escenas en movimiento [22].

Se pueden distinguir dos tipos de videos, el analógico y el digital. El video analógico es aquel que usa señales que varían continuamente en el tiempo, cuya amplitud representa información de sonido o de intensidad luminosa. El VHS o los dispositivos de cinta magnética son medios donde se almacenan este tipo de señales. Un video digital presenta una señal codificada, donde se usan variables discretas que permiten mayor calidad y contraste que una señal analógica. El video digital ofrece la posibilidad de compresión de datos y es posible su edición a través del tratamiento de las imágenes o sonido que la conforman sin la necesidad de seguir un orden predeterminado. Esto permite la implementación de algoritmos sofisticados para su procesamiento. Además, ofrece la posibilidad de ser trabajado en tiempo real. En este punto es importante entender dos conceptos importantes:

Formato.- El formato de un video es la forma en la que el video es guardado en nuestro ordenador. Estos pueden ser AVI, AVI2, WMV, RealVideo, QuickTime, entre otros. Esto sería equivalente a guardar una imagen en .jpg, .gif, .bmp, .tiff etc.

Códec.- Codificador / Decodificador. Software que codifica y decodifica el video.

2.1.2. Imagen

Una señal de video digital puede estar compuesta de una secuencia de imágenes que representan una zona específica en un determinado instante de tiempo. Una imagen es definida como una matriz de $M \times N$ píxeles, donde cada pixel representa una función $f(x, y)$, donde x, y son coordenadas espaciales y f es la función de intensidad que representa a las coordenadas [22].

Para su procesamiento una imagen puede estar representada en escala de grises [23] o en escala de colores [24], según el espacio de color con el cual se trabaje.

2.2. Espacios de color

Los diversos objetos que percibimos a nuestro alrededor son representados por diversos componentes de color, brillo, iluminación, entre otros. Un color puede ser representado desde diversos espacios de color, donde cada uno de estos espacios define el modelo de composición del mismo. Algunos espacios representarán mejor que otros los diversos colores que puede percibir el ser humano.

Algunos de los espacios de color más conocidos son:

1. Espacio de color RGB.- Este espacio representa los tres colores primarios: rojo, verde y azul. Se trata de un modelo orientado a *hardware* y es conocido por su uso para visualización de imágenes en monitores [25].

2. Espacio de color HSV.- Este espacio está formado por los componentes *Hue*, que describe al color puro; saturación, que representa la cantidad de luz blanca que hay en un tono; y *value*, que se refiere al brillo de la imagen. Este modelo es usado para análisis de imágenes [25].
3. Espacio de color YCrCb.- Este espacio está orientado a *hardware* y está formado por los componentes de luminancia o brillo (Y) y de crominancia (Cr y Cb) de la imagen[25].
4. Espacio de color CIE L*a*b*.- Este espacio de color representa mejor los colores que puede percibir el ojo humano. Está conformado por los componentes de luminancia ($L^*=0$: Negro, $L^*=100$: Blanco), la posición entre magenta y verde ($a^*>0$: Magenta, $a^*<0$: Verde), y la posición entre amarillo y azul ($b^*>0$: Amarillo, $b^*<0$: Azul).

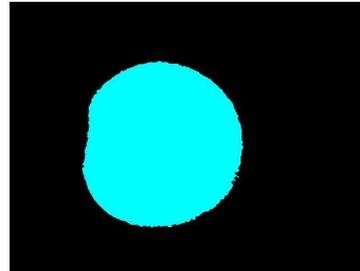
2.3. Segmentación de imágenes

La segmentación es el proceso de partición de imágenes en sub regiones [22]. Para ello, se han desarrollado diversas metodologías basadas en principios tales como:

1. Discontinuidad.- Algoritmo para tratar imágenes monóchromas, enfocado en detectar cambios bruscos de intensidad (como en los bordes), y en base a ello diferenciar regiones [22].
2. Regiones homogéneas.- Algoritmo para tratar imágenes monóchromas, enfocado en detectar regiones con el mismo patrón de similitud, y en base a ello diferenciarlo del resto [22].



(a) Imagen original, sin segmentar.



(b) Imagen segmentada

Figura 2.2: Ejemplo de segmentación de una tapa circular amarilla

Una de las herramientas usadas en la segmentación de imágenes son los histogramas.

Un histograma $H(x)$ es una representación gráfica donde el eje vertical representa la frecuencia con la que puede ocurrir o aparecer la variable x [23]. En el procesamiento de imágenes, un histograma puede ser usado para la segmentación de imágenes y reconocimiento de objetos a través de la definición previa de un umbral.

Sobre la base de estos histogramas, se pueden realizar muchos análisis según sea la naturaleza de los mismos (unimodal [26], bimodal, multimodal [27] (ver Figura 2.3)). Además, al obtener

parámetros como la “*probability density function*” o el análisis de histogramas acumulados, se podrán analizar los cambios en las escenas ante efectos como la oclusión.

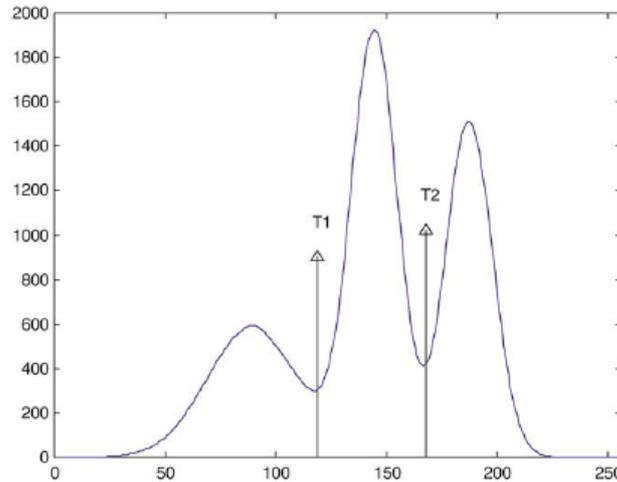


Figura 2.3: Histograma multimodal [27]

La utilidad de un sistema que segmenta y reconoce objetos está definida por su robustez ante cambios tales como transformaciones de similaridad (escala, rotación y traslación en dos dimensiones), transformaciones de afinidad (escala, rotación, traslación y transvección en dos dimensiones), transformaciones proyectivas, entre otros. A continuación se describe brevemente algunos de los conceptos más utilizados en la segmentación de imágenes:

1. Regiones locales y de correspondencia.- Algunos métodos de segmentación exigen la definición de regiones locales dentro de un cuadro y así, en base a histogramas de color poder reconocer determinado objeto [28]. Otros métodos también proponen el uso regiones de correspondencia [29]; es decir, en base a características globales (centroide, centro de masa, etc.) o locales (definido en función del objeto en particular que se analizará) establecer relaciones entre dos cuadros consecutivos. A partir de los resultados de dichas relaciones, se podrá realizar acciones según los objetivos que se planteen.

En el presente estudio se analizarán regiones locales y cuando el objeto de estudio sea detectado, se establecerán relaciones a lo largo del tiempo analizando parámetros como el centroide, lo que permitirá su seguimiento.

2. Descriptores.- También se han realizado estudios que tienen como objetivo el reconocimiento de múltiples objetos [30]. Para ello, se hacen uso - en algunos casos - de descriptores que sean invariantes a las transformaciones y cambios antes mencionados. Es importante mencionar que buscando reconocer un objeto, en muchos casos se busca detectar puntos de control cuyas características sean invariantes a las transformaciones mencionadas [31], [32], [33]. Para esto, se usan descriptores geoméricamente invariantes en un espacio bidimensional y que sea invariantes incluso a transformaciones proyectivas.

3. Acerca de otros métodos de segmentación.- Existen otros métodos de segmentación como k-means, regiones de crecimiento, contornos activos, entre otros; algunos de los cuales parten de un punto inicial y en función a características (como color, posición, etc.), se van expandiendo hasta cierto punto donde las condiciones establecidas no se cumplan más, lo que indica el fin de la segmentación.

Los métodos de reconocimiento mencionados anteriormente buscan la detección de diversos tipos de objetos, incluso cuyos bordes no necesariamente representen a alguna figura geométrica conocida. En el caso del presente estudio, el contorno de la esfera que es objeto de análisis es una circunferencia, lo que permite simplificar su reconocimiento.

2.4. Seguimiento de objetos

Para el seguimiento de objetos, se han desarrollado teorías que están relacionadas al desarrollo de modelos, ya sean estos estáticos o adaptativos [34].

1. Modelos estáticos.- Se delimita el campo de estudio y, sobre las hipótesis que se pueden plantear en función a los cambios esperados se toman decisiones. En este caso, si sucede algún cambio inesperado el sistema no podría tomar alguna acción correcta.
2. Modelos adaptativos.- La naturaleza predictiva y recursiva de estos modelos buscan aprender de las situaciones o características que acontecieron y, en base a ello poder predecir algún cambio determinado.

Sobre estos conceptos se desarrollan diversas teorías como el “Forward Backward Error”; o las que están relacionadas con el Learned Lineal Predictor [35], donde se establece la existencia de un conjunto de predictores locales del movimiento. Así mismo, el método de “Sum of Square Difference” (SSD) es la base de muchos de los métodos desarrollados hoy en día, dado que este modelo permite analizar la variación existente entre datos obtenidos de un conjunto analizado. La magnitud de esta variación indicará los posibles cambios a los que está sujeto el conjunto de datos.

También se han desarrollado aplicaciones más complejas [36], en las que se plantea el uso de un seguidor KLT (Kanade-Lucas-Tomasi), que con ayuda de descriptores SIFT (Scale Invariant Feature Transform [31]) y el uso de *optical flow*, permitirá la realización de un trabajo óptimo, ya que puede manejar diversos movimientos de cámara como paneos, variaciones de la apariencia humana, desorden, oclusión, entre otros. Esto se logra dado que este tipo de modelos plantea el reconocimiento de la acción que genera el movimiento. Para ello, segmenta las imágenes por homografía identificando las principales características de la misma. También se tienen las aplicaciones desarrolladas en torno al modelo TLD (tracking learning - detection) [34], en la que desarrolla un mecanismo bajo el nombre de Predator que es capaz de detectar, reconocer y seguir objetos (como la punta de los dedos, caras, entre otros) con mucha rapidez y dando al usuario la posibilidad de acceder a varias opciones.

Algunos de los métodos más conocidos en procesos que implican el seguimiento de objetos hacen uso de técnicas como el filtro de Kalman [37], [38], que permite estimar el estado de un sistema lineal basado en un modelo de predicción - actualización. Esto le permite el desarrollo de aplicaciones con histogramas unimodales y el manejo de procesos en paralelo. Otro modelo utilizado es aquel que corresponde al filtro de partículas [20].

2.5. Síntesis de la investigación

En el presente estudio se abordará el tema de la detección y seguimiento de un objeto esférico de una señal de video en tiempo real, manifestado en el desarrollo de aplicaciones que cumplan con los siguientes requerimientos:

1. Rapidez.- Esta cualidad permite el reconocimiento del objeto y la toma de acciones respectivas en un tiempo adecuado, lo cual permitirá el funcionamiento normal y correcto del sistema en tiempo real. La toma de decisiones debe darse en un tiempo tal que sea menor al existente entre dos cuadros consecutivos de una señal de video; es decir, menor a 30 cuadros por segundo.
2. Exactitud.- Esta cualidad exige que la aplicación desarrollada represente exactamente a la esfera que se está monitoreando, la cual es representada a través de su centroide. La exactitud del sistema será definido cuantitativamente en el capítulo 4 con los resultados obtenidos en función de la distancia entre el objeto esférico y la cámara de video.
3. Robustez.- Esta cualidad le permite al sistema poder trabajar bajo determinadas condiciones ambientales y soportar cambios en posición (acercamiento, alejamiento, desplazamiento vertical y horizontal). Estas condiciones de funcionamiento y los requerimientos del sistema serán definidos cuantitativamente en el capítulo 4 con los resultados obtenidos de las pruebas realizadas. No obstante, se considera que el sistema para su correcto funcionamiento, trabajará en condiciones de iluminación y ambiente de una oficina tradicional.
4. Optimización.- Al ser un sistema que plantea una solución inspirada en el modelo del Play Station 3, pero sin el uso de dispositivos como acelerómetros o los procesadores CELL; el diseño y desarrollo del sistema buscará la optimización de recursos a nivel de *hardware* y *software*, cumpliendo de igual forma con los requerimientos de la aplicación.

La calidad del sistema que se va a desarrollar no solo estará limitada por la detección y seguimiento del objeto esférico, sino también por hacerlo bajo los parámetros de velocidad, precisión, robustez y optimización ya explicados.

Capítulo 3

DETECCIÓN Y SEGUIMIENTO DE UN OBJETO ESFÉRICO EN UN VIDEO DIGITAL EN TIEMPO REAL

El objetivo del presente estudio es el desarrollo de un sistema que permite la interacción entre ser humano y computador a través de la ejecución de aplicaciones orientadas al entretenimiento del usuario, haciendo para ello uso de las herramientas que proporcione el procesamiento de imágenes y las facilidades que brinda el uso de herramientas informáticas como el lenguaje de programación C. Esto se logrará a través de la detección y seguimiento automático de un objeto esférico en tiempo real en un video digital; para ello se plantean los siguientes objetivos específicos:

1. Desarrollo de un sistema en tiempo real que opere con una cámara de video que trabaje a 10 cuadros por segundo, sin presentar retardos.
2. Desarrollo de un sistema que detecte automáticamente la posición y tamaño de la esfera dentro un cuadro tras ser calibrado.
3. Cálculo de histogramas como método que permita la obtención de umbrales adecuados que segmenten correctamente la imagen.
4. Implementación de filtros (máscara de Gauss, Laplace) que faciliten el procesamiento de la imagen, permitiendo un seguimiento rápido y efectivo.
5. Implementación del sistema en lenguaje C, haciendo uso de la librería ffmpeg sobre el sistema operativo Linux.

El planteamiento de la solución está inspirado en el sistema Play Station Move que permite el seguimiento de la esfera que posee su controlador de movimiento. Además, al ser este un sistema que interactuará directamente con el usuario, deberá ser desarrollado bajo condiciones de tiempo de procesamiento y exactitud que permita una interacción aceptable con el usuario.

3.1. Descripción del sistema

3.1.1. Modelo teórico del sistema

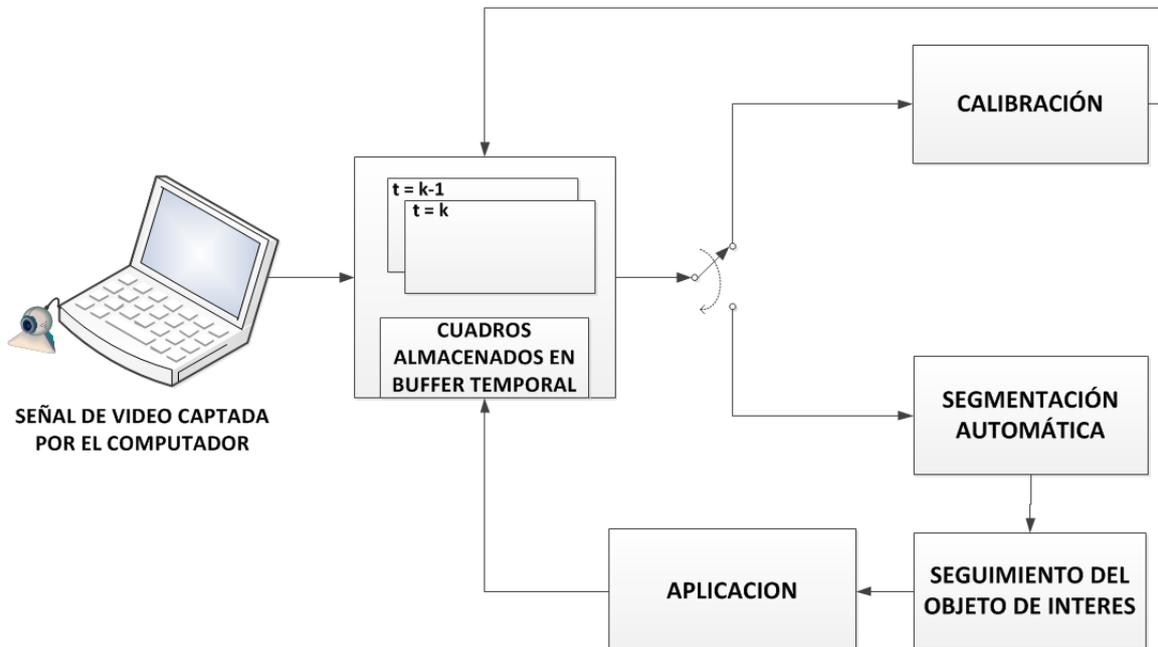


Figura 3.1: Modelo teórico del sistema a desarrollar

En la Figura 3.1 se puede observar el modelo teórico del sistema a desarrollar, el cual realizará lo siguiente:

1. El sistema obtendrá los cuadros correspondientes a la señal de video y almacenará el cuadro de análisis en un buffer temporal. Este buffer será actualizado cada vez que el sistema obtenga un nuevo cuadro.
2. Luego, el sistema buscará en primera instancia ser calibrado, es decir, definir los parámetros de color (este es el más importante), tamaño, posición y resolución espacial del objeto a seguir. Para ello, examinará continuamente una región de análisis local donde se deberá posicionar inicialmente a la esfera. Cuando esta región presente parámetros adecuados, el sistema se habrá calibrado y estos parámetros serán almacenados.
3. Después, la imagen se segmentará automáticamente utilizando los datos que han sido almacenados en la fase previa de calibración, es decir, se resaltará el objeto de interés (esfera) con respecto al espacio que lo rodea.
4. Cuando se haya detectado la esfera, se procederá a extraer características de la misma tales como tamaño y posición dentro de la región local donde se encuentre. Estos parámetros no necesariamente coincidirán con aquellos que se obtuvieron producto de la calibración,

alguna variación con respecto a estos parámetros indicará que el objeto ha experimentado cambios que el sistema deberá analizar.

5. Con las características extraídas y, luego de ser comparadas con los valores que les precedieron, el sistema analizará los cambios que experimentó la esfera y, producto de esa comparación, actualizará la posición y tamaño de la región de interés local dentro del cuadro de análisis. Después, esta actualización se manifestará en el desarrollo de alguna aplicación, ya sea a través de la sobre-imposición de una imagen o la realización algún juego que permita la interacción con el computador, entre otros.

El reconocimiento de un objeto involucra el análisis de los cambios a los cuales puede estar expuesto el mismo: rotación, traslación, etc. Por esta razón, delimitar a una esfera las características de forma del objeto a seguir permite simplificar el procesamiento a realizar, ya que conociendo la forma del contorno del objeto (una circunferencia), el análisis se enfocará en dos aspectos: cambios en posición y cambios en el tamaño del objeto esférico. Estos datos podrán ser usados en el desarrollo de determinadas aplicaciones.

Para el desarrollo del presente sistema se armó un controlador manual con una esfera de 10 cm (la que será nuestro objeto a seguir), como se puede apreciar en la Figura 3.2.



Figura 3.2: Controlador manual a usar en el sistema

Este dispositivo, además de la esfera, consta de una linterna, desde donde el usuario podrá sujetar el controlador y así realizar los movimientos que desee. La linterna iluminará la esfera con el objetivo de atenuar los problemas que generan los efectos de la iluminación externa y generar mayor contraste con el medio que rodea a la esfera, haciendo más fácil su detección.

En este punto es importante resaltar que a diferencia del Play Statio Move donde la iluminación es activa, es decir, el sistema usa luz de longitud de onda auto-regulable para maximizar el contraste con el fondo; en el sistema desarrollado en el presente trabajo se usa iluminación pasiva, lo que limita el alcance del sistema a operar bajo determinadas condiciones de iluminación externa como se detallará en el capítulo 4.

3.1.2. Lógica de la programación

Partiendo del hecho de que un correcto uso de técnicas de programación permitirá desarrollar una aplicación rápida, la cual debe tener un tiempo de respuesta casi imperceptible en la interacción con el usuario, es que se ha implementado el sistema usando el lenguaje C sobre el sistema operativo Linux buscando cumplir con los objetivos antes expuestos. En la Figura 3.3 se muestra las diversas fases que componen la lógica de programación implementada para realizar lo descrito en el modelo teórico general del sistema (Figura 3.1):



Figura 3.3: Diagrama de bloques: Obtención y procesamiento de un cuadro

1. Obtención del cuadro.- Esta función se encarga de obtener adecuadamente los datos de un cuadro, usando para ello información correspondiente al codificador y formato correctos para la interpretación de los datos obtenidos de la cámara Logitech QuickCam Orbit AF (ver Figura 3.4), que presenta las siguientes características técnicas [39]:
 - a) Enfoque automático del sistema.
 - b) Sensor de alta resolución: 2 megapíxeles con tecnología RightLight 2.
 - c) Profundidad de color: 24 bits de color verdadero.
 - d) Captura de video de hasta 1600 x 1200 píxeles (calidad HD).
 - e) Velocidad de cuadro de hasta 30 cuadros por segundo.
 - f) Conexión vía USB 2.0 de alta velocidad.
 - g) Cámara óptica Carl Zeiss.

El alcance del presente estudio no aborda el análisis y estudio de la cámara de video escogida, ya que el sistema deberá estar en la capacidad de funcionar con cualquier tipo de cámara, siempre y cuando sea compatible con el sistema operativo con el que se esté trabajando (Linux en este caso). Además, deberá tener los controladores necesarios que permitan decodificar adecuadamente la señal de video a través de la librería ffmpeg. Con todo esto, se podrá obtener la imagen en el espacio de color deseado y a una velocidad de trabajo que sea mayor o igual a 10 fps.



Figura 3.4: Cámara de video

La cámara Logitech que se usó sí es compatible con el sistema operativo Linux y hace uso del controlador UVC que proporciona la herramienta *video for linux* (v4l), con la cual se podrá procesar adecuadamente la señal de video.

2. Conversión del cuadro a valores adecuados.- Esta función convierte los valores del cuadro a sus equivalentes para el espacio de color con el que se trabajará.
3. Análisis de movimiento .- Esta es la función principal del presente estudio. Se encarga de analizar lo sintetizado en la Figura 3.1. Para ello, analizará si el sistema ha sido o no calibrado. De serlo, dará como resultado la posición y tamaño de la esfera.
4. Desarrollo de aplicación .- Con los datos obtenidos de la función anterior se desarrollarán aplicaciones de acuerdo al gusto del usuario. Para el presente estudio se desarrollaron cuatro aplicaciones que se mostrarán en el capítulo 4.
5. Mostrar imagen.- Esta función se encarga de mostrar a través de una interfaz gráfica la imagen original con los cambios producto del desarrollo de la aplicación.

Se resalta una vez más la importancia de la función correspondiente al análisis del movimiento, ya que ahí está el objeto de interés de este estudio.

Además, como se mencionó, se busca que el tiempo de procesamiento esté dentro de valores aceptables; por lo que es importante conocer sobre qué computadora, sistema operativo o programas se desarrollarán para el presente sistema. A continuación, se describirán los componentes principales para el desarrollo del sistema que, sumados al controlador manual, la cámara de video y el programa en sí, componen el sistema en su totalidad.

El computador

El computador utilizado para la realización del presente estudio es una Sony Vaio cuyo procesador posee las siguientes características: Intel(R) Core(TM) 2 Duo CPU T5450 Q @ 1.67GHz. (ver Figura 3.5)



Figura 3.5: Computador a utilizar

Sistema operativo:

Este sistema será desarrollado sobre el sistema operativo GNU/Linux 2.6.38-8 generic, versión # 42 de la distribución Ubuntu 11.04. Además, para futuros estudios, esta plataforma puede ser fácilmente utilizada con otras técnicas de programación, como la implementación de hilos, las cuales deben ejecutarse con una prioridad para el sistema operativo y así reducir el tiempo de ejecución del sistema.

Lenguaje de programación

Se programará en lenguaje C por la familiaridad que se tiene con él al ser un lenguaje de alto nivel. También por las facilidades que brinda para usar códigos o bibliotecas existentes y para desarrollar programas modulares o aplicaciones.

Librería ffmpeg

Para poder procesar la señal de video se hace necesario tener una herramienta que tenga los formatos, códecs, filtros, funciones que permitan procesar la señal de video, de tal forma que pueda ser trabajado de forma óptima. Para ello, se usará la librería ffmpeg [40] que provee todas las herramientas descritas anteriormente.

Esta tesis hace uso de programas que fueron desarrollados en el laboratorio de DSP (Digital Signal Processing) de la PUCP. Estos programas se encargarán de recibir la señal de video que se capture de la cámara web. Además, se encargará de encontrar el códec y formato adecuado para procesar los cuadros que la componen hasta que, finalmente obtengamos la imagen correspondiente en el espacio de color que se desee. Dicha imagen será almacenada en determinado espacio de memoria, de donde se empezará su procesamiento.

A continuación se procederá con la descripción de los bloques que conforman el modelo teórico planteado.

3.2. Adquisición de la imagen

Para la obtención de la imagen de una señal de video se hará uso de la librería ffmpeg. El sistema será iniciado indicando por línea de comandos donde se encuentra el dispositivo que permitirá el registro de imágenes y el controlador que permitirá codificar la señal (video for linux). Posteriormente la librería ffmpeg inicializará datos como:

1. Tamaño de la imagen a extraer del video digital: 640 x 480 píxeles.
2. Controlador a utilizar para la codificación de la señal de video: Video for linux (v4l).
3. Velocidad de captura de la señal de video: 10 cuadros por segundo (fps).
4. Espacio de color: RGB de 24 bits.

Con estos datos la librería procederá a interpretar la señal de video, y encontrará un formato y un códec que permitirá el procesamiento del video digital. También se encargará de reservar espacios de memoria donde se almacenarán los datos obtenidos de cada imagen. De esta forma, después de haber inicializado el sistema con los parámetros antes explicados, se estará a la espera de recibir una señal de la cámara de video indicando que los datos que vendrán a continuación corresponden a los de un cuadro.

Cuando un cuadro sea recibido en su totalidad y almacenado en el espacio de memoria asignado, se procederá con la ejecución de las demás fases. Estas son calibración, segmentación, seguimiento y desarrollo de una aplicación.

3.3. Calibración

En el proceso de calibración se busca detectar parámetros de color que representen a la esfera, a partir del cual el sistema pueda funcionar automáticamente. Cuando el sistema es calibrado, además de los parámetros de color y tamaño (previamente definidos), se obtienen otros que serán importantes para el posterior análisis. Estos parámetros son:

1. Resolución espacial.
2. Posición (en tres dimensiones) de la esfera.

Estos parámetros estarán definidos por la región local de análisis que se estableció como punto de partida para la calibración y por el conocimiento a priori del tamaño de la esfera. Esta región tiene un tamaño de 60 x 60 píxeles. La calibración se dará cuando la esfera se encuentre a una distancia frontal de 90 cm de la cámara de video (según mediciones realizadas). Por último, se recuerda que la esfera tiene un diámetro de 10 cm. Con estos datos y el parámetro de color de la esfera en valores adecuados que indiquen que el sistema ha sido calibrado, se podrá realizar el cálculo de los dos parámetros mencionados. A continuación se muestra el diagrama de bloques de la etapa de calibración.

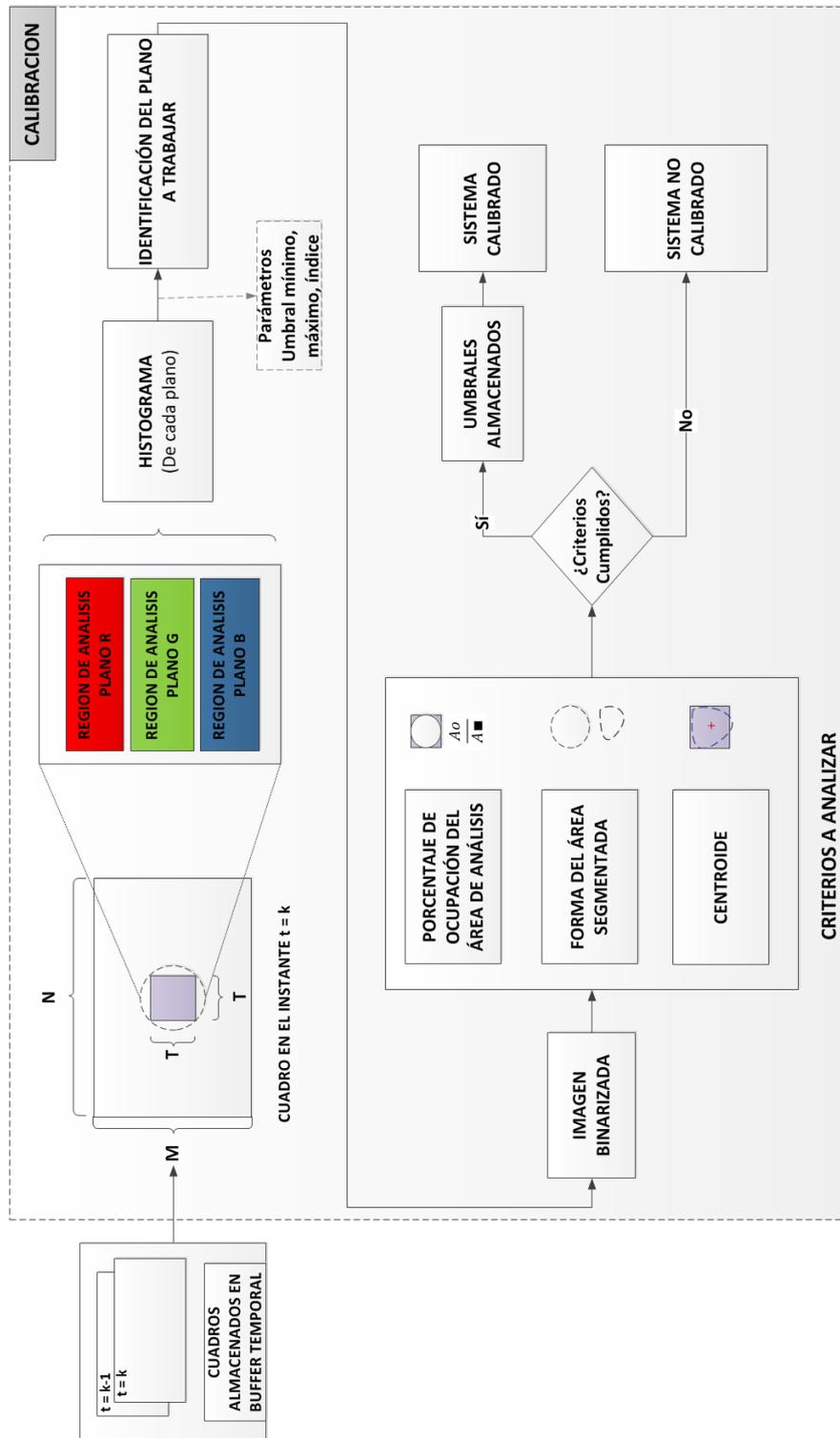


Figura 3.6: Diagrama de bloques: Calibración

3.3.1. Región local de análisis

Dada una imagen $I(x, y)$, $x \in [0, M]$, $y \in [0, N]$, de dimensiones $M \times N$ píxeles, se define una región local como aquel espacio dentro de la imagen que representa la ubicación del objeto a seguir:

$$I_{local}(x, y), x \in [0, T], y \in [0, T], \text{ donde } T < \min(M, N) \quad (3.1)$$

Inicialmente esta región local estará ubicada en el centro de la imagen como se puede apreciar en la Figura 3.7. Además, se calculará la resolución espacial al momento de la calibración, dado que es conocido el tamaño de la región local de análisis y de la circunferencia.

$$\text{Resolución Espacial} = \frac{60 \times 60 \text{ píxeles}}{10 \times 10 \text{ cm}^2} = 6 \text{ pix/cm.} \quad (3.2)$$



Figura 3.7: Zona de calibración: Región de análisis local



Figura 3.8: Caso ideal: Esfera ocupando la región de análisis local

La importancia del uso de regiones locales radica en poder disminuir el tiempo de procesamiento, ya que resulta más rápido analizar una parte de la imagen que analizar la totalidad de la misma. Además, permite limitar la búsqueda del objeto esférico en las regiones adyacentes a la región definida. Este tipo de solución es abordado en [28], en la que se definen regiones de análisis

y sobre las cuales se aplicarán histogramas de color para la búsqueda de determinados objetos. Tomando de base esta metodología, se analizará la región local definida.

Cabe resaltar que cuando el sistema sea calibrado, el análisis en regiones locales se seguirá manteniendo, ya que las regiones locales representan al objeto y, por ende, su análisis continuo permite el registro de posibles cambios de tamaño y posición debido a los movimientos de la esfera.

Segmentación por color:

En principio se realizará un análisis enfocado en segmentación por color sobre la región de análisis local definida. Este sistema, más allá de enfocarse en analizar los efectos de la luminancia, brillo u otros, la detección y seguimiento de la esfera se centrará en analizar los parámetros de color que tenga la misma y a partir de ello podrá segmentarla.

Una imagen en el espacio RGB estará compuesto por 3 planos, cada uno de tamaño $M \times N$ píxeles y representan los componentes de color primario: rojo, verde y azul. Debido a que se trabajará con regiones locales, el tamaño de cada uno de estos planos será de $T \times T$ píxeles. El proceso iniciará con el análisis para cada plano de la región de análisis local, lo cual permitirá obtener un histograma para cada espacio de color. Luego, para un mejor análisis se aplicará una máscara de Gauss sobre este gráfico con el objetivo de suavizar las curvas que presente y eliminar los cambios bruscos.

Tal como se propone en [26] y [27], el objetivo de la realización de estos histogramas es la identificación de objetos dentro del área analizada que presenten aproximadamente el mismo rango de color, es así que se busca obtener los mínimos y máximos locales. Estos son umbrales que definirán el rango dentro del cual se encontrará el objeto de estudio. Para lograr esto, se aplican los pasos descritos en [27]. A continuación, se procederá a describir los modelos matemáticos usados para realizar dicha segmentación.

Cálculo del histograma:

Para todo $I_{local}(a, b)$, $a \in [0, T]$, $b \in [0, T]$ se plantea:

$$H(x) = \begin{cases} H(x) + 1 & \text{si } I_{local}(a, b) = x \\ H(x) & \text{si } I_{local}(a, b) \neq x \end{cases} \quad (3.3)$$

Donde $x \in [0, G]$, $G =$ Valor máximo de los píxeles en cada espacio de color de la imagen dada.

Suavizado del histograma:

Como se explicó, se aplicará una máscara de Gauss [27]. La longitud de la máscara, al que se

denominará *bins* [27], influirá en qué tanto sea suavizado el histograma. Un número muy elevado de *bins* producirá que se presente una única campana en el histograma; es decir, solo se podría identificar una única región. Un número de *bins* muy bajo no producirá un cambio en el histograma original. Por ello, el objetivo es obtener resultados que permitan distinguir regiones claramente diferenciadas entre sí dentro de la región local de análisis. Para efectos del presente sistema se ha usado 21 *bins*, dado que con este valor se han obtenido resultados aceptables al momento de diferenciar los objetos dentro de la región de análisis.

Se aplicará la convolución entre el histograma original y la máscara gaussiana descrita, de la cual se obtiene la siguiente expresión:

$$\bar{H} = H * W_g. \quad (3.4)$$

Donde $*$ representa la convolución.

Extracción de mínimos y máximos locales:

- *Mínimos locales o valles* ($V(n)$), i será un mínimo local si: $\bar{H}(i) < \bar{H}(i+1)$ y $\bar{H}(i) < \bar{H}(i-1)$.
- *Máximos locales o picos* ($P(n)$), i será un máximo local si: $\bar{H}(i) > \bar{H}(i+1)$ y $\bar{H}(i) > \bar{H}(i-1)$.

Con los umbrales obtenidos se procederá a definir los límites dentro de los cuales se deberá encontrar la región a segmentar. Una región estará definida por el rango de valores que parta del mínimo local anterior a un máximo local y luego, hasta el mínimo local posterior al mismo máximo local. De tal forma que para n regiones: $R(1) \dots R(n)$, se tienen n picos y $n+1$ valles.

En caso de reconocer varias regiones dentro de un mismo histograma, se escogerá aquella que tenga aquel máximo local que presente mayor valor correspondiente a la función histograma. Esto se da porque se parte del supuesto de que en caso la esfera se encuentra dentro de la región de calibración, esta ocuparía la mayor parte de la misma, manifestándose en una mayor concentración de puntos dentro del histograma respectivo. Al tomar la región que presente el máximo local de mayor valor, se espera que represente a la esfera, de no serlo el sistema no se calibrará.

Plano a trabajar:

De efectuar los pasos descritos anteriormente a cada plano de color RGB, se obtendrán 3 regiones pertenecientes a cada uno de estos. Para el desarrollo del sistema se trabajará con aquel plano cuya región presente un mayor valor dentro del histograma. Para ello, se parte de la hipótesis de que un objeto poseerá un componente de color más fuerte en alguno de los 3 planos de color RGB, este análisis pretende identificar aquel plano. El plano con el cual se trabajará será denominado *PlanoQ*, además se definirán los umbrales U_{min} y U_{max} . U_{min} para el punto de inicio (mínimo

local anterior al máximo local) y U_{max} para el punto final (mínimo local posterior al máximo local) de la región local definida dentro del histograma.

En la Figura 3.9 se muestra el análisis realizado a cada uno de los planos del espacio de color RGB a través del cálculo de histogramas.

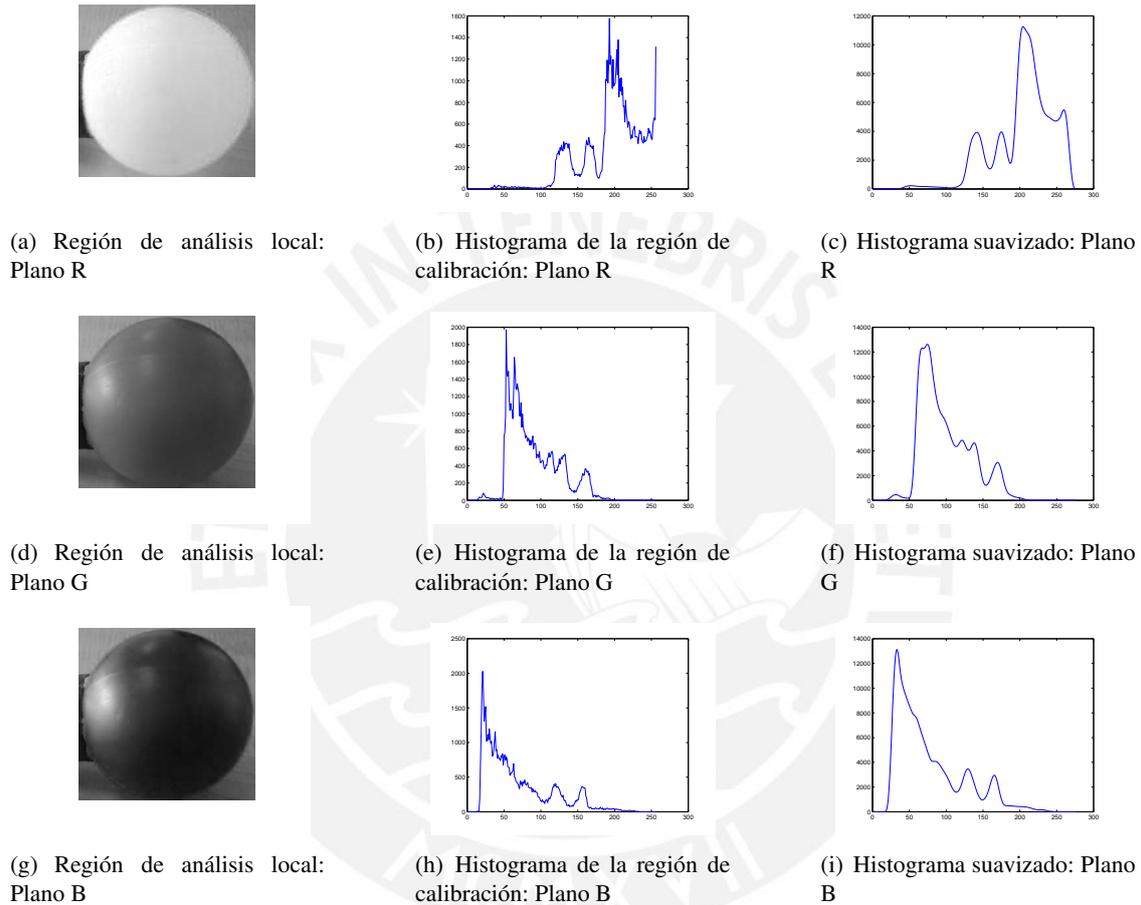


Figura 3.9: Análisis en la región local para cada uno de los 3 planos del espacio de color RGB

Imagen binarizada:

Con el plano definido y los umbrales que se obtuvieron se procederá a umbralizar la región de análisis local. Para ello, se define la imagen binarizada como: $I_{bin}(x, y)$, $x \in [0, T]$, $y \in [0, T]$.

$$I_{bin}(x, y) = \begin{cases} 0, & \text{si } I_{local}(x, y) \Rightarrow U_{min} \quad \text{y} \quad I_{local}(x, y) \leq U_{max} \\ 255, & \text{si } I_{local}(x, y) < U_{min} \quad \text{y} \quad I_{local}(x, y) > U_{max} \end{cases} \quad (3.5)$$

Además, cada píxel de la región dentro de la imagen original que no forma parte de la región local de análisis tendrá el valor de 255 para efectos del proceso de binarización.

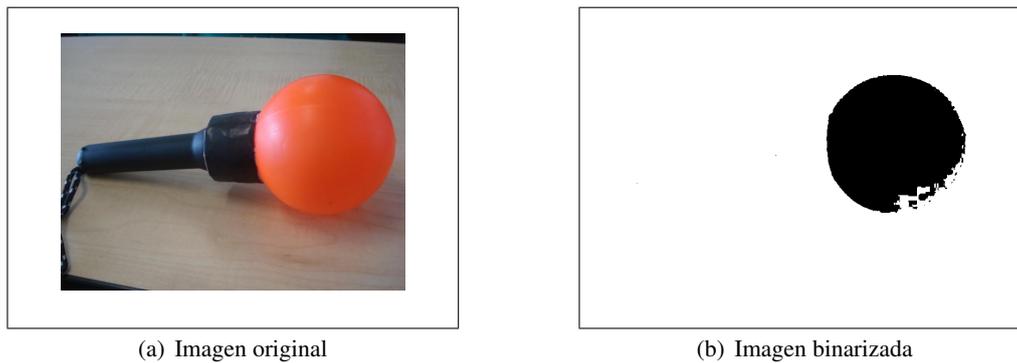


Figura 3.10: Binarización de una imagen con los umbrales obtenidos a partir del análisis de histogramas.

3.3.2. Criterios de calibración

El siguiente paso consiste en identificar si la región segmentada representa o no a la esfera a seguir. Para ello se plantean los siguientes criterios:

1. Porcentaje de ocupación.
2. Forma del área segmentada.
3. Centroide de la región segmentada.

Porcentaje de ocupación:

La región de análisis local se caracteriza por ser una región cuadrada que debe contener a la esfera en su totalidad, de tal forma que en un caso ideal la imagen binarizada que se obtendría se asemejaría a lo mostrado en la Figura 3.11(a). En la Figura 3.11(b) se observa el instante en el que la esfera ocupa la región de análisis, lo cual calibra el sistema.



Figura 3.11: Zona que debe ocupar la esfera en la calibración

La esfera en dos dimensiones se analizará como un círculo. En ese sentido, el porcentaje de ocupación que debe ocupar el círculo dentro del cuadrado está definido por las siguientes ecuaciones:

$$A_{\text{cuadrado}} = T \times T = T^2 \tag{3.6}$$

$$A_{\text{círculo}} = \pi \frac{T^2}{2} \tag{3.7}$$

$$\text{Porcentaje de Ocupación} = \frac{A_{\text{círculo}}}{A_{\text{cuadrado}}} * 100 = 78,5398 \% \tag{3.8}$$

El área segmentada deberá ocupar el 78.54 % del área de la región local para que sea considerada como potencial alternativa a representar la esfera. No obstante, si el sistema se limita exactamente a comparar los parámetros obtenidos con este valor, sin considerar el margen de error propio de la segmentación, probablemente el sistema no llegue a calibrarse aún cuando se den todas las condiciones del caso. Por esta razón, se considera un rango de análisis que estará comprendido entre el 75 % y 80 % del área de ocupación total. El área de la región segmentada deberá estar dentro de dicho rango para que pueda cumplir con este criterio de calibración.

Se escogió dicho rango porque con él se obtuvieron regiones segmentadas que representaban a la esfera de manera mucho mejor. Un rango más amplio incidía en errores como la detección de falsos objetos que no tienen relación con la esfera. Por otro lado, un rango más pequeño hacía al sistema más selectivo, siendo más difícil de calibrar.

Forma del área segmentada:

La información que proporcione el área de la imagen segmentada no será suficiente dado que, aunque habiéndose cumplido el criterio de porcentaje de ocupación, puede que la imagen segmentada no corresponda con la esfera sino con cualquier otro objeto cuya área ocupe el rango de análisis antes mencionado. Por este motivo, se deberá analizar otro parámetro que caracterice a una esfera. El parámetro a analizar para comprobar la correspondencia con la forma de la región segmentada serán los bordes. En el presente estudio se procede a detectar los bordes usando una máscara de Laplace de 3x3 (ver Figura 3.12(a)). El filtro se aplicará para cada píxel de una región de igual tamaño que la máscara propuesta (ver Figura 3.12(b)), teniendo en consideración que al momento de aplicar la máscara sobre de los bordes de la imagen, esta se completará con 0 alrededor de la misma para evitar problemas al momento de las operaciones.

0	-1	0
-1	4	-1
0	-1	0

(a) Máscara de Laplace

P_1	P_2	P_3
P_4	P_5	P_6
P_7	P_8	P_9

(b) Región de 3x3 píxeles sobre la cual se aplicará la máscara

Figura 3.12: Regiones de trabajo para la detección de bordes

La aplicación de la máscara sobre la imagen estará definida por la siguiente ecuación:

$$P_5 = (4 * P_5) - (P_2 + P_4 + P_6 + P_8) \quad (3.9)$$

Como producto de la aplicación de este filtro se obtendrá una matriz donde se almacenarán las coordenadas de la posición de cada uno de los puntos que representan los bordes. Esta información será utilizada para el cálculo aproximado del centro y del radio de la circunferencia. Para ello se aplicará el método de “Least Square Circle Fit” (descrito en [41]), cuyos modelo matemático plantea básicamente lo siguiente:

Dada la función: Ecuación de la circunferencia, se tiene:

$$g(u, v) = (u - u_c)^2 + (v - v_c)^2 - R^2 \quad (3.10)$$

De donde:

u : Diferencia entre la coordenada x de un punto del borde y la media de todas las coordenadas en dicho eje.

v : Diferencia entre la coordenada y de un punto del borde y la media de todas las coordenadas en dicho eje.

u_c : Coordenada x del centro que se desea estimar.

v_c : Coordenada y del centro que se desea estimar.

R : Radio de la circunferencia que se desea estimar.

La ecuación de la circunferencia será derivada con respecto a las variables de las coordenadas en los ejes X, Y y al radio. Las expresiones resultantes serán igualadas a cero, estas ecuaciones son:

$$\sum_i g(u_i, v_i) = 0 \quad (3.11)$$

$$\sum_i u_i g(u_i, v_i) = 0 \quad (3.12)$$

$$\sum_i v_i g(u_i, v_i) = 0 \quad (3.13)$$

El desarrollo de estas ecuaciones permitirá la estimación de las coordenadas y el radio de la circunferencia (esto puede ser revisado en [41]). Básicamente, se busca encontrar aquella varianza que sea mínima, es decir, encontrar aquella curva que aproxime de mejor forma los puntos de los bordes haciendo que la distancia entre estos puntos sea mínima y estimando así el centro de la circunferencia. Partiendo de este análisis se puede calcular aquella diferencia máxima entre los puntos del centro y los puntos del borde de la circunferencia, representando esa distancia a la longitud del radio. De esta manera, se habrán estimado los parámetros de la circunferencia.

Con estos datos se verificará que el centro se aproxime al centro del cuadrado del área de calibración (con un error de ± 5 unidades) y el radio se aproxime al valor de la mitad de la longitud del cuadrado (con un error de ± 2 unidades). Se definieron estos valores como rango de error permitido porque al considerar un rango mayor para cada caso, se obtenían aproximaciones equivocadas.

Centroide de la región segmentada:

El último criterio de análisis corresponde al cálculo del centroide (centro de masa) de la región segmentada, el cual deberá coincidir con el centro del cuadrado (con un error de ± 5 unidades por lo antes mencionado). Este criterio complementa al criterio de la forma del área segmentada, dado que permite corregir posibles errores de estimación que presente el método de “least square circle fit”. Estos errores pueden surgir ante falsas estimaciones que podría realizar el método en mención al tener una cantidad de puntos que no permita realizar una estimación correcta (incluso habiendo cumplido con el primer criterio de calibración). Además, este criterio asegurará que la región segmentada presente un área homogénea.

Finalmente, de estar correctamente calibrado el sistema, se procede a almacenar en memoria las siguientes variables:

- Umbrales mínimo y máximo usados en la segmentación de la región de análisis local.
- Umbral que contiene al índice cuyo valor de la función histograma es el mayor.
- Indicador del plano (R o G o B) con el cual se trabajará.
- Radio y centro de la esfera calibrada.

Con estos datos el sistema estará calibrado luego, se procederá con la segmentación automática de las regiones de análisis en función de los parámetros obtenidos.

3.3.3. Situaciones presentes en la calibración del sistema

Durante el proceso de calibración la esfera puede ser situada en diferentes ubicaciones. De manera general, en la Figura 3.13 se muestran los posibles escenarios donde la esfera estará situada tomando como referencia la región de análisis local. A continuación, se procederá a explicar como reaccionará el sistema ante los casos planteados considerando que para que la esfera sea calibrada debe cumplir con los criterios anteriormente explicados.

En la Figura 3.13(a) se observa que la esfera está dentro de la región de análisis local. No obstante, el sistema no se calibrará dado que se puede estimar que el porcentaje de ocupación del área cuadrada será menor a la requerida.

3.3. CALIBRACIÓN

33

En la Figura 3.13(b) se observa que la esfera no se encuentra dentro de la región de análisis por lo que en primera instancia no cumplirá con el criterio de porcentaje de ocupación.

En la Figura 3.13(c) se observa que la región de análisis está dentro de la esfera. Este caso probablemente ocurra cuando la esfera esté muy cerca de la cámara ocupando gran parte de su rango de visión. En este caso, el porcentaje de ocupación no está dentro de los rangos establecidos, sino por el contrario será mayor, llegando incluso al 100 %.

En la Figura 3.13(d) se observa que la esfera ocupa parcialmente la región de análisis. En este caso se deben resaltar principalmente dos aspectos: 1) Se percibe que el porcentaje de ocupación no está dentro de los límites permitidos. 2) La forma del área segmentada dentro de la región de análisis no corresponde con el de una circunferencia, sino probablemente con la forma de un cuarto de la misma. Ambos aspectos hacen que el sistema no sea calibrado.

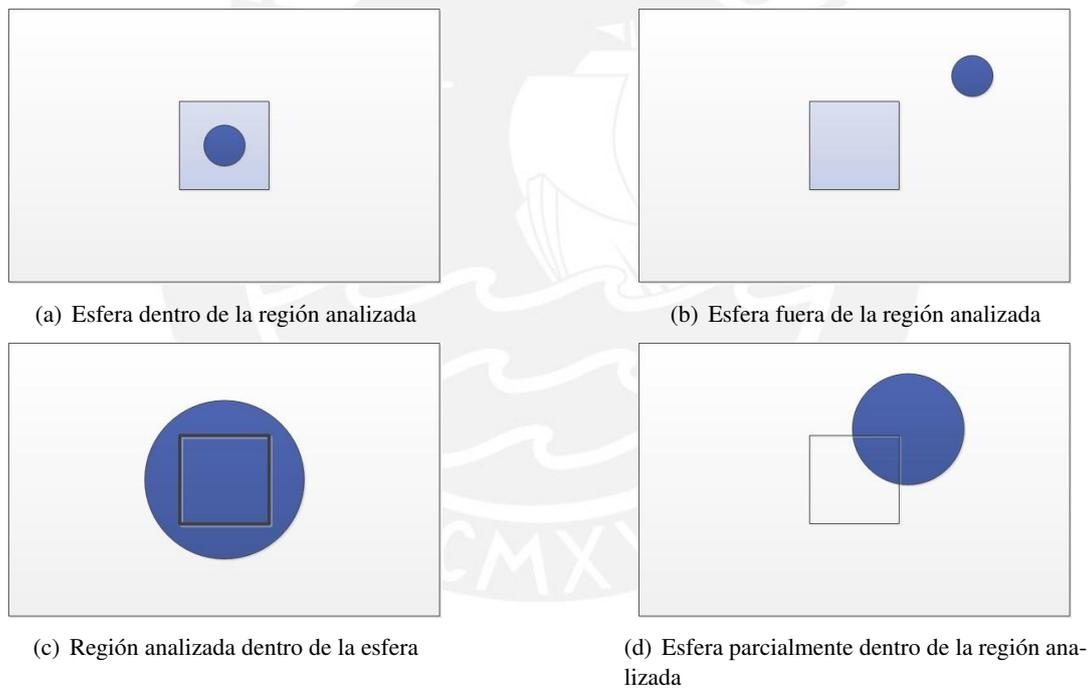
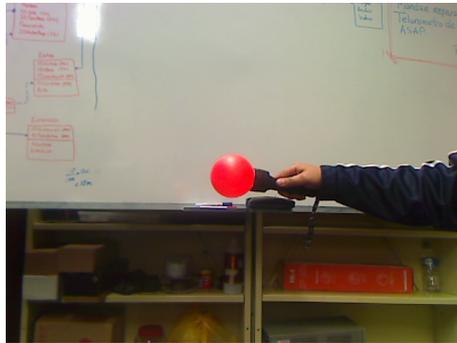


Figura 3.13: Posibles ubicaciones de la esfera en relación a la región de análisis local

Por otro lado, se realizaron pruebas segmentando esferas de los colores naranja, morado y verde dentro de la región de análisis local establecida inicialmente sobre un fondo uniforme de color claro (blanco). Es importante mencionar que el uso de estos colores fueron definidos aleatoriamente.

En la Figura 3.14 se muestran los resultados obtenidos. Se destaca que la esfera de color naranja es la mejor segmentada por el sistema, dado que obtiene los mejores resultados en cuanto a forma y porcentaje (encontrándose dentro del rango establecido). La esfera morada no logra obtener valores de porcentaje de área dentro del rango establecido y es evidente que la esfera verde

no cumple con los requisitos establecidos dado que se pierde y se confunde con los parámetros de color que presenta el fondo.



(a) Cuadro original dentro del cual sucede la calibración: Esfera naranja



(b) Imagen segmentada de la región de análisis local. Porcentaje de ocupación = 75.28 %



(c) Cuadro original dentro del cual sucede la calibración: Esfera morada



(d) Imagen segmentada de la región de análisis local de 3.14(c).



(e) Cuadro original dentro del cual sucede la calibración: Esfera verde



(f) Imagen segmentada de la región de análisis local de 3.14(e).

Figura 3.14: Calibración: Esfera de colores naranja, morado y verde.

Finalmente, cuando el sistema sea activado analizará uno a uno los cuadros obtenidos siguiendo los pasos previamente descritos hasta que encuentre un cuadro que cumpla con los criterios explicados. A partir de ese momento el sistema estará calibrado y se procederá al seguimiento del objeto detectado, pasando previamente por la etapa de segmentación automática.

3.4. Segmentación automática

Uno de los objetivos principales del presente estudio es la segmentación automática de la esfera, lo que implica que el sistema estará en la capacidad de detectar al objeto, reconocerlo y segmentarlo a partir de los datos obtenidos en la calibración, sin necesidad de que haya intervención manual dentro del sistema por parte del usuario. En la Figura 3.15, se observan las fases desarrolladas en la presente etapa.

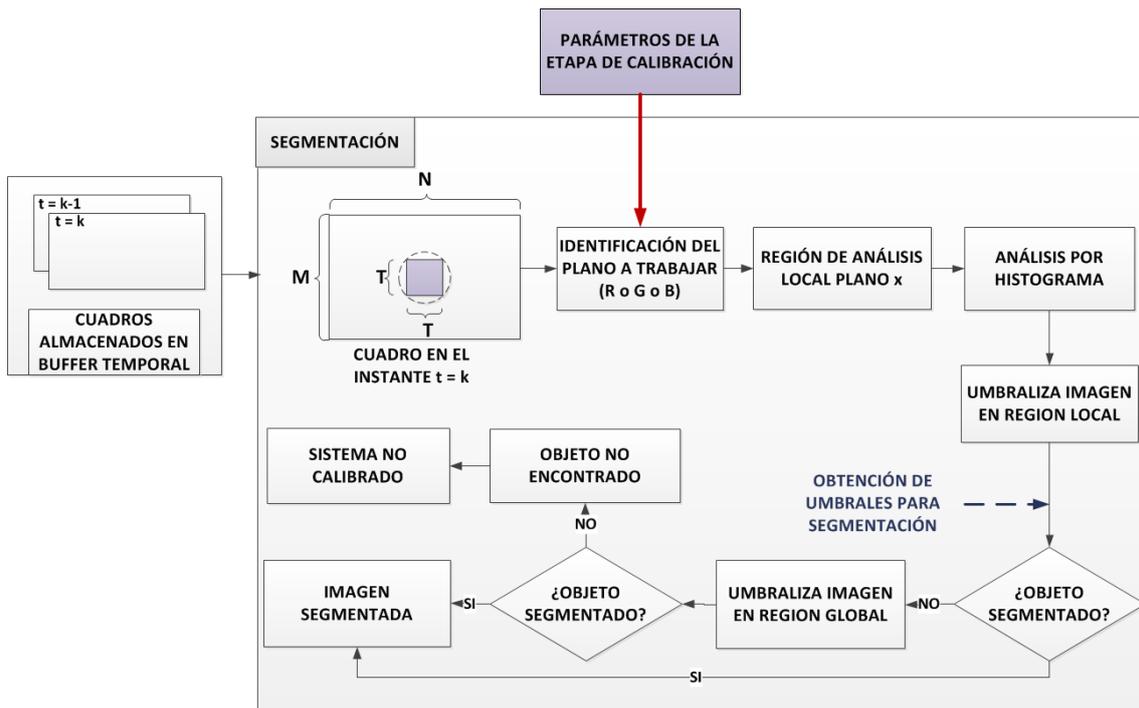


Figura 3.15: Diagrama de bloques: Segmentación automática

Las fases mostradas en la Figura 3.15 están basadas en los mismos modelos teóricos descritos en la etapa de calibración, incluso se usa información obtenida de la misma como se explica a continuación:

3.4.1. Identificación del plano a trabajar

Esta fase se inicia con la obtención de la región local de análisis en sus tres planos (RGB). No obstante, con el objetivo de utilizar menos recursos y hacer que el tiempo de procesamiento sea más rápido, se trabajará con un solo plano. Dicho plano fue identificado en la etapa de calibración.

3.4.2. Segmentación automática de la esfera

Habiendo definido el plano sobre el cual se trabajará, se procederá a obtener los umbrales para la segmentación del objeto a través del método de histograma. Para ello, se usará un dato obtenido de la etapa de calibración que se denominará índice de calibración ($I_{calibracion}$). Este dato es aquel

índice que presentaba un mayor valor dentro del histograma obtenido cuando la esfera fue calibrada. Además, este índice representa la información de color en el plano trabajado de la esfera detectada.

Para este análisis se parte del supuesto de que el $I_{calibración}$ definirá cual es la región que representa a la esfera. Para ello, se hallarán los mínimos locales anterior y posterior a dicho índice dentro del histograma de análisis. Esto se puede observar en la Figura 3.16. Cabe resaltar que el máximo local de la región que represente a la esfera no tiene necesariamente por qué coincidir con el $I_{calibración}$.

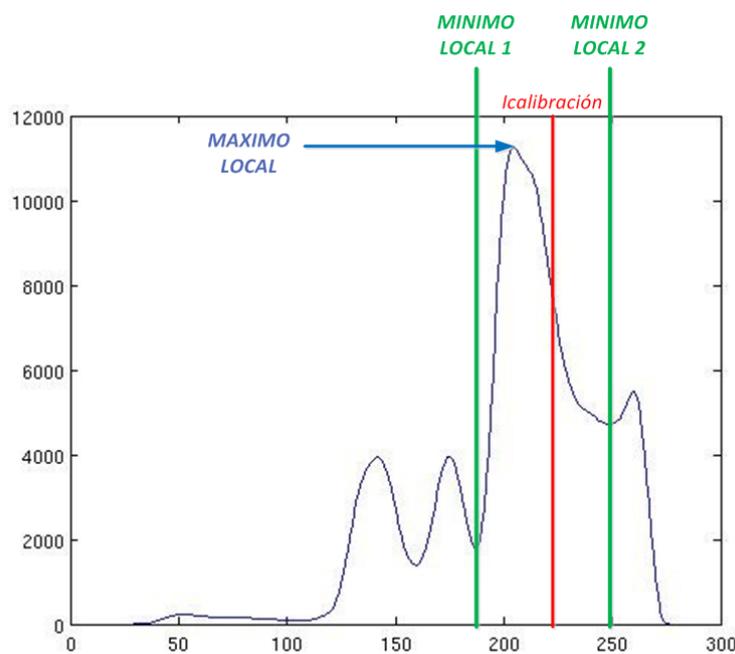


Figura 3.16: Detección de umbrales en la segmentación automática

Con los umbrales que definen la región de análisis se procederá con la segmentación del objeto con el desarrollo de la ecuación 3.3.1, para luego obtener los bordes de dicha región aplicando la máscara de Laplace. Dicha información será analizada en la fase de seguimiento.

3.4.3. Segmentación de la imagen en la región global de análisis

Puede darse el caso que habiendo obtenido los umbrales para la segmentación y umbralizando la región de análisis local no se haya encontrado el objeto de análisis. Esto puede darse en alguno de los siguientes casos:

1. Ante condiciones de iluminación que cambien bruscamente (en este caso ya no servirían los parámetros de color que se hallaron en la etapa de calibración).

2. El objeto haya salido fuera del rango de observación de la cámara tal que no pueda ser hallada ni analizando la imagen en su totalidad. En casos como este el sistema se descalibrará, por lo que se hará necesario ubicar el objeto dentro del rango de visión de la cámara y calibrar nuevamente el sistema.
3. Considerando que el sistema está trabajando a una velocidad de 10 cuadros por segundo, es decir, que entre cada cuadro hay un tiempo de 100 ms, si el objeto de análisis es movido a una velocidad tal que cuando se obtenga el siguiente cuadro, el objeto no se encuentre dentro de la región de análisis local pero sí cerca de dicha región; entonces, el objeto podrá encontrarse analizando una región global de análisis siguiendo los pasos descritos en la Figura 3.17.

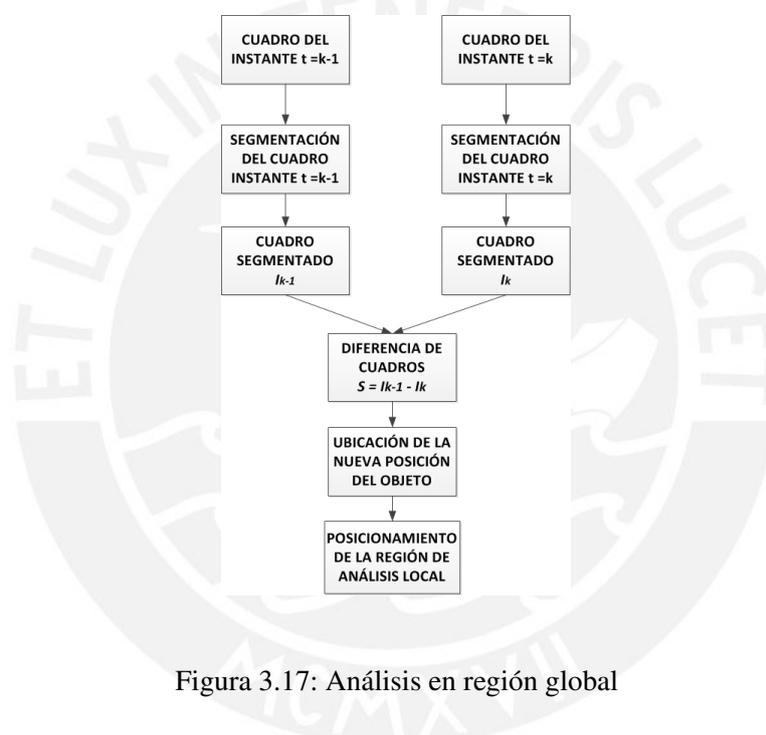


Figura 3.17: Análisis en región global

Para el análisis de la región global se tomará en cuenta lo siguiente:

- El tamaño de la región de análisis global está definido por el conocimiento a priori de la velocidad máxima a la que puede desplazarse la esfera dentro de dos cuadros consecutivos de una señal de video. Con el conocimiento de este dato y de la resolución espacial se podrá delimitar una región de búsqueda mayor a la región de análisis local, sin llegar a ser necesariamente toda la imagen. Como se recuerda se busca evitar el análisis de toda una imagen, lo que implicaría mayor tiempo de procesamiento y probabilidad de error.
- Se analizarán las dos últimas regiones globales de análisis que se obtuvieron cuando el objeto de interés dejó de ser encontrado en la región de análisis local (instantes $t = k - 1$ y $t = k$). El análisis se efectuará a través del método de histograma con los datos de los umbrales y plano a trabajar obtenidos en la etapa de calibración. Esto permitirá segmentar las regiones globales de análisis en su totalidad.

- Al segmentar las regiones globales de análisis es probable que se encuentren regiones que no pertenezcan a la esfera pero que compartan parámetros de color similares. Para eliminar este efecto se parte del hecho de que en su mayoría estas regiones son estáticas, es decir, no tendrán mayores variaciones en el tiempo porque corresponden a objetos que están en una misma posición. Por esta razón, si se restan las imágenes segmentadas de ambos cuadros, las regiones correspondientes a esos objetos estáticos deben eliminarse.
- Hasta el momento se habrá obtenido un cuadro que represente la diferencia de las imágenes segmentadas de los dos últimos cuadros obtenidos. Se espera que en este cuadro se representen dos regiones que corresponden a la posición de la esfera para los instantes $t = k - 1$ y $t = k$, y algunas pequeñas sub regiones que pueden corresponder a diversos objetos o puntos las cuales serán ignoradas cuando sea calculado el centroide. Entonces, como es de interés obtener la posición de la esfera en el instante $t = k$, se procederá a ignorar la región de la esfera para el instante $t = k - 1$, cuya posición sí es conocida.
- Finalmente, al haber obtenido la región correspondiente a la esfera para el instante $t = k$ se calculará el centroide y sobre este se ubicará la región de análisis local con el último dato del radio correctamente obtenido. Aquí, se procederá a segmentar la región de análisis local para luego obtener los bordes y con ello estimar los parámetros de centro y radio de la esfera. Las Figuras 3.18, 3.19, 3.20, 3.21 muestran algunos ejemplos del método explicado:

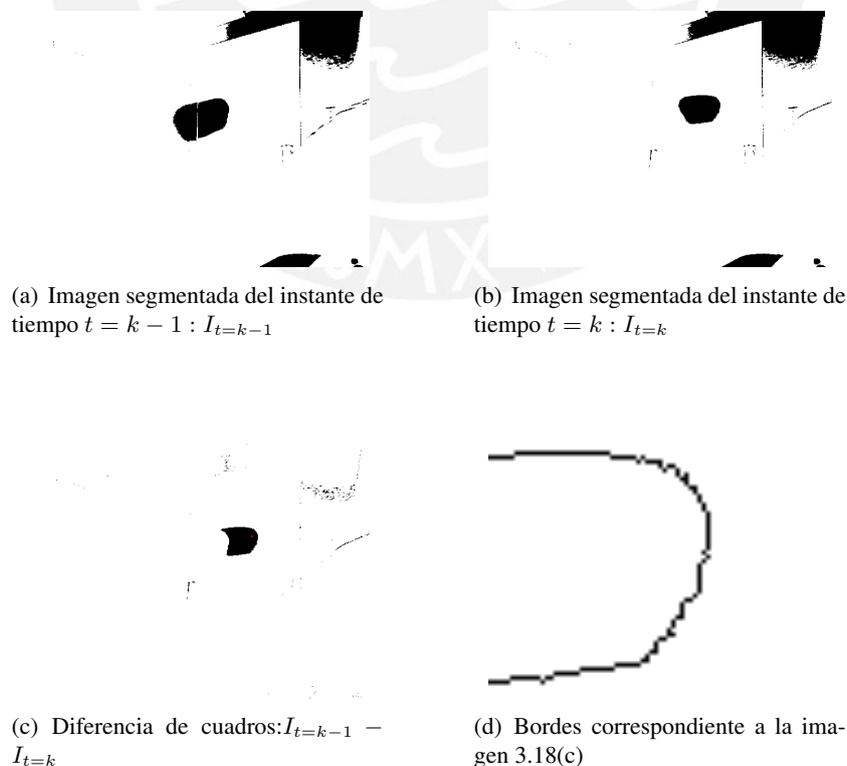


Figura 3.18: Ejemplo 1: Seguimiento de la esfera a través regiones globales de análisis

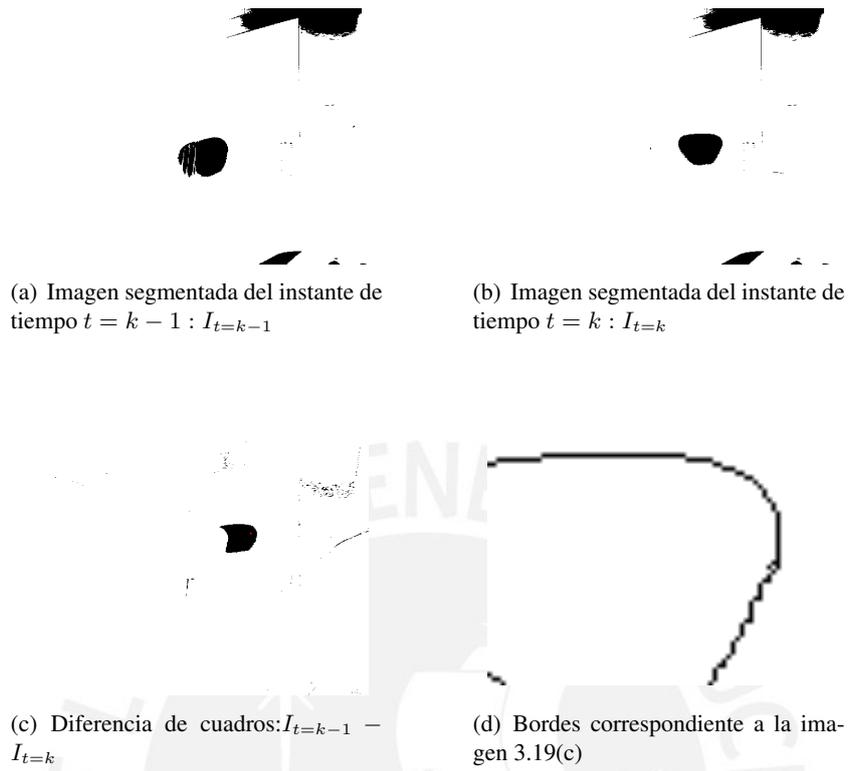


Figura 3.19: Ejemplo 2: Seguimiento de la esfera a través regiones globales de análisis

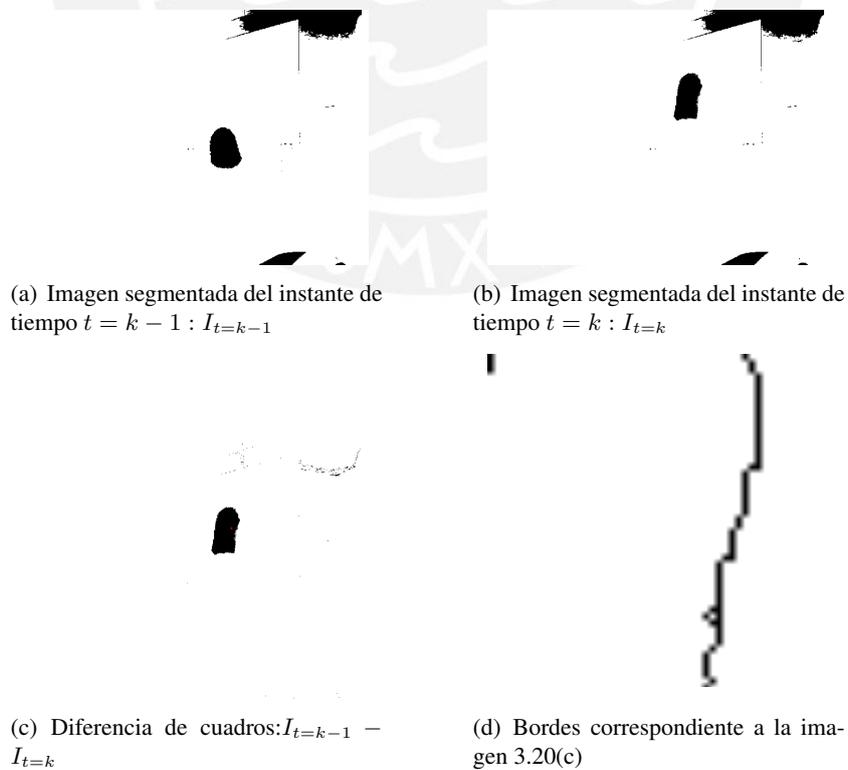


Figura 3.20: Ejemplo 3: Seguimiento de la esfera a través regiones globales de análisis

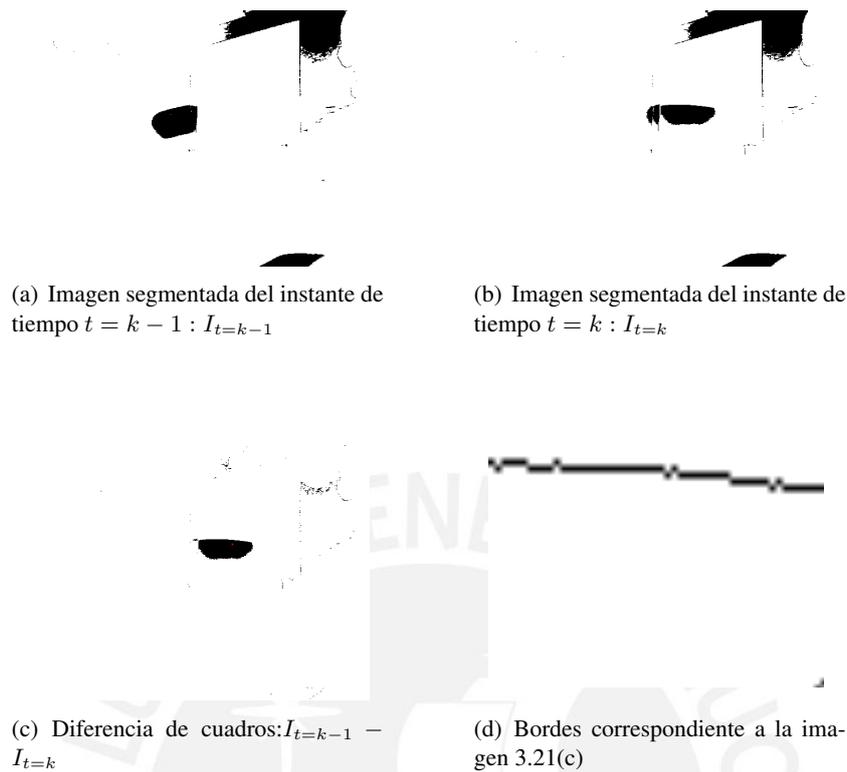


Figura 3.21: Ejemplo 4: Seguimiento de la esfera a través regiones globales de análisis

En las Figuras 3.18 y 3.19 se observa que los bordes obtenidos al final del análisis planteado permitirán definir la forma de una esfera, probablemente no del mismo tamaño que la original, pero sí uno que se aproxime. No obstante, en las Figuras 3.20 y 3.21 los bordes obtenidos no garantizan que el algoritmo “least square circle fit” pueda lograr una representación correcta de la circunferencia, cayendo en la necesidad de calibrar nuevamente el sistema, tal como se pudo comprobar luego durante las pruebas.

Finalmente, en la Figura 3.22, se muestran las imágenes segmentadas dentro de la región de análisis local para tres instantes de tiempo consecutivos posterior a la calibración. Esto, como resultado de haber aplicado los pasos descritos en la Figura 3.15. Como se observa, se logra identificar de forma adecuada los bordes de la región, lo que permitirá un mejor cálculo de los parámetros del centro y radio para el seguimiento.

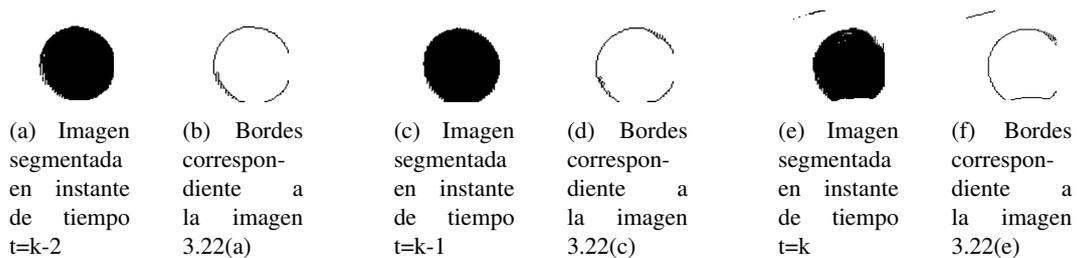


Figura 3.22: Segmentación automática de la esfera de color naranja

3.5. Seguimiento del objeto

En esta etapa se procederá a obtener los parámetros correspondientes al centro y radio de la circunferencia estimada partiendo del análisis de bordes de la región segmentada. Con estos parámetros se definirá la posición y tamaño de la nueva región de análisis local (como se muestra en la Figura 3.23), la cual será usada en el desarrollo de diversas aplicaciones.

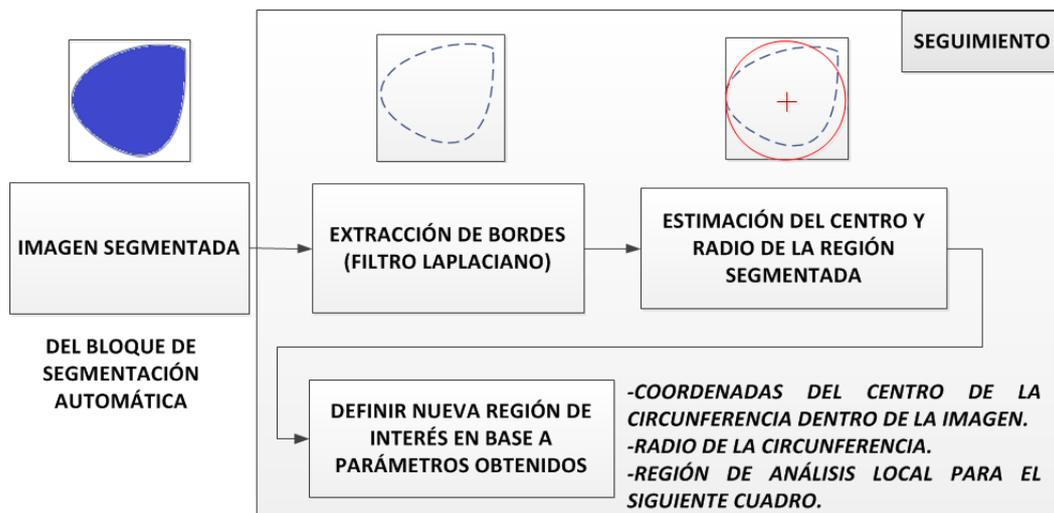


Figura 3.23: Diagrama de Bloques: Seguimiento

3.5.1. Extracción de bordes

Para esto se tiene la imagen segmentada dentro de la región de análisis local producto de la fase de segmentación automática. Luego, se procederá a obtener los bordes de dicha región usando la máscara de Laplace como se describió en la etapa de calibración.

3.5.2. Estimación del centro y radio de la región segmentada

Con la información de los bordes se procederá a estimar el centro y radio a los que corresponderían los puntos de dicha región usando el método de “least square circle fit” [41] descrito en la etapa de calibración, desarrollando para ello las ecuaciones 3.10, 3.11, 3.12 y 3.13.

Los parámetros obtenidos pueden que no se encuentren necesariamente dentro de la región de análisis local, lo que indica la necesidad del establecimiento de una nueva región de análisis partiendo de los parámetros obtenidos. Una vez se haya realizado esto, se habrá logrado el seguimiento del objeto de interés. En la Figura 3.24 se observa el análisis que se realizó a partir de la información de bordes dentro de una región de análisis local para la estimación de la circunferencia correspondiente.

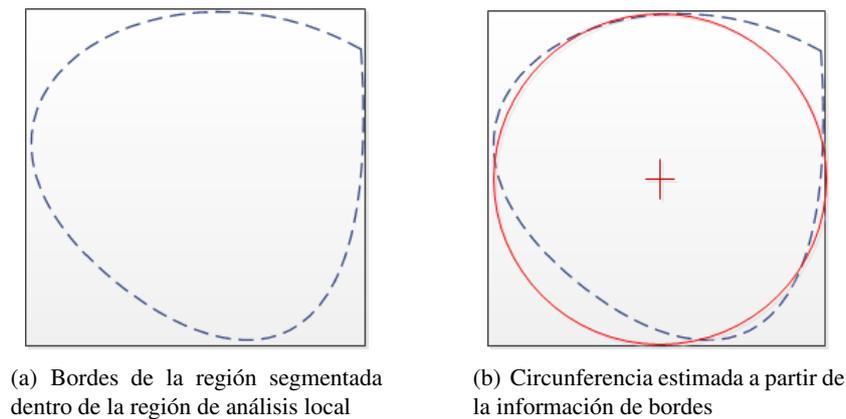


Figura 3.24: Estimación de la circunferencia dentro de una región local de análisis

No obstante, no todas las regiones a estimar se presentarán como el caso de la Figura 3.24, en otros casos la circunferencia a estimar abarcará regiones fuera de la región de análisis local como es el caso de la Figura 3.25.

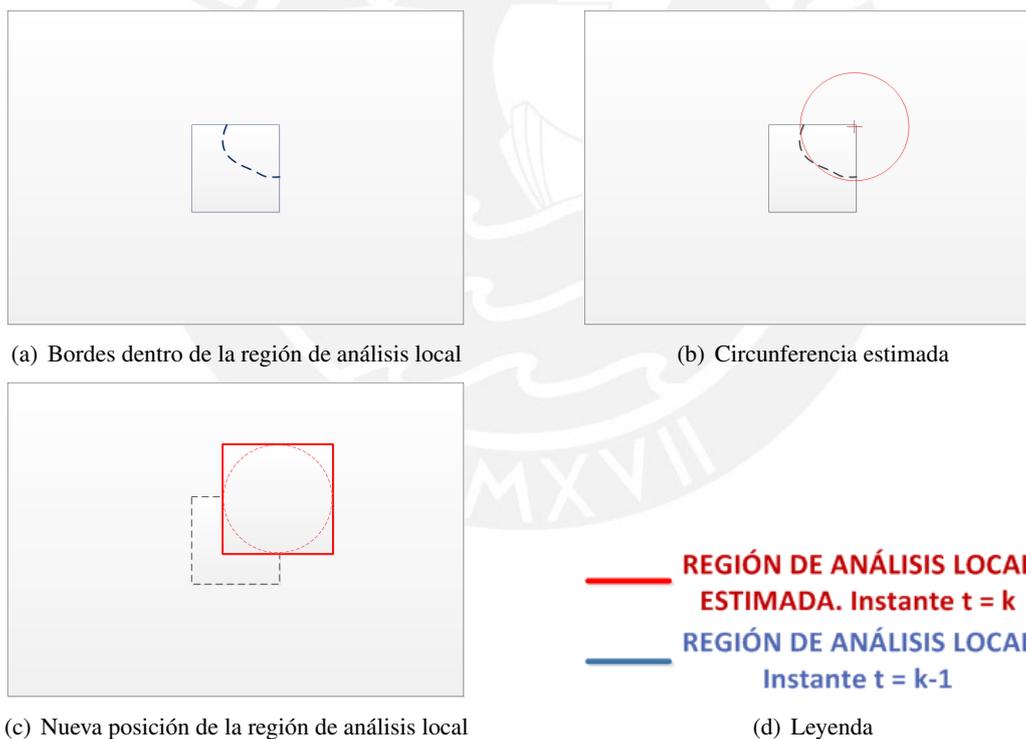


Figura 3.25: Seguimiento del objeto esférico analizando los bordes de la región segmentada

En la Figura 3.25(a) se muestra la posición inicial de la región de análisis local y el borde de la región segmentada dentro de dicha región. En la Figura 3.25(b) se observa la estimación de la circunferencia a partir de los datos de bordes, para finalmente, en la Figura 3.25(c) y teniendo como referencia los parámetros de la circunferencia estimada, definir la posición y tamaño de la nueva región de análisis local. Se cumple así con el objetivo del seguimiento del objeto.

En conclusión, se analizará la variación de la región de análisis local en dos aspectos: tamaño y posición. El primero estará ligado a comprobar si la región creció o se hizo más pequeña, lo que sería equivalente a notar si el objeto se acercó o alejó con respecto de la posición de la cámara. El segundo aspecto estará ligado al cambio de posición de la esfera, ello definirá la dirección a la cual el objeto se ha desplazado. Estos datos serán importantes de cara al desarrollo de las diversas aplicaciones.



Capítulo 4

EVALUACIÓN DEL SISTEMA DE DETECCIÓN Y SEGUIMIENTO DE UN OBJETO ESFÉRICO

En este capítulo se presentan los diversos resultados que se obtuvieron al aplicar el modelo teórico descrito en el capítulo 3 a través de diversas aplicaciones que involucran el seguimiento de un objeto esférico.

Para medir la eficiencia del sistema, en principio, se analiza la exactitud que se obtuvo al realizar el seguimiento a lo largo de las aplicaciones realizadas. Para ello, se analiza cómo reacciona el sistema ante cambios a una profundidad constante y variable. Además, se analiza la velocidad de desplazamiento a la cual se deberá mover el controlador manual con el objetivo de que siempre pueda ser analizado correctamente por el sistema desarrollado. También, se medirá el tiempo de procesamiento de la ejecución del algoritmo descrito en el presente estudio, lo que a su vez, delimitará la velocidad (fps) a la cual se recomienda trabajar la cámara. Finalmente, se definen los límites bajo los cuales el sistema operará adecuadamente, según la magnitud de error que se encuentre.

4.1. Descripción de las aplicaciones desarrolladas

Se han desarrollado 4 aplicaciones, las que procederemos a describir a continuación:

4.1.1. Aplicación 1: Sobreposición de una circunferencia

En la presente aplicación se busca sobreponer a la esfera una circunferencia que posea las mismas dimensiones (píxeles) y posición que la esfera vista en la interfaz gráfica. Una muestra de la presente aplicación la observamos en la Figura 4.1.



Figura 4.1: Aplicación 1: Sobre imposición de una circunferencia

4.1.2. Aplicación 2: Trayectoria del centro de la circunferencia

En la presente aplicación se busca dibujar la trayectoria del centro de la esfera a lo largo del tiempo mientras dure la ejecución de la aplicación, como se muestra en la Figura 4.2.

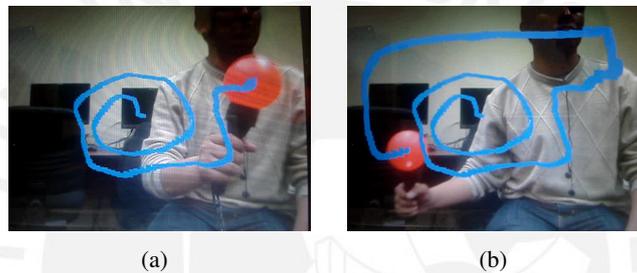


Figura 4.2: Aplicación 2: Trayectoria del centro de la circunferencia

4.1.3. Aplicación 3: Sobreposición del logo de la PUCP

En la presente aplicación se busca sobreponer a la esfera el logo de la PUCP, tal que esta posea las mismas dimensiones (píxeles) y la misma posición que la esfera vista en la interfaz gráfica durante la ejecución de la aplicación. En la Figura 4.3 se aprecia un ejemplo de lo descrito.



Figura 4.3: Aplicación 3: Sobreimposición del logo de la PUCP

4.1.4. Aplicación 4: Seguimiento de la trayectoria de la circunferencia

En la siguiente aplicación se busca a partir de la visualización de una imagen en color azul - y cuando el sistema sea calibrado - limpiar aquellos espacios que representan la trayectoria de la esfera en su totalidad, observando así la imagen original del cuadro (ver Figura 4.4). Esta

aplicación está inspirada en aquellos espejos o ventanas que producto de la suciedad no permiten visualizar lo que hay a nuestro alrededor y que, al limpiarse, permite ver lo que hay en realidad.

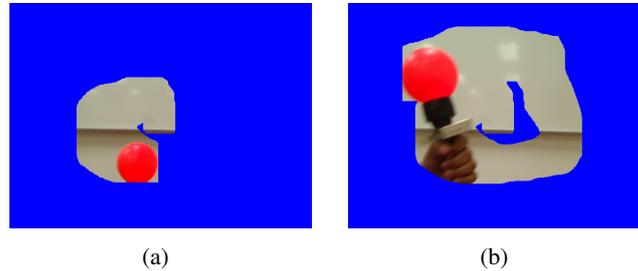


Figura 4.4: Aplicación 4: Seguimiento y limpieza de la trayectoria de la circunferencia

4.2. Consideraciones de las pruebas realizadas

4.2.1. Montaje de los dispositivos para realización de las pruebas

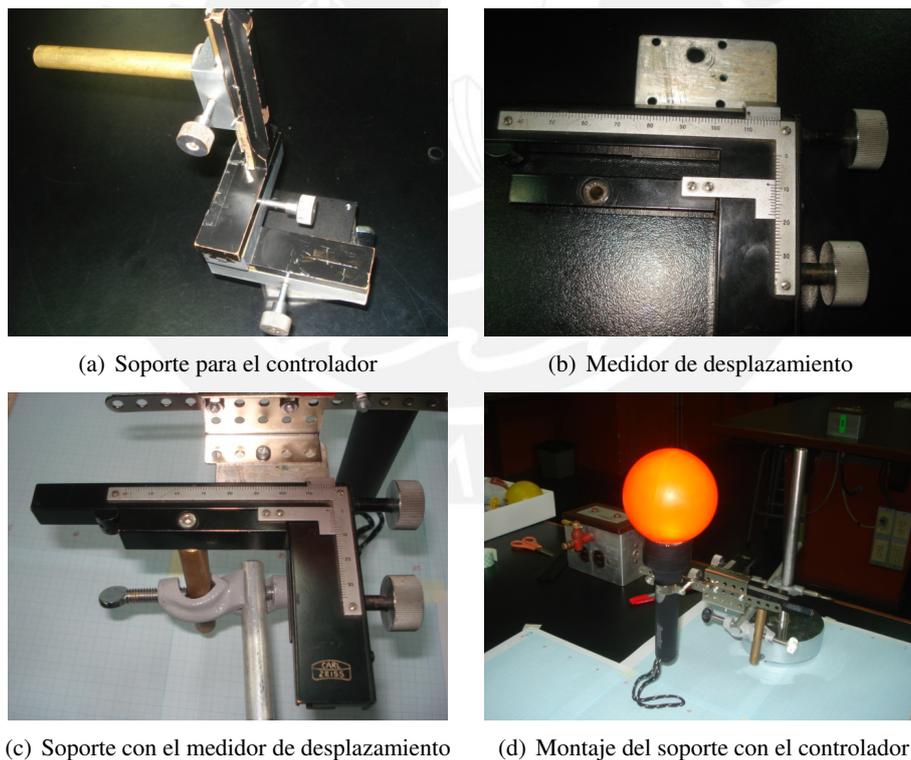


Figura 4.5: Montaje de estructura para control de movimiento del controlador manual

A continuación se describirá qué herramientas y/o procedimientos se usaron con el objetivo de medir los desplazamientos realizados con el controlador manual, para ello, este controlador fue colocado sobre una pequeña estructura de madera (Figura 4.5(a)). Esta estructura, proporcionada por el Laboratorio de Física de la PUCP, se caracteriza por permitir fijar en posiciones establecidas el controlador manual. Además, esta estructura posee un dispositivo de la marca Carl Zeiss [42]

que permite desplazar el controlador manual en distancias de hasta un máximo de 7 cm (Figura 4.5(b)). Este dispositivo es conectado al soporte principal (Figura 4.5(c)), donde además el controlador manual será colocado y fijado (Figura 4.5(d)). Así se tendrá el dispositivo listo para ser utilizado y desplazado.

Además, se usó papel milimetrado colocado sobre una mesa dentro del laboratorio de Física de la PUCP con el objetivo de medir la distancia entre la cámara y el controlador manual (Figura 4.6(a)). Es importante resaltar que no solo se realizaron pruebas teniendo al controlador manual en posición vertical (Figura 4.5(d)), sino esta también fue colocada en posición horizontal como se observa en la Figura 4.6(b). Después, se procedió a colocar la cámara y el soporte sobre el papel debidamente marcado usando cinta métrica. Es así que finalmente se obtuvo una disposición física de los dispositivos que permiten medir la exactitud del sistema (ver Figuras 4.6(c) y 4.6(d)).

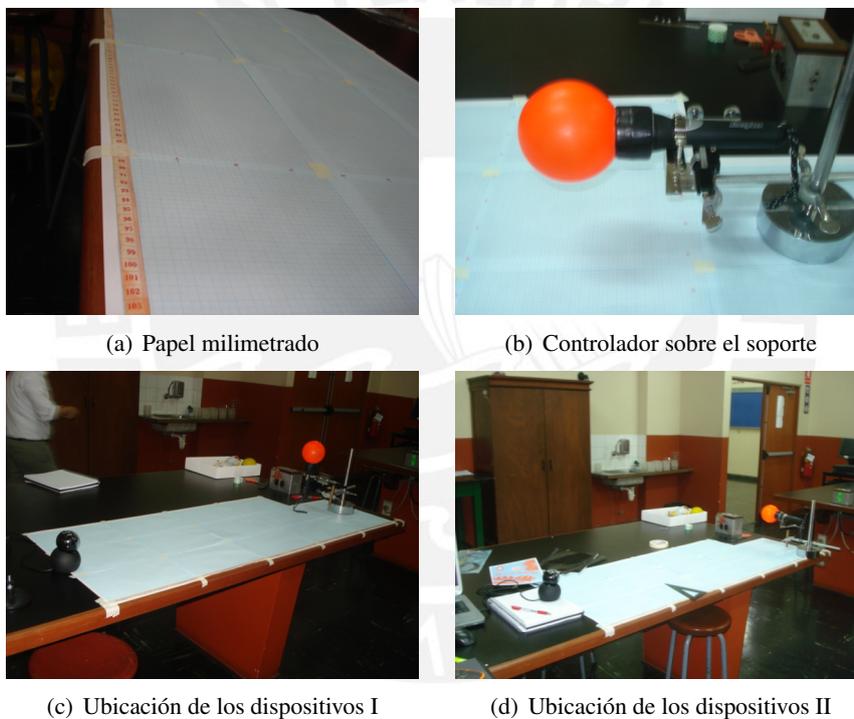


Figura 4.6: Disposición física de dispositivos para realización de pruebas

4.2.2. Condiciones bajo las cuales se realizaron las pruebas

En este punto es importante entender las condiciones sobre las cuales se realizaron las pruebas, lo que a posteriori determinará las condiciones sugeridas sobre las cuales se recomienda ejecutar el sistema desarrollado.

- Ambiente de desarrollo: Las pruebas se realizaron en el Laboratorio de Física de la PUCP (2do. Piso).
- Iluminación del ambiente: Las pruebas fueron desarrolladas en un ambiente cuyo nivel de iluminación corresponde al de un estándar de oficina. En las pruebas desarrolladas, por lo

general, se tenía el ambiente parcialmente iluminado. En otros casos, se hizo con las luces totalmente apagadas, lo que luego permitió constatar el correcto funcionamiento del sistema en ambas condiciones. Además, se tuvo la oportunidad de verificar la operatividad del sistema en otros ambientes de oficina que presentaban mejor iluminación y el sistema respondió también de forma positiva. Finalmente, se sugiere emplear el sistema en condiciones de iluminación estándar de oficina o similares.

- **Computador:** Las pruebas fueron realizadas sobre el computador descrito en el capítulo 3 (ver Figura 3.5). Considerando que en la actualidad existen computadores más potentes, esto nos hace suponer que el sistema ejecutado en un computador más potente permitiría obtener mejores resultados en aspectos como la velocidad de respuesta.
- **Sistema Operativo:** El sistema fue desarrollado sobre el sistema operativo GNU/Linux 2.6.38-8 generic, versión # 42 de la distribución Ubuntu 11.04.
- **Software de Programación:** El sistema se desarrolló en lenguaje C, haciendo uso de las librerías del sistema y de la librería ffmpeg.
- **Cámara:** Se usó una cámara Logitech (ver Figura 3.4) compatible con el sistema.
- **Velocidad de movimiento:** Las pruebas desarrolladas fueron realizadas a una velocidad baja, la cual no excede el 1 m/s (más adelante se explicará la obtención de este ratio), esto dado porque el dispositivo que permite medir el desplazamiento del sistema (Figura 4.5(b)) no permitía desplazamientos más rápidos del controlador.

4.2.3. Metodología de las pruebas realizadas

Las pruebas fueron realizadas con el objetivo de medir principalmente dos aspectos: exactitud del sistema y tiempo de procesamiento de las aplicaciones desarrolladas. Para la medición de la exactitud del sistema se hizo énfasis en analizar la respuesta del sistema en dos situaciones. Cuando a una profundidad constante se realizan desplazamientos horizontales, y cuando se varía la profundidad (acercar o alejar el controlador) manteniendo el objeto en el mismo punto en las coordenadas “X” y “Y” viéndolo frontalmente. La medición del tiempo de procesamiento del sistema entre cuadro y cuadro permitirá verificar que el sistema pueda trabajar al menos a 10 fps. (cómo se planteó en los objetivos del presente trabajo) bajo distintas condiciones y aplicaciones desarrolladas.

Para ello, se realizaron alrededor de 50 pruebas en total, donde las 4 aplicaciones desarrolladas fueron sometidas a diversas pruebas y diferentes distancias entre el controlador y la cámara de video, por lo general, a distancias frontales de 20, 45, 90 y 120 cm. En algunos casos las pruebas fueron desarrolladas con las luces totalmente apagadas; en otras, con las luces del lugar parcialmente encendidas; y en otras, con las luces encendidas en su totalidad. En cualquiera de estos casos, el sistema respondió de forma positiva.

A continuación se muestran los resultados de la ejecución de las aplicaciones desarrolladas bajo las condiciones descritas. Adicionalmente, se colocarán algunas imágenes representativas de algunas de las aplicaciones lo cual permitirá un mejor entendimiento de lo que se realizó en cada prueba. En caso se desee revisar la totalidad de contenido multimedia (videos o fotos) de las pruebas realizadas y el detalle de las mismas, estas se encuentran disponibles en el ANEXO 1.

4.3. Análisis de exactitud del sistema

En la presente sección, y con ayuda de los instrumentos proporcionados por el laboratorio de Física de la PUCP, se mide la exactitud que presenta el sistema. Para ello, se desea medir la respuesta del sistema en las tres dimensiones conocidas: “X”, “Y” y “Z”, relacionando esta última con la profundidad; por eso - y tal como se mencionó - se medirá la respuesta del sistema a una profundidad constante y luego a una profundidad variable.

4.3.1. Profundidad constante - Movimiento en el eje “X” y eje “Y” constante

Las pruebas realizadas a una profundidad constante permite medir la exactitud que presenta el sistema para movimientos en los ejes “X” y “Y”, es decir, en dos dimensiones y para un plano. No obstante, los instrumentos y el medidor de desplazamiento proporcionado por el laboratorio de Física solo permitió medir los movimientos en el eje “X” (Movimientos horizontales: izquierda y derecha viendo frontalmente al usuario).

Considerando la lógica de programación desarrollada y que la resolución espacial es la misma tanto para el eje “X” como para el eje “Y” a una misma profundidad, es que los resultados que se obtengan para el eje “X” también se generalizarán para el eje “Y”. Finalmente, cabe resaltar que todos los desplazamientos realizados son de 7 cm a lo largo del eje “X”. A continuación se procederá a describir las pruebas realizadas:

Pruebas realizadas con luz encendida

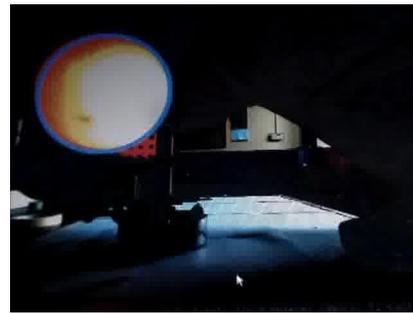
En la Figura 4.7 se muestra dos instantes de tiempo para el desarrollo de las aplicaciones 1 y 3, donde la esfera es ubicada a 20 cm de distancia de la cámara. En ambas aplicaciones se observa como el objeto se ha desplazado horizontalmente.

También se realizaron pruebas a una profundidad de 90 cm de la cámara. En la Figura 4.8 se observa el desarrollo de la aplicación 2 para dos instantes de tiempo distintos, en dicha aplicación se puede observar el trazo de la trayectoria descrita.

Finalmente, se realizaron pruebas a una profundidad de 115 cm. En la Figura 4.9 se muestra la ejecución de la aplicación 3, aunque en este caso - y por la profundidad - el desplazamiento es un poco más complicado de percibir en la interfaz gráfica.



(a) Aplicación 1: Instante inicial



(b) Aplicación 1: Objeto desplazado

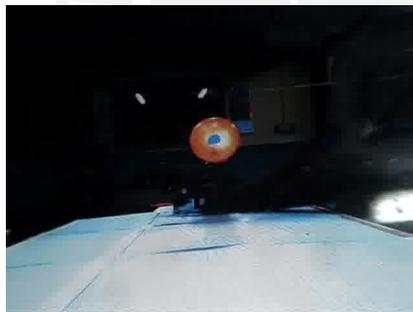


(c) Aplicación 3: Instante inicial



(d) Aplicación 3: Objeto desplazado

Figura 4.7: Aplicaciones 1 y 3: Movimiento del objeto a 20 cm



(a) Aplicación 2: Instante inicial



(b) Aplicación 2: Objeto desplazado

Figura 4.8: Aplicación 2: Movimiento del objeto a 90 cm



(a) Aplicación 3: Instante inicial



(b) Aplicación 3: Objeto desplazado

Figura 4.9: Aplicación 3: Movimiento del objeto a 115 cm

Pruebas realizadas con luz apagada

En la Figura 4.10 se muestra la ejecución de las aplicaciones 3 y 4 a una profundidad de 45 cm de la cámara de video cuando la luz del ambiente de trabajo fue apagada.

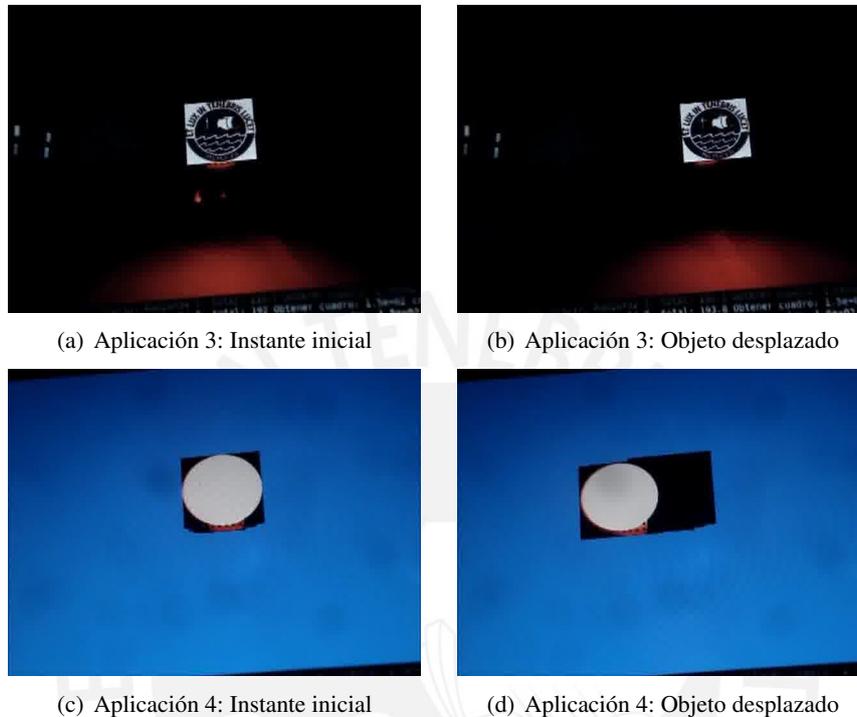


Figura 4.10: Aplicaciones 3 y 4: Movimiento del objeto a 45 cm en ambiente oscuro

En la Figura 4.11 se muestra el comportamiento de la aplicación 1 al ser ejecutada en un ambiente oscuro a 85 cm de distancia de la cámara de video.

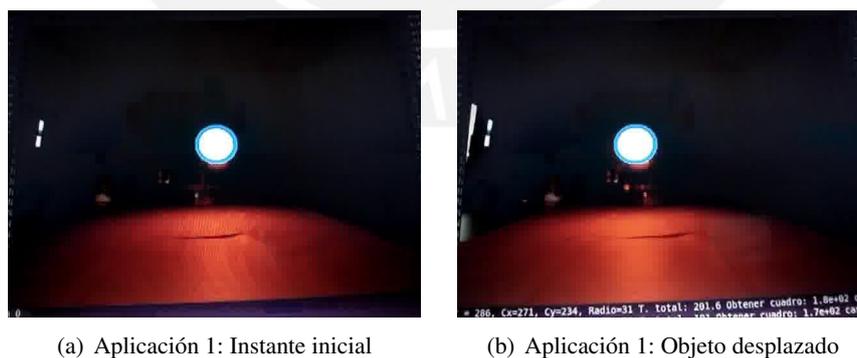


Figura 4.11: Aplicación 1: Movimiento del objeto a 85 cm en ambiente oscuro

Resultados

A continuación, se muestran los resultados obtenidos tras haber procesado la data del sistema, para esto se analizan las variaciones experimentadas a lo largo de los tres ejes.

Análisis en el eje “X”:

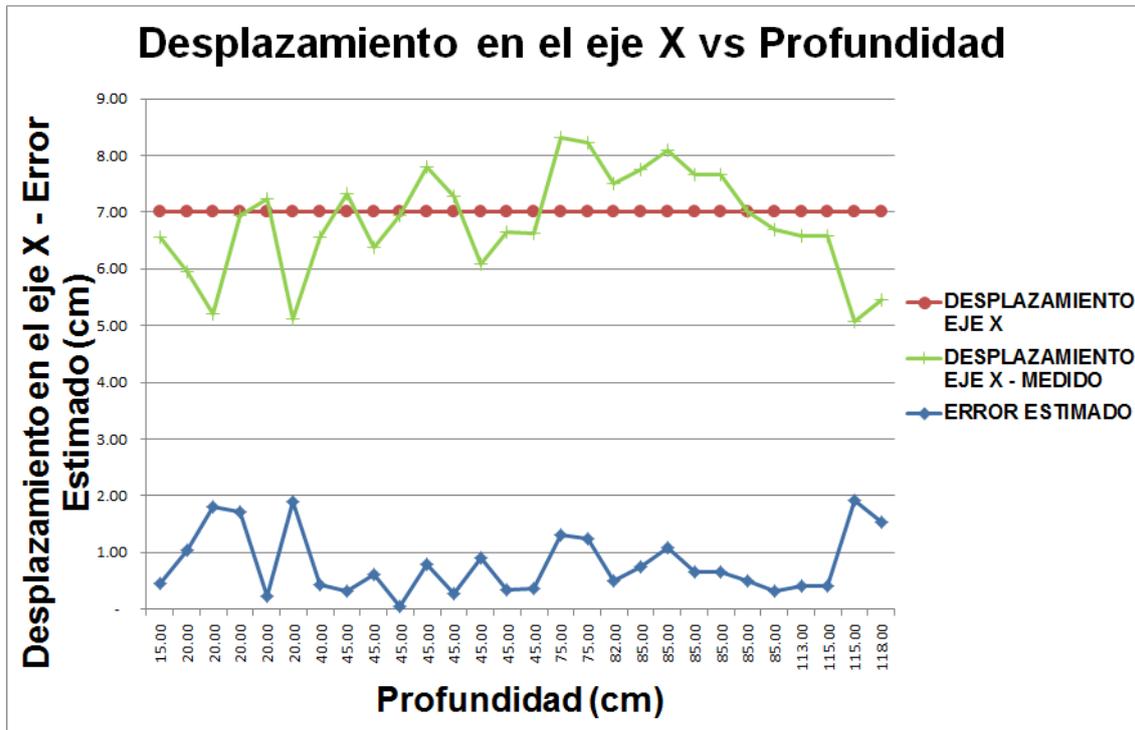


Figura 4.12: Comportamiento del sistema en el eje “X”

El desplazamiento que se experimenta a lo largo del eje “X” para las diversas aplicaciones es de 7 cm. En la Figura 4.12 se observa el comportamiento de la variable desplazamiento en el eje “X” tras la ejecución de las diversas aplicaciones para distintas profundidades. Como se muestra, todas las pruebas tienden a registrar un desplazamiento de 7 cm con un margen máximo de error de 2 cm. No obstante, si limitamos el trabajo del sistema a una profundidad de entre 40 cm y 115 cm se puede observar un error máximo de 1.5 cm.

Análisis en el eje “Y”:

Al no mover la esfera de forma intencional a lo largo del eje “Y”, se esperaría que el desplazamiento sea de 0 cm. Sin embargo, y para efectos de las pruebas realizadas, esto no se ha dado así tal como se puede apreciar en la Figura 4.13, donde se observa que para profundidades menores e iguales a 45 cm el punto en el eje “Y” se encuentra entre 3 y 5 cm por encima de la posición inicial de la calibración. Esto se da porque durante la calibración la esfera no se encuentra en el mismo eje focal que la cámara de video, existiendo cierta oblicuidad entre el plano de la imagen y el plano físico XY. Por este motivo, cuando la esfera se acerca a la cámara, esta tiende a colocarse arriba de la posición inicial (un ejemplo de este tipo de casos se observan en las Figuras 4.7 y 4.10). Por el contrario, cuando la esfera se aleja de la cámara (profundidades mayores e iguales a

110 cm), la esfera decrece en su tamaño y tiende a colocarse debajo de la posición inicial. Como es de esperarse, a profundidades de 85 cm (donde aproximadamente se da la calibración) el error es mínimo y no mayor a 1 cm.

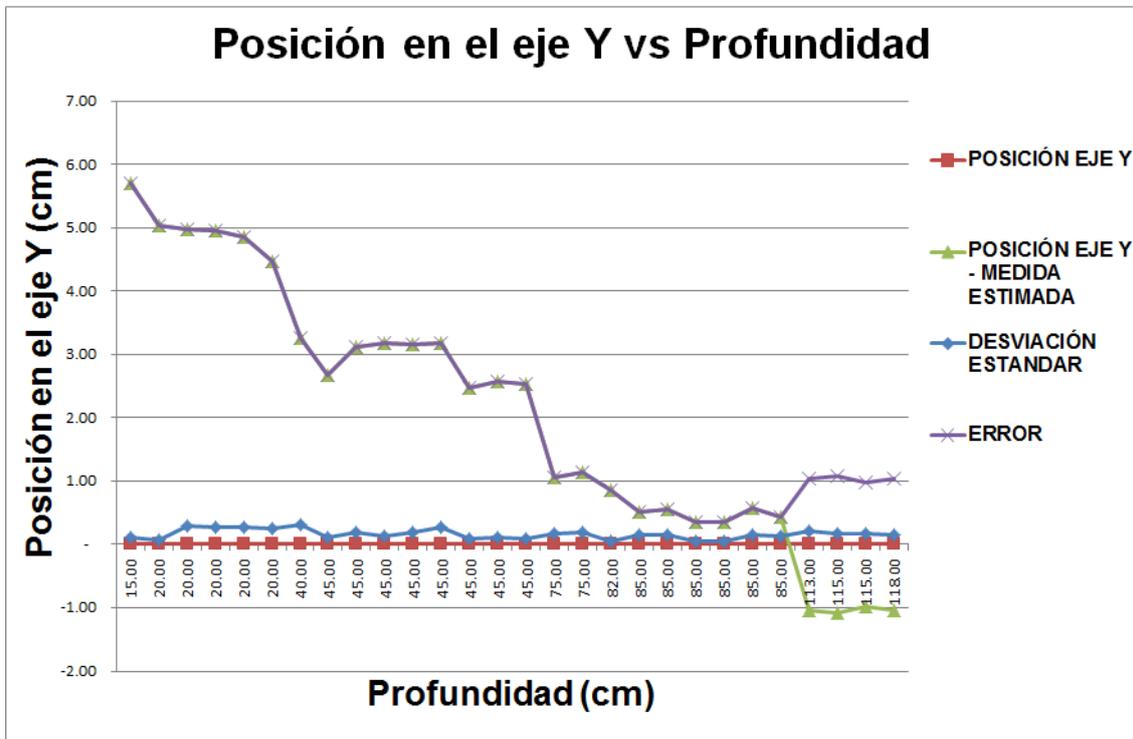


Figura 4.13: Comportamiento del sistema en el eje “Y”

Es importante notar los valores bajos que presenta la desviación estándar. Esto significa que, si bien es cierto la posición de la esfera en el eje “Y” cambió, cuando se inicia el movimiento horizontal, los valores obtenidos alrededor de esta nueva posición difieren poco entre sí, lo que refleja la estabilidad del sistema en este punto. Además, este caso permite plantear una sugerencia de cara a trabajos futuros, en los cuales se deberá plantear un modelo matemático que permita definir teóricamente la posición en el eje “Y” ante cambios como los descritos anteriormente. Finalmente, se concluye que para profundidades entre 20 y 120 cm el sistema en el eje “Y” presenta un error máximo de 5 cm, considerando que si se trabaja a 90 cm, este error máximo no es mayor a 1 cm.

Análisis en el eje “Z” - Profundidad:

El eje “Z” está relacionado con la profundidad de la esfera con respecto a la cámara de video. A priori se conoce la profundidad a la que se realiza la prueba (esto con ayuda del papel milimetrado y reglas utilizadas), por lo que se espera que cuando el sistema sea ejecutado, este muestre resultados cercanos a la profundidad teórica. En la Figura 4.14 se muestra la respuesta del sistema para profundidades que van desde 15 cm hasta los 115 cm. Aquí se observa que el comportamiento de la variable en el eje “Z” para las aplicaciones desarrolladas tiende a ser el esperado.

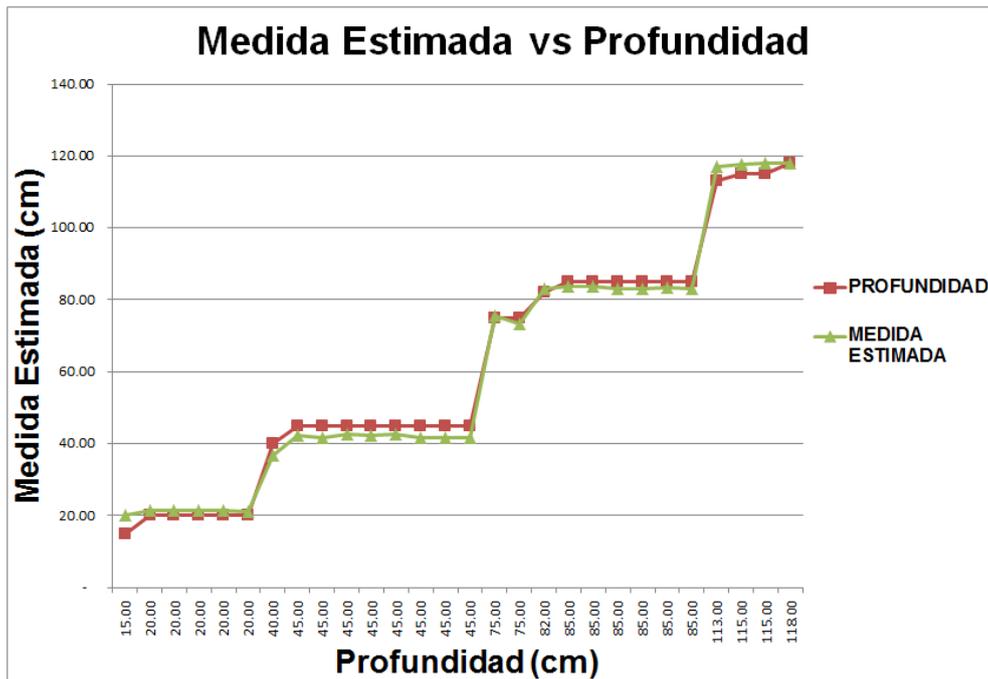


Figura 4.14: Comportamiento del sistema en el eje “Z” - Profundidad

Para poder observar mejor la magnitud del error, se tiene la Figura 4.15, dónde se observa que el error para este eje es de aproximadamente 4 cm entre profundidades de 20 cm a 118 cm. En cuanto a la desviación estándar, se observa que esta presenta valores muy variantes (con valores de 2 cm, 3 cm, e incluso mayores a 4 cm) sin un parámetro fijo que permita deducir, por ejemplo, entre qué profundidades la desviación estándar es baja y, así poder tener mayor confianza en el valor de la profundidad que se esté midiendo con el sistema.

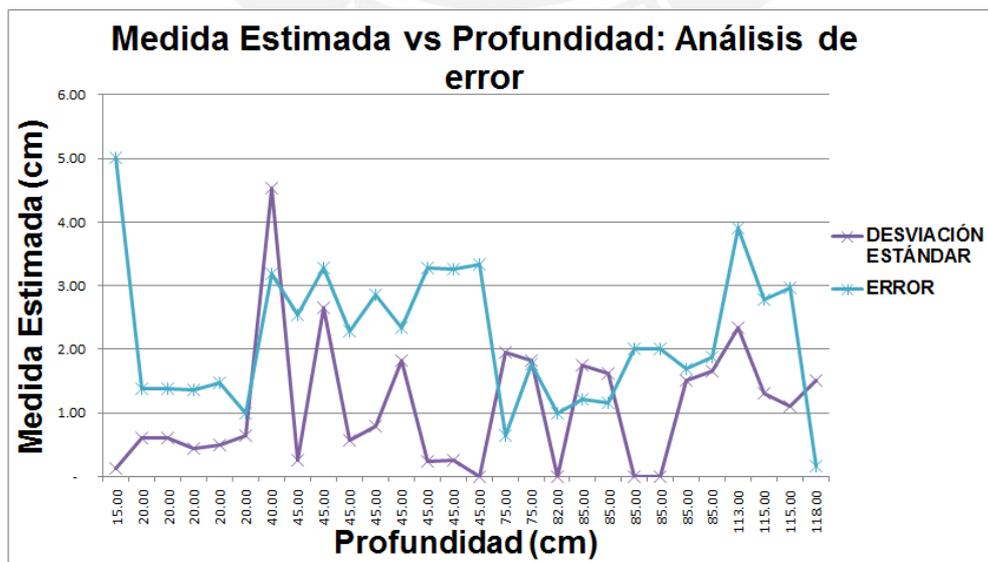


Figura 4.15: Análisis de error en el eje “Z”

Es importante resaltar que el error en el eje “Z” (profundidad) es mayor que para el eje “X” y similar al eje “Y” (dependiendo la profundidad a la cual se trabaje), lo que de alguna forma induce a pensar que se deberían implementar técnicas que permitan mejorar estos resultados, esto a través del desarrollo de modelos matemáticos o el uso de *hardware* adicional que permita tener una medida más exacta de la profundidad. A continuación se procede a evaluar cómo responde el sistema cuando es la profundidad la que ahora varía a través de movimientos de acercamiento y alejamiento con respecto a la cámara de video.

4.3.2. Profundidad variable - Ejes “X” y “Y” constantes

Las pruebas realizadas a una profundidad variable permite medir la exactitud que presenta el sistema para movimientos en el eje “Z”, asociado esto con la medición de la profundidad. Para este caso no se realizará movimiento alguno en los ejes “X” y “Y”. No obstante, es posible que estos varíen al acercarse o alejarse a la cámara ya que el movimiento no está siendo realizado en el mismo eje focal de la cámara. Ante esta situación, los pequeños desplazamientos que se den serán medidos con el objetivo de medir la exactitud de este error. A continuación, se muestran las pruebas realizadas en el presente caso.

Pruebas realizadas con luz encendida

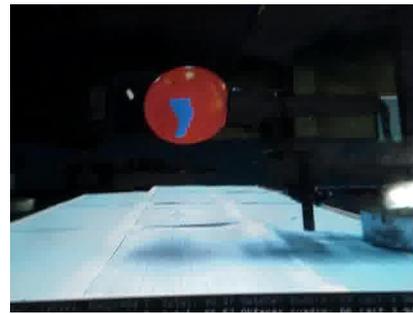
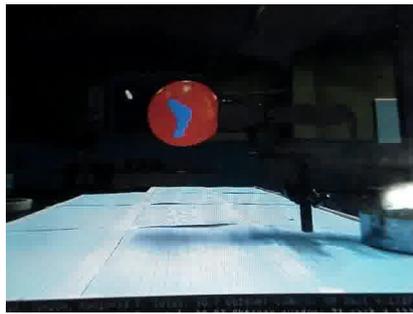
En primera instancia se realizaron pruebas con luz encendida partiendo de una profundidad de 25 cm con respecto a la cámara. En la Figura 4.16 se observa cómo es que para la aplicación 3, el objeto se termina acercando hasta quedar a una distancia de 18 cm aproximadamente. En este caso se puede observar como hay un desplazamiento con respecto al centro de la esfera en el eje “Y”, esto como consecuencia de que el movimiento no es realizado en el eje focal de la cámara de video.



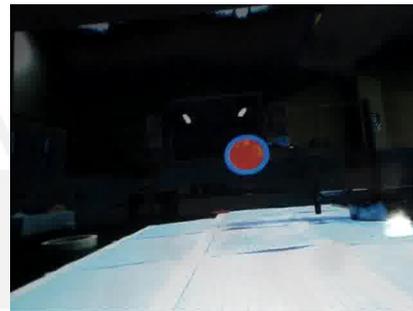
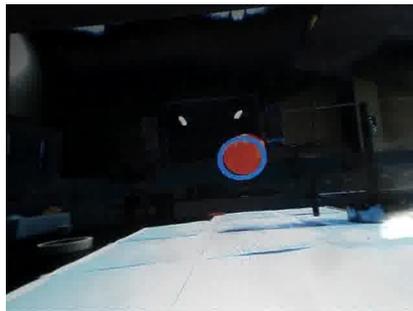
(a) Aplicación 3: Esfera inicialmente a 25 cm (b) Aplicación 3: La esfera se acerca a la cámara

Figura 4.16: Aplicación 3: Acercamiento de la esfera

En las Figuras 4.17(a) y 4.17(b) se observa el desarrollo de la aplicación 2, en la cual la esfera se acerca de 45 cm a 38 cm aproximadamente de distancia hacia la cámara. En las Figuras 4.17(c) y 4.17(d) se muestra el desarrollo de la aplicación 1, en la cual la esfera se aleja de la cámara partiendo desde una posición inicial de 85 cm.



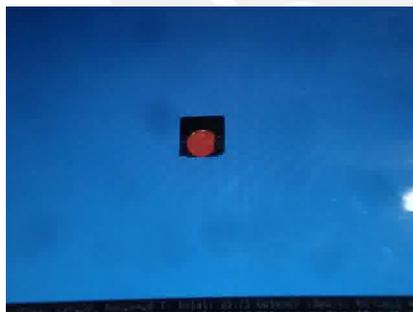
(a) Aplicación 2: Esfera inicialmente a 45 cm (b) Aplicación 2: La esfera se acerca a la cámara



(c) Aplicación 1: Esfera inicialmente a 85 cm (d) Aplicación 1: La esfera se aleja de la cámara

Figura 4.17: Aplicación 2 y 1: La esfera se acerca y aleja de la cámara respectivamente

También se realizaron pruebas a profundidades mayores. En la Figura 4.18 se observa el desarrollo de la aplicación 4, en la cual la esfera se aleja de la cámara partiendo de una distancia de 135 cm.



(a) Aplicación 4: Esfera inicialmente a 135 cm (b) Aplicación 4: La esfera se aleja de la cámara

Figura 4.18: Aplicación 4: La esfera se aleja de la cámara

Pruebas realizadas con luz apagada

En este caso también se realizaron pruebas en un ambiente totalmente oscuro. En la Figura 4.19 se muestra el desarrollo de las aplicaciones 1 y 3. Aquí la esfera se acerca hacia la cámara partiendo de una distancia de 85 cm.



(a) Aplicación 1: Esfera inicialmente a 85 cm (b) Aplicación 1: La esfera se acerca a la cámara



(c) Aplicación 3: Esfera inicialmente a 85 cm (d) Aplicación 3: La esfera se acerca a la cámara

Figura 4.19: Aplicaciones 1 y 3: La esfera se acerca a la cámara en un ambiente oscuro

Resultados

A continuación se muestran los resultados obtenidos de haber procesado la data del sistema. Por lo tanto, se analizarán las variaciones experimentadas a lo largo de los tres ejes.

Análisis en el eje “X”:

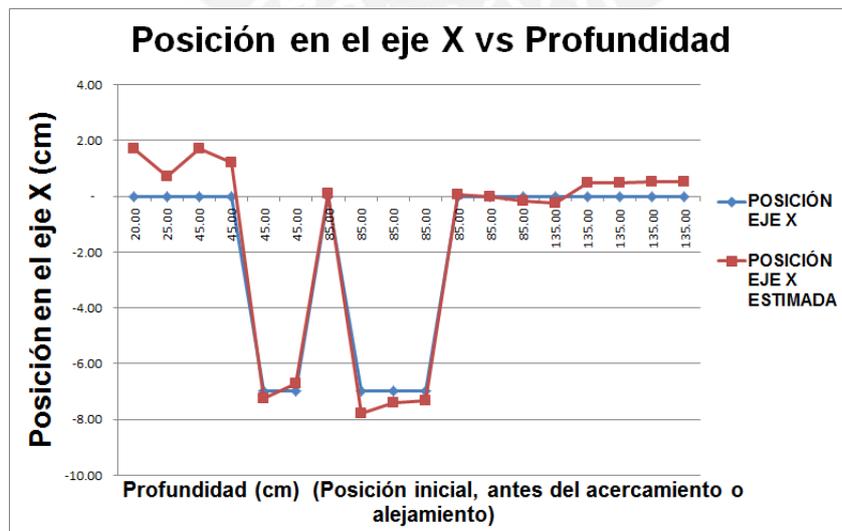


Figura 4.20: Comportamiento del sistema en el eje “X”

La esfera, al estar ubicada en el mismo punto del eje “X” dónde se da la calibración y al no experimentar desplazamientos, presenta una medida que es igual a 0 (posición inicial “X” = 0 cm), como se observa en la Figura 4.20. Además, se tomaron pruebas para 5 casos donde la esfera estuvo desplazada a 7 centímetros del eje “X” de la calibración. A pesar de esto, también se obtuvieron aproximaciones adecuadas.

En la Figura 4.21 se observa que el error máximo que se tiene para las mediciones hechas a distintas profundidades no es mayor a 1.8 cm.

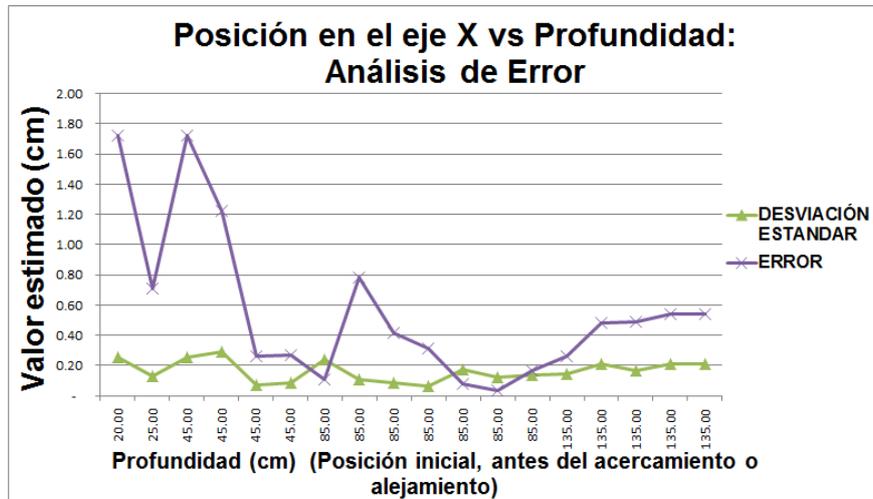


Figura 4.21: Análisis de error en el eje “X”

Análisis en el eje “Y”:

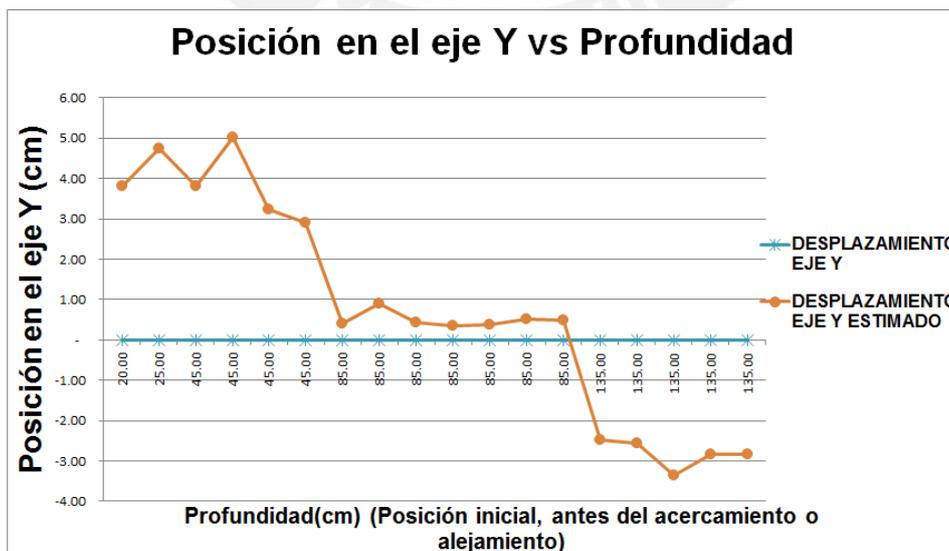


Figura 4.22: Comportamiento del sistema en el eje “Y”

4.3. ANÁLISIS DE EXACTITUD DEL SISTEMA

En la Figura 4.22 se observa el comportamiento del sistema en el eje “Y”. Aquí, de forma análoga a lo explicado para el eje “X”, no se espera que exista variación en este eje (posición inicial “Y” = 0 cm), salvo para los casos cuando la esfera se acerca o aleja haciendo que varíe su posición en este eje como cuando se analizó las variaciones en el eje “Y” para pruebas realizadas a una profundidad constante.

En la Figura 4.23 se observa que el sistema presenta un error máximo de 5 cm para profundidades menores a 45 cm y mayores a 115 cm. Para pruebas con profundidades cercanas a 85 cm (cerca a la distancia de calibración) el error es menor a 1 cm.

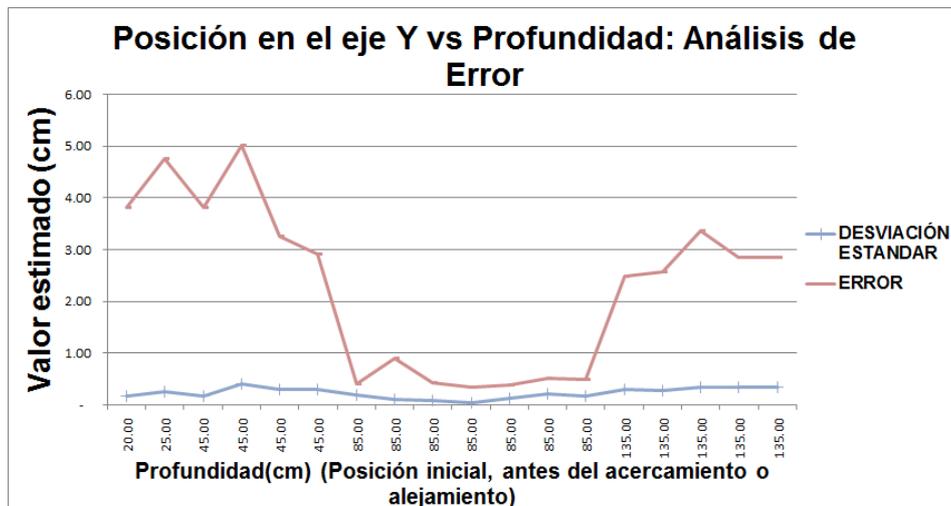


Figura 4.23: Análisis de error en el eje “Y”

Análisis en el eje “Z” - Cambios en profundidad:

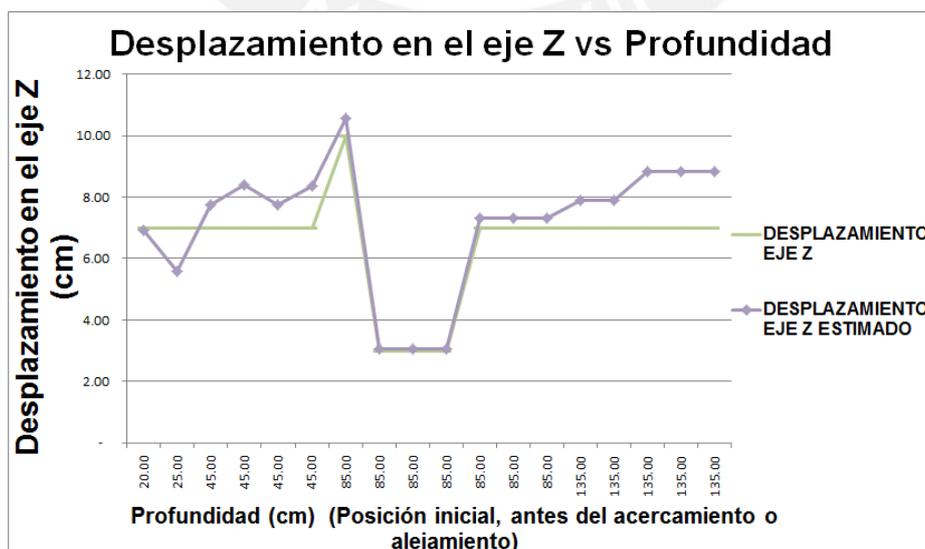


Figura 4.24: Comportamiento del sistema en el eje “Z” - Profundidad

Para las pruebas desarrolladas, por lo general, se experimentaron desplazamientos de 7 cm, salvo para algunos casos en los cuales fueron de 3 cm o 10 cm (este último con un medidor de mayor alcance que solo se pudo utilizar una ocasión). En la Figura 4.24 se muestran los resultados obtenidos en cuanto al desplazamiento en el eje “Z” (profundidad). Es importante mencionar que en el eje “X” de las Figuras 4.24 y 4.25 se tomó como referencia la profundidad sobre la cual se inicia el movimiento de desplazamiento en el eje “Z” para cada prueba realizada.

En la Figura 4.25 se observa que el error máximo no es mayor a 2 cm para las pruebas realizadas a profundidades de 20 cm a 135 cm. Este valor de error tiene relación con lo desarrollado en las pruebas mostradas, es decir, se tienen valores de errores que van desde 1 cm hasta los 5 cm.



Figura 4.25: Análisis de error en el eje “Z” - Profundidad

4.4. Análisis de la velocidad del movimiento

Al inicio del presente capítulo se mencionó que la velocidad de ejecución de las pruebas en mención eran relativamente lentas, entendiéndose por “lenta” que por lo general no supere la velocidad de 1 m/s. La velocidad con la que se desplaza el controlador del movimiento y la velocidad de procesamiento del sistema (frames por segundo) son dos parámetros relacionados que definirán la correcta operatividad del sistema, esto a través de su sensibilidad y qué tan rápida es la respuesta ante los movimientos del usuario. En cuanto al último parámetro es de conocimiento que uno de los objetivos de este trabajo es que el sistema opere a 10 fps, lo cual garantiza una respuesta adecuada y rápida. Sin embargo, es importante definir qué tan rápido puede el usuario mover el controlador manual a fin de poder delimitar una velocidad máxima de movimiento del controlador a la cual el sistema podrá responder adecuadamente.

En principio es importante saber cuál es la velocidad máxima en teoría a la cual el sistema podrá trabajar adecuadamente. Este caso se dará cuando - entre dos cuadros consecutivos - la

esfera no se encuentre dentro de su región de análisis local, sino, apenas haya salido de la misma. Esto se puede apreciar en la Figura 4.26.

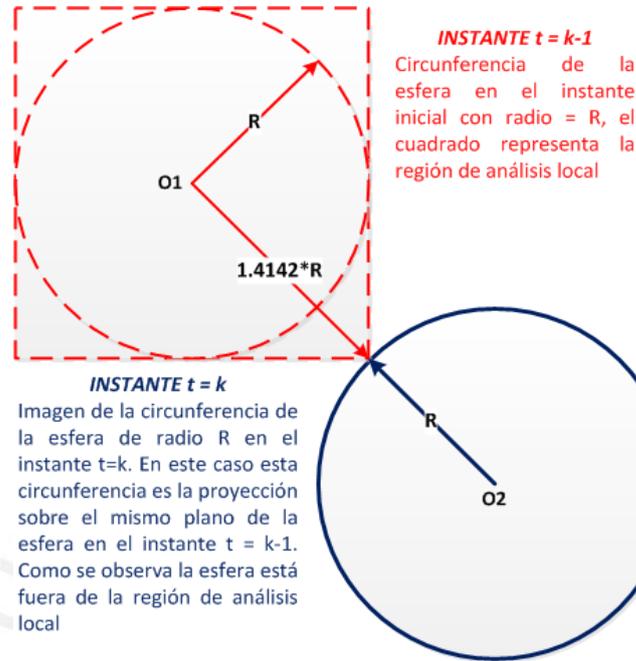


Figura 4.26: Velocidad máxima del controlador: Caso de análisis

Para este caso de análisis, la situación más crítica se dará cuando el desplazamiento se realice a una misma profundidad. A partir de lo anterior y de acuerdo con lo que se muestra en la Figura 4.26, se plantean las siguientes ecuaciones:

$$\text{Distancia recorrida}_{O_1O_2} = \text{Velocidad}_{max} * \text{Tiempo} \quad (4.1)$$

$$R + R\sqrt{2} = V_{max} * (100\text{ms}) \quad (4.2)$$

$$R(1 + 1,4142) = V_{max} * 0,1\text{s} \quad (4.3)$$

$$5\text{cm} * 2,4142 = V_{max} * 0,1\text{s} \quad (4.4)$$

$$0,05\text{m} * 2,4142 = V_{max} * 0,1\text{s} \quad (4.5)$$

$$V_{max} = 1,2071\text{m/s.} \quad (4.6)$$

Donde $R = \text{radio} = \text{Radio de la circunferencia (5 cm)}$

De la ecuación descrita se observa que el movimiento de la esfera no debería ser mayor a 1.21 m/s con el objetivo de tener un seguimiento correcto de la esfera. Esta es la velocidad teórica máxima de desplazamiento del controlador manual. Sin embargo, es importante entender cuál es el rango de velocidades de desplazamiento comúnmente observadas en videojuegos con seguimiento de objetos, la cuál estará limitada por la velocidad máxima de movimiento de la mano.

Según OSHA esta velocidad es de 160 cm/s, y es también conocida como constante de velocidad de la mano, descubierta por el estudioso sueco L. Lobl [43]. Considerando que 160 cm/s equivale a 1.6 m/s y aplicando la ecuación 4.1, el sistema desarrollado tendría que trabajar a 13 fps para asegurar un correcto seguimiento. El presente sistema trabaja a 10 fps por lo que la diferencia no es significativa.

En la Figura 4.27. se muestra la velocidad máxima a la cual se desplazó el controlador manual para las pruebas realizadas y descritas hasta el momento. Como se observa, estos valores prácticamente no sobrepasaron el valor teórico, lo cual tiene mucha coherencia, ya que en ningún momento se perdió el seguimiento de la esfera. Además, las aplicaciones acabaron por decisión del usuario y no por problemas en el seguimiento. Incluso, tampoco alcanzaron la velocidad máxima de la mano.

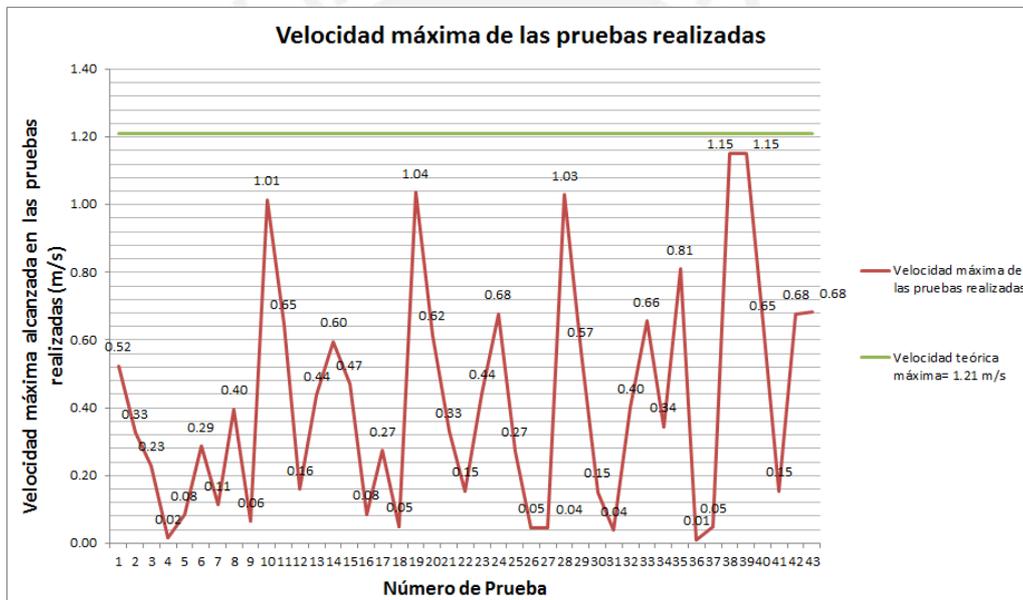


Figura 4.27: Velocidad de movimiento de las pruebas realizadas con ayuda del medidor de desplazamiento Carl Zeiss

El dispositivo medidor de desplazamiento no permitía mover el controlador a velocidades mayores, ante esto, se realizaron algunas pruebas adicionales sin el medidor de desplazamiento, moviendo el controlador manual de forma rápida y calculando la velocidad instantánea a la cual el sistema lo dejaba de seguir. La medición de velocidad se hace a través del cálculo del desplazamiento de la esfera entre dos cuadros consecutivos. Los resultados de la misma se muestran en la Figura 4.28, donde se observan diversas velocidades a las cuales el sistema deja de seguir la esfera. En principio, esta volatilidad se explica con lo rápido que se puede mover el controlador manual haciendo complicada la tarea del sistema de poder registrar una marca fidedigna. También se explica por el margen de error que tiene el sistema al calcular los desplazamientos de la esfera.

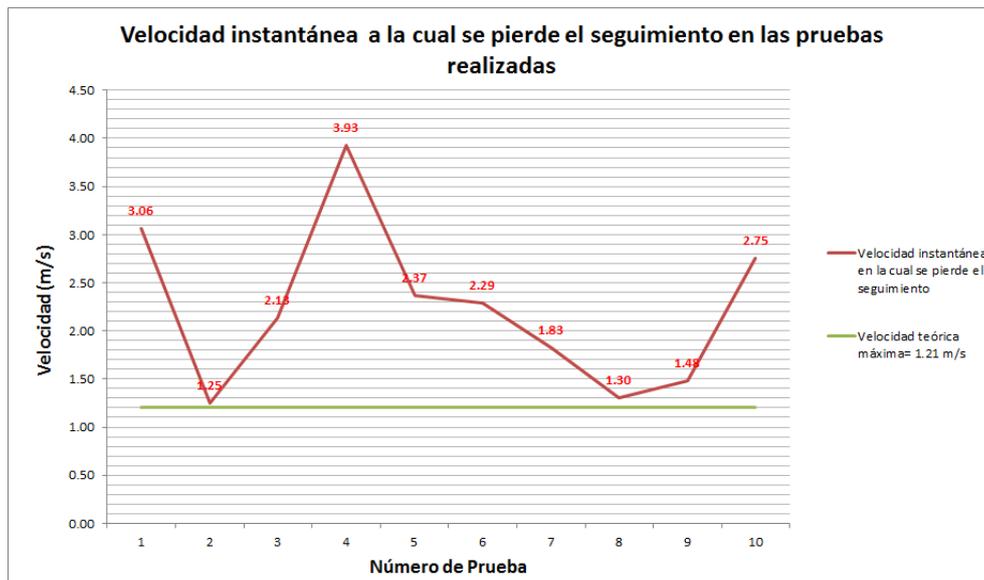


Figura 4.28: Velocidad de movimiento de las pruebas realizadas sin el medidor de desplazamiento Carl Zeiss

De la Figura 4.28, y tomando como referencia las marcas más bajas de las pruebas realizadas: 1.25 m/s, 1.3 m/s y 1.48 m/s, las cuales son cercanas a la velocidad máxima teórica calculada, se puede definir que la velocidad máxima a la cual el controlador manual deberá ser desplazado para asegurar un correcto funcionamiento del sistema es de 1.2 m/s.

4.5. Tiempo de procesamiento de las aplicaciones desarrolladas

En la presente sección se procederá a evaluar el sistema analizando el rendimiento que se tenga con la ejecución de determinadas aplicaciones. Para la evaluación del tiempo de procesamiento es necesario conocer cuáles son las fases que se desarrollaron durante la programación. Esto se muestra en la Figura 3.3, donde se resalta la fase de análisis de movimiento. Esta fase es el eje central del presente estudio ya que permite analizar los cambios que experimenta la esfera, además, permite indicar cuál es su ubicación y tamaño, lo que hace posible realizar el seguimiento respectivo. Los tiempos utilizados en cada una de las fases mostradas en la Figura 3.3 serán denotados de la siguiente forma:

T_1 : Tiempo en ms de obtención de un cuadro.

T_2 : Tiempo en ms para la conversión de valores del cuadro obtenido a valores adecuados para el espacio de color a usar

T_3 : Tiempo en ms de análisis de movimiento (Obtención de coordenadas del centro y del radio).

T_4 : Tiempo en ms de desarrollo de la aplicación.

T_5 : Tiempo en ms para mostrar la imagen.

T_t : Tiempo total en ms.

A continuación, se muestran los resultados obtenidos para la ejecución de las pruebas para cada una de las aplicaciones descritas anteriormente. En las figuras que se mostrarán se puede apreciar la relación de las fases descritas previamente.

4.5.1. Aplicación 1 : Sobreposición de una circunferencia - Análisis de Tiempo

Para la presente aplicación, en la Figura 4.29 se muestra el resultado de la ejecución de un total de 14 pruebas, de las cuales se observa que la mayoría tiene un tiempo de ejecución de 100 ms, lo que ratifica la correcta operatividad del sistema al trabajar a 10 fps.

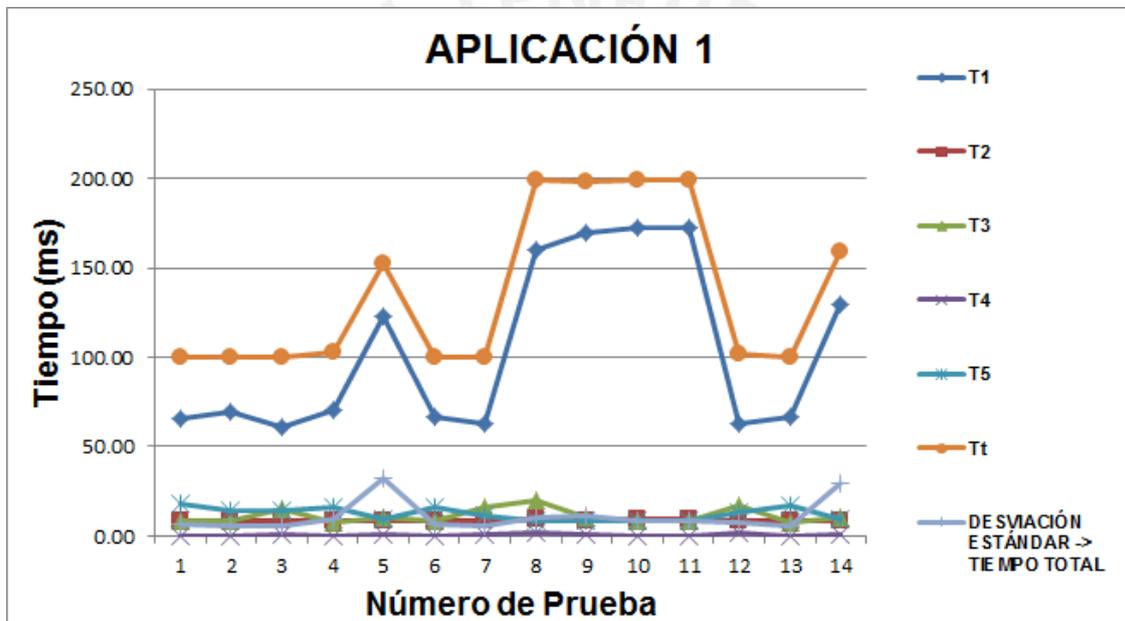


Figura 4.29: Análisis de tiempo de ejecución - Aplicación 1

No obstante, se observa la existencia de pruebas que trabajan a 150 ms e incluso a 200 ms como es el caso de las pruebas 8, 9, 10 y 11. Esto se debe a que esas pruebas fueron ejecutadas cuando las luces estuvieron apagadas en su totalidad, haciendo que el tiempo de captura de la imagen por parte de la cámara sea también mayor, como se puede observar en la curva correspondiente al tiempo de captura de cada cuadro. En el caso de las pruebas cuya duración es de 150 ms esto se debe a que durante la ejecución de la aplicación y a la mitad de la misma, las luces fueron apagadas, lo que hace que en promedio el sistema haya trabajado en un nivel intermedio.

Además, de la Figura 4.29 se puede observar que el tiempo de ejecución de la aplicación es prácticamente constante e independiente de las condiciones externas.

APLICACION 1	T1(ms)	T2(ms)	T3(ms)	T4(ms)	T5(ms)	Tt(ms)	Desv. Estándar
Prueba 1	65.37	8.32	8.19	0.38	17.78	100.05	6.35
Prueba 2	69.15	8.52	8.52	0.27	14.36	100.01	5.68
Prueba 3	60.81	8.31	15.33	1.25	14.35	100.01	5.44
Prueba 4	70.30	8.42	7.78	0.28	16.01	102.82	9.29
Prueba 5	123.04	8.90	10.78	0.66	9.45	152.91	31.92
Prueba 6	66.93	8.54	8.25	0.35	15.83	99.89	6.30
Prueba 7	62.38	8.19	16.48	1.32	11.57	100.00	5.16
Prueba 8	159.62	9.30	19.95	1.52	8.23	198.71	10.40
Prueba 9	169.16	8.94	10.46	0.66	8.90	198.66	11.65
Prueba 10	172.91	9.21	8.52	0.36	8.39	198.68	8.83
Prueba 11	172.91	9.21	8.52	0.36	8.39	198.68	8.83
Prueba 12	63.07	8.24	16.70	1.43	12.88	102.00	7.65
Prueba 13	66.67	8.29	7.82	0.33	16.85	99.99	5.20
Prueba 14	129.61	8.91	10.57	0.64	9.34	159.15	29.10

Cuadro 4.1: Tabla - Tiempos en Aplicación 1

4.5.2. Aplicación 2: Trayectoria del centro de la circunferencia - Análisis de tiempo

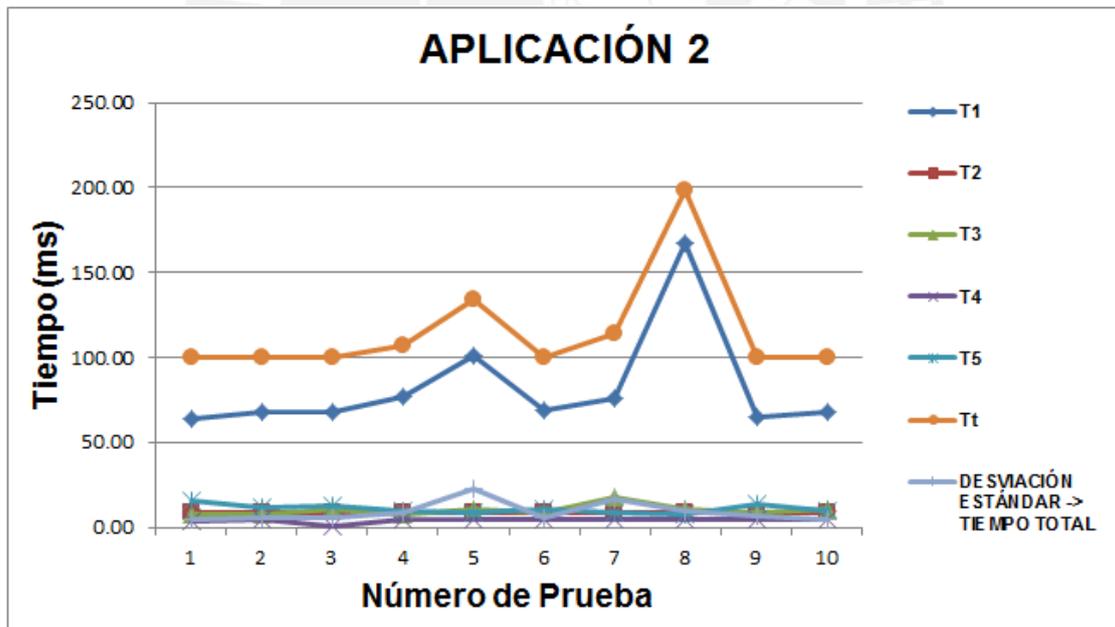


Figura 4.30: Análisis de tiempo de ejecución - Aplicación 2

Para el caso de la aplicación 2 mostrada en la Figura 4.30, se observa que las pruebas también se ejecutan dentro de los 100 ms esperados. Salvo las pruebas 5 y 8, cuyos tiempos de procesamiento son mayores y se debe a que fueron trabajadas con las luces apagadas, se considera que el sistema trabaja correctamente a 10 fps.

APLICACION 2	T1(ms)	T2(ms)	T3(ms)	T4(ms)	T5(ms)	Tt(ms)	Desv. Estándar
Prueba 1	64.18	8.23	8.03	4.05	15.75	100.33	4.79
Prueba 2	68.09	8.59	8.01	4.17	11.25	100.04	6.00
Prueba 3	67.72	8.29	10.64	0.68	12.57	99.95	5.80
Prueba 4	76.44	8.66	7.94	4.21	9.83	107.12	8.37
Prueba 5	101.45	8.57	10.35	4.29	8.94	133.96	22.95
Prueba 6	68.66	8.43	8.18	4.13	10.58	100.00	5.74
Prueba 7	75.57	8.42	17.38	4.28	8.28	113.92	16.56
Prueba 8	167.18	8.79	10.60	4.22	7.90	198.73	9.18
Prueba 9	65.32	8.48	8.27	4.12	13.80	100.00	6.73
Prueba 10	67.68	8.37	10.50	4.14	9.29	100.03	4.95

Cuadro 4.2: Tabla - Tiempos en Aplicación 2

4.5.3. Aplicación 3: Sobreposición del logo de la PUCP - Análisis de tiempo

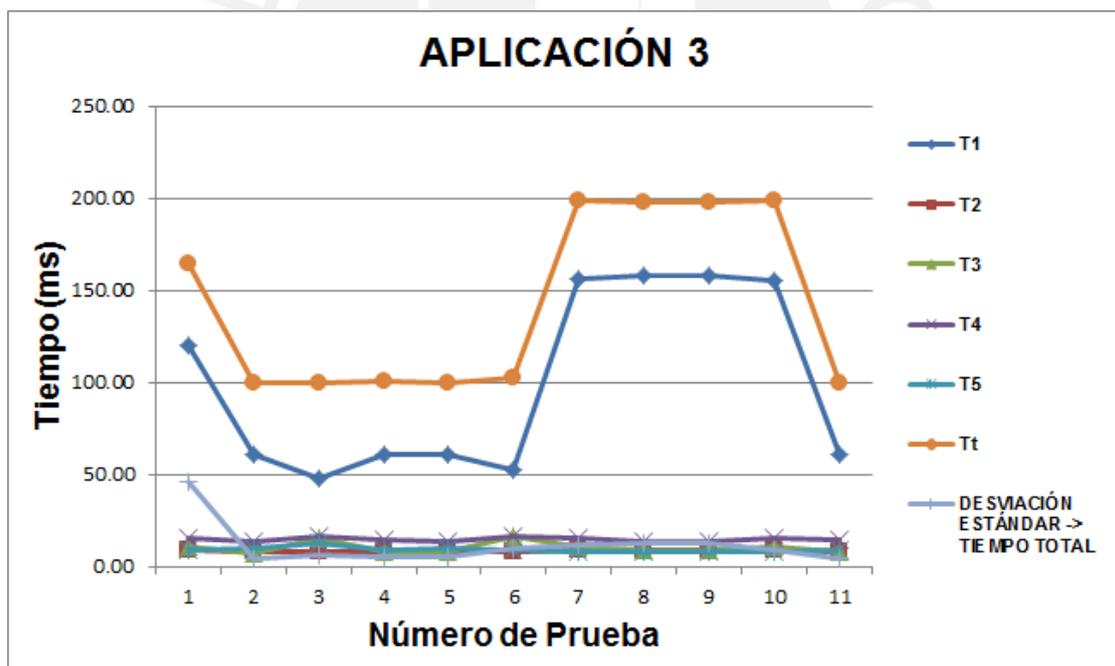


Figura 4.31: Análisis de tiempo de ejecución - Aplicación 3

Para el caso de la aplicación 3 (Figura 4.31), esta también se encuentra trabajando dentro de los 100 ms. esperados, salvo las pruebas 1, 7, 8, 9 y 10 que fueron trabajadas con las luces apagadas y presentan un tiempo de procesamiento mayor a 150 ms.

APLICACION 3	T1(ms)	T2(ms)	T3(ms)	T4(ms)	T5(ms)	Tt(ms)	Desv. Estándar
Prueba 1	119.91	8.81	11.20	15.51	8.97	164.38	45.83
Prueba 2	60.77	8.25	7.43	13.75	9.64	99.93	4.20
Prueba 3	47.82	8.19	15.16	16.07	12.88	100.05	6.51
Prueba 4	61.18	8.33	7.65	14.38	8.73	100.25	4.86
Prueba 5	60.71	8.30	7.92	13.68	9.47	100.02	5.03
Prueba 6	52.29	8.18	16.48	16.45	9.11	102.56	10.05
Prueba 7	156.14	8.66	10.68	15.48	7.69	198.53	11.21
Prueba 8	158.33	8.94	8.54	13.86	7.97	198.10	12.78
Prueba 9	158.33	8.94	8.54	13.86	7.97	198.10	12.78
Prueba 10	155.65	9.10	11.02	15.45	7.71	198.68	8.81
Prueba 11	60.64	8.27	7.88	14.18	9.13	100.05	4.34

Cuadro 4.3: Tabla - Tiempos en Aplicación 3

4.5.4. Aplicación 4: Seguimiento y limpieza de la trayectoria de la circunferencia - Análisis de tiempo

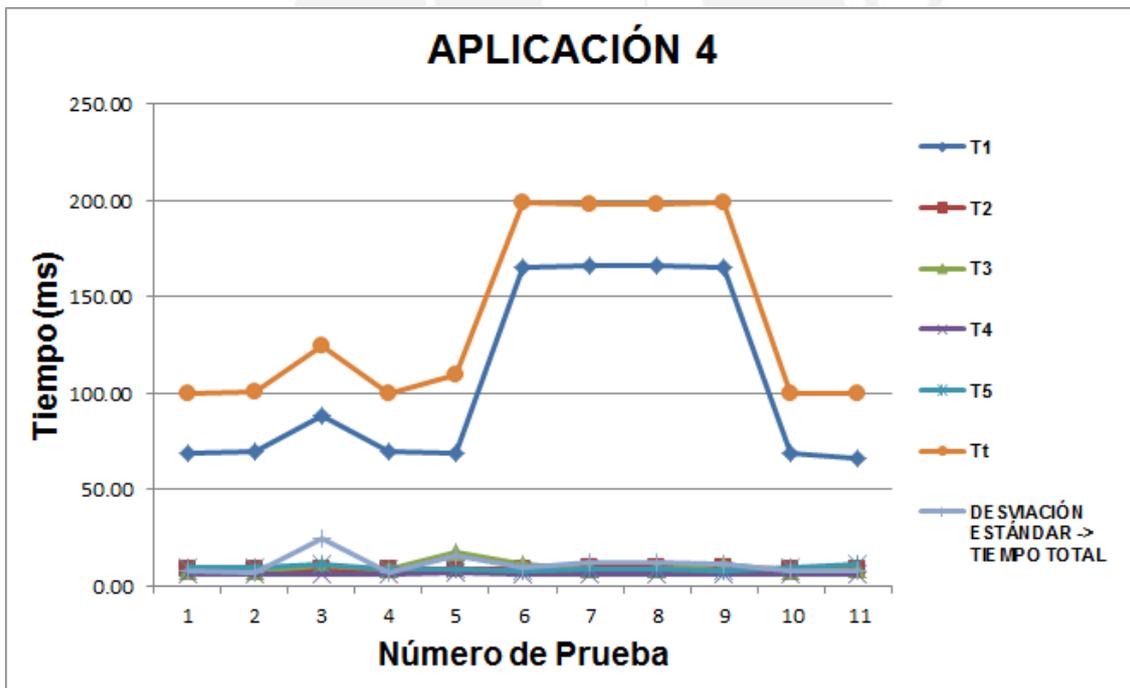


Figura 4.32: Análisis de tiempo de ejecución - Aplicación 4

Finalmente para la aplicación 4 (Figura 4.32), las pruebas que fueron trabajadas con las luces encendidas trabajan dentro de los 100 ms esperados, salvo las pruebas 3, 6, 7, 8 y 9 que fueron trabajadas con escenarios oscuros y presentan un tiempo de procesamiento mayor a 100 ms.

APLICACION 4	T1(ms)	T2(ms)	T3(ms)	T4(ms)	T5(ms)	Tt(ms)	Desv. Estándar
Prueba 1	68.53	8.66	7.80	5.63	9.43	100.02	7.65
Prueba 2	69.16	8.63	7.94	5.69	8.93	100.37	6.87
Prueba 3	88.22	8.53	10.13	5.92	11.22	124.10	24.73
Prueba 4	69.40	8.58	8.25	5.63	8.15	100.00	6.77
Prueba 5	68.24	8.69	17.72	6.57	8.21	109.41	15.20
Prueba 6	165.49	8.86	10.79	5.95	7.98	198.76	9.58
Prueba 7	165.83	9.24	8.71	5.95	8.23	198.24	12.04
Prueba 8	165.83	9.24	8.71	5.95	8.23	198.24	12.04
Prueba 9	164.83	9.06	11.02	5.90	7.86	198.61	11.03
Prueba 10	68.53	8.66	7.80	5.63	9.43	100.02	7.65
Prueba 11	66.40	8.46	8.01	5.58	11.50	99.98	7.28

Cuadro 4.4: Tabla - Tiempos en Aplicación 4

4.6. Análisis económico

El presupuesto de los recursos tanto humanos como tecnológicos, usados tanto para *hardware* como para *software*, está estructurado de la siguiente forma:

Artículo	Cantidad	Precio unidad (\$)	Total (\$)
S.O Linux Ubuntu 11.04	1	0.00	0.00
Librería FFMPEG	1	0.00	0.00
Programa Geanix	1	0.00	0.00
Computador Personal	1	1500.00	1500.00
Cámara Logitech	1	70.00	70.00
Controlador Manual	1	10.00	10.00
Programador	1	11000.00	11000.00
Gastos menores	1	20.00	20.00
TOTAL			12600.00

Cuadro 4.5: Presupuesto utilizado para el desarrollo del proyecto

A partir del costo mostrado, se resalta lo provechoso y útil que resulta trabajar con el sistema operativo Linux, ya que permite abaratar costos, además de todas las facilidades que brinda a nivel de software.

4.7. Discusión de resultados

A lo largo del presente capítulo, se ha venido mostrando y comentando los resultados obtenidos luego de la ejecución de diversas pruebas realizadas a diferentes profundidades de la cámara de video (entre 20 cm y 135 cm). Con estas pruebas se buscó medir la capacidad del sistema en exactitud y tiempo. En ese sentido, se enfocará el siguiente análisis bajo esas dos perspectivas.

4.7.1. Acerca de la exactitud

Se realizaron pruebas que consideraban movimientos a una profundidad constante y a una profundidad variable. Para el primer caso, los desplazamientos que se hicieron en el eje “X” fueron medidos por el sistema con un error de 2 cm. Por otro lado, para el eje “Y” se observó que, al no estar en el eje focal de la cámara, esta realizaba desplazamientos que no son medidos por el sistema, registrando un error máximo de 5 cm. Para el caso de la profundidad, esta fue medida con un error de 4 cm. No obstante, la desviación estándar es muy variable lo que refleja que el sistema para medir la profundidad no es muy preciso. En el segundo caso, para medir los desplazamientos en el eje “Z” se tomó como referencia que los ejes “X” y “Y” permanecerán constantes, para la variable en el eje “X” no hubo problema al presentar un error de 1.8 cm, pero, para las variaciones en el eje “Y” se observaron las mismas dificultades que para las pruebas a profundidad constante. En cuanto al desplazamiento en el eje “Z”, se observa la presencia de un error máximo de alrededor 2 cm.

Finalmente, se resalta que el grado de exactitud del sistema al funcionar entre 20 cm y 135 cm es de 5 cm. En caso se limite más el rango de operación del sistema para operar a una profundidad de trabajo de entre 50 cm y 100 cm, este error podría ser ahora de alrededor de 2 cm aproximadamente. El que estos márgenes sean aceptables o no dependerá de la aplicación que se desee desarrollar y los objetivos que persigan. Para las aplicaciones desarrolladas en el presente trabajo se puede considerar como márgenes adecuados, ya que para estas el nivel de exactitud no es fundamental, sino que la utilidad de las aplicaciones desarrolladas pasa más por aspectos visuales y demostrativos.

4.7.2. Acerca del tiempo de respuesta

Se planteó que si el sistema puede funcionar al menos a 10 fps, se estaría asegurando un grado de comodidad aceptable para la interacción usuario - computador. A través de la ejecución de las aplicaciones desarrolladas se observa que bajo condiciones de iluminación adecuada se asegura la respuesta del sistema dentro de estos tiempos.

4.7.3. Acerca de las condiciones de funcionamiento

Finalmente, partiendo de las condiciones bajo las cuales se desarrollaron las pruebas de este trabajo, se define que el sistema operará adecuadamente a 10 fps en ambientes de preferencia cerrados e iluminados (condiciones de iluminación de una oficina tradicional) y el movimiento del controlador manual no debería de exceder de 1.2 m/s para que el sistema pueda seguir la esfera.

Conclusiones

El sistema desarrollado brinda un seguimiento con una exactitud tal que asegura un error máximo de 5 cm en la ubicación de las coordenadas de la esfera a lo largo de sus 3 dimensiones, esto para la ejecución de aplicaciones que sean desarrolladas a una distancia de entre 20 cm a 135 cm de profundidad con respecto a la cámara de video. Este error puede ser aún menor si se limita más la profundidad de trabajo o si se analiza eje por eje. Con respecto a la rapidez del sistema, se ha obtenido un tiempo de procesamiento tal que permite a la cámara trabajar a 10 cuadros por segundo. Si bien es cierto, sistemas como el Play Station 3 trabajan a velocidades mayores, es necesario mencionar que esta hace uso de otros componentes en *hardware* y *software* que facilitan esta tarea, componentes que no son objeto de estudio de esta tesis.

Se logró implementar correctamente las etapas de calibración y seguimiento, a través de la obtención y correcto uso de parámetros como umbrales de color, resolución espacial, entre otros, y del correcto manejo de las regiones de análisis local (o global en caso se necesite). Esto hizo posible el ahorro de tiempo y otros recursos. Así mismo, el sistema fue desarrollado adecuadamente sobre la plataforma Linux haciendo uso del lenguaje de programación C y mediante el uso de la librería FFMPEG. El sistema operará adecuadamente en ambientes de preferencia cerrados e iluminados y la velocidad del movimiento del controlador manual no debe de exceder de 1.2 m/s.

Por otro lado se comprobó que la implementación del algoritmo de “least square circle fit” permite un ajuste óptimo para la obtención de parámetros de una circunferencia. Para ello se parte de la información de bordes que la etapa de segmentación proporciona. También se logró segmentar exitosamente la esfera, a través de la segmentación por color (usando el espacio de color RGB). Se detectó la región correspondiente a la esfera y se obtuvo los puntos necesarios que permitían una correcta estimación. También el sistema está en la capacidad de segmentar automáticamente la esfera después de haber superado la etapa de calibración

Por último, a pesar de haber obtenido un tiempo de procesamiento adecuado, este puede ser mejorado y se puede obtener una velocidad más acorde a las exigencias del mercado actual. Esto pone en evidencia la importancia de la implementación de algoritmos y técnicas de programación más sofisticados como método que permita obtener un tiempo de procesamiento más óptimo y hacer un uso más eficiente de los recursos del computador.

Recomendaciones

En cuanto al *hardware*, se pueden establecer diversas mejoras. Una de ellas es partir de la implementación del sistema en computadores cuyos procesadores tengan un mejor rendimiento al momento de realizar operaciones matemáticas o incluso en procesadores que presenten varios núcleos que faciliten la programación en paralelo.

En cuanto a la programación se pueden implementar diversas mejoras como el uso de instrucciones SIMD [44] o el uso de “*threads*” que permitan optimizar el tiempo de ejecución.

Así mismo se sugiere analizar el impacto que podrían tener la implementación de este sistema partiendo de otros espacios de color como: YUV, CIEL*a*b*, entre otros. Aquí se podría observar el impacto directo que tienen otros factores importantes en el sistema como es la iluminación.

En cuanto al seguimiento, existen en la actualidad métodos que permiten la estimación de parámetros valiéndose del aprendizaje de estados anteriores. Uno de estos métodos puede ser la implementación del filtro de Kalman, cuya implementación podría suponer la obtención de mejoras en cuanto al tiempo de procesamiento y precisión en la estimación de parámetros.

Como futuro tema de estudio, se podría implementar aspectos como la lectura de datos que podrían proporcionar sensores como los acelerómetros y, en conjunto con estos datos, añadir funciones adicionales al sistema y analizar mejor aspectos como el cambio en rotación, velocidad del objeto, incluso, generalizar el sistema ya no solo para la detección de objetos esféricos.

Bibliografía

- [1] Sony. Playstation 3 sales reach 50 million units worldwide. Accedido 19 de junio del 2011. Sony Computer Entertainment Inc. [Online]. Available: http://www.scei.co.jp/corporate/release/110415_e.html
- [2] (2010, Marzo) Research shows average number of hours per week spent on online gaming has grown since 2009. Accedido el 19 de junio del 2011. [Online]. Available: http://www.npd.com/press/releases/press_100302.html
- [3] S. I. A. Asaad A. M. Al-Salih, "Image segmentation in a stationary camera-based system for computer vision," *Computer Vision and Information Technology*, vol. -, pp. 34–43, 2010.
- [4] M. Z. Elias Malamas, Euripides Petrakis, "A survey on industrial vision systems, applications and tools," *Image and Vision Computing*, vol. 21, pp. 171–188, 2003.
- [5] P. F. Alcantarilla, "Sistema de monitorización y control de tráfico en carrera," Ph.D. dissertation, Universidad de Alcalá, 2006.
- [6] Sony. Play station move. Accedido el 1 de mayo 2011. Sony. [Online]. Available: <http://us.playstation.com/ps3/playstation-move/>
- [7] C. J. Murray. (2007, February) Ps3: Fast and cool. Design News. [Online]. Available: http://www.designnews.com/document.asp?doc_id=219798
- [8] P. Rodriguez, "Optical motion tracking in earthquake-simulation shake table testing: Preliminary results," *Asilomar Conference on Signals, Systems, and Computers (ASILOMAR)*, pp. 371–375, 2009.
- [9] (2011, January) Getting into the game. Accedido el 27 de abril del 2011. Popular Mechanics [serial online]. [Online]. Available: <http://search.ebscohost.com>
- [10] Microsoft. Xbox 360. Accedido el 1 de mayo del 2011. Microsoft Co. [Online]. Available: <http://www.xbox.com/en-US/xbox360/consoles?xr=shellnav>
- [11] Wii. Accedido el 1 de mayo del 2011. Nintendo Co. [Online]. Available: <http://www.nintendo.com/wii/console/controllers>

- [12] L. Alexande. (2011, Marzo) Microsoft: Kinect hits 10 million units, 10 million games. Accedido el 18 de junio del 2011. Gamasutra. [Online]. Available: <http://www.gamasutra.com/view/news/33430/>
- [13] N. Group. 2009 video game software sales across top global markets experience declines for console and portable platforms. Accedido el 19 de junio del 2011. NPD Group. [Online]. Available: http://www.npd.com/press/releases/press_100203a.html
- [14] Sony. Play station 3. Accedido el 13 de abril del 2011. Sony. [Online]. Available: <http://scei.co.jp/corporate/release/100311e.html>
- [15] G. NPD. (2011, Febrero) The npd group provides insight into category sales trends for kids during the 2010 holiday season. Accedido el 19 de junio del 2011. NPD Group. [Online]. Available: http://www.npd.com/press/releases/press_110224a.html
- [16] L. Nintendo Co, “Consolidated sales transition by region,” Nintendo, Tech. Rep., 2011.
- [17] L. Nintendo Co., “Consolidated financial highlights,” Nintendo Co., Ltd., Tech. Rep., 2009.
- [18] (2009) Third quarter 2009 sales results from top global video games software markets released. Accedido el 18 de junio del 2011. NPD Group. [Online]. Available: http://www.npd.com/press/releases/press_091105.html
- [19] A. F. David Armando Insuasti, Julian Quiroga, “Detección y seguimiento de vehículos automotores en video,” *XIII Simposio de tratamiento de Señales, Imágenes y Visión Artificial*, 2008.
- [20] A. M. Vega, “Seguimiento visual en playstation 3,” Master’s thesis, Universidad Rey Juan Carlos, 2010.
- [21] E. F. Jeffrey Snyder, *Digital Image Processing and Analysis: Human and Computer Vision Applications with CVIPtools, Second Edition*, by Scott E. Umbaugh, J. Electron, Ed. Imaging 20, 039901, Sep 15, 2011.
- [22] R. E. W. Rafael C. Gonzalez, *Digital Image Processing (3rd Edition)*, P. H. S.-. 013168728X, Ed. Prentice Hall, 3rd edition (August 31, 2007),.
- [23] S. H. Cha, “Comprehensive survey on distance/similarity measures between probability density functions,” *International Journal of Mathematical Models and Methods un Applied Sciences*, vol. 4, pp. 300 – 307, 2007.
- [24] D. B. M. Swain, “Color indexing,” *International Journal of Computer Vision*, vol. 7, pp. 11–32, 1991.
- [25] K. N. N. Douglas Chai, “Face segmentation using skin-color map in videophone applications,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 551–564, 1999.

- [26] P. L. Rosin, “Unimodal thresholding,” *Pattern Recognition*, pp. 2083–2096, 2001.
- [27] Y.-L. C. Jeng-Horng Chang, Kuo-Chin Fan, “Multi-modal gray level histogram modeling and decomposition,” *Image and Vision Computing*, vol. 20, pp. 203–216, 2002.
- [28] J. M. S. Obdrzalek, “Object recognition using local affine frames on distinguished regions,” *British Machine Vision Conference*, vol. 1, pp. 113 – 122, 2002.
- [29] D. J. Ronen Basri, “Recognition using region correspondences,” *International Journal of Computer Vision*, vol. 25(2), pp. 145 – 166, 1997.
- [30] B. S. K. Mikolajczyk, B. Leibe, “Multiple object class detection with a generative model,” *Computer Vision and Pattern Recognition*, vol. 1, pp. 26–36, 2006.
- [31] D. L. M. Brown, “Invariant features from interest point groups,” *British Machine Vision Conference*, vol. 1, pp. 656–665, 2002.
- [32] A. Z. Josef Sivic, “Video data mining using configurations of viewpoint invariant regions,” *Computer Vision and Pattern Recognition*, vol. 1, pp. 488–495, 2004.
- [33] V. G. V. Ferrari, T. Tuytelaars, “Simultaneous object recognition and segmentation from single or multiple model views,” *International Journal of Computer Vision*, vol. Vol. 67, No. 2., pp. 159–188, 2006.
- [34] J. M. Z. Kalal, K. Mikolajczyk, “Face-tld: Tracking-learning-detection applied to faces,” *International Conference on Image Processing*, pp. 3789 – 3792, 2010.
- [35] T. S. K. Zimmermann, J. Matas, “Tracking by an optical sequence of linear predictors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31 - 04, pp. 677 – 692, 2009.
- [36] K. H. Uemura, S. Ishikawa, “Feature tracking and motion compensation for action recognition,” in *British Machine Vision Conference*, 2008.
- [37] D. Simon, “Kalman filtering,” *Embedded Systems Programming*, pp. 72–79, 2011.
- [38] G. B. G. Welch, “An introduction to the kalman filter,” University of North Carolina at Chapel Hill Chapel Hill, NC, USA, Tech. Rep., 1995.
- [39] Quickcam orbit af. Accedido el 24 de noviembre del 2011. Logitech. [Online]. Available: <http://www.logitech.com/en-us/435/3480>
- [40] Ffmpeg. Accedido el 6 de abril del 2011. [Online]. Available: <http://www.ffmpeg.org/ffmpeg.html>.
- [41] D. Coope, “Circle fitting by linear and nonlinear least squares,” *Journal of Optimization Theory and Applications*, vol. 76 N2, pp. 381–388, 1993.
- [42] (2012) Accedido el 01 de diciembre del 2012. [Online]. Available: <http://www.zeiss.es/>

- [43] C. R. Asfahl, *Seguridad Industrial y Salud*, 4th ed., M. Horton, Ed. Prentice Hall, Mexico., 2000, pagina 317.
- [44] Simd. Accedido el 25 de noviembre del 2011. [Online]. Available: <http://www.simd.org/>



ANEXOS

- ANEXO 1: Pruebas realizadas para el análisis de exactitud y tiempo.
- ANEXO 2: Código fuente del programa en lenguaje C.

