

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS E INGENIERÍA



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

**ROBOT MÓVIL CON VISIÓN ESTEREOSCÓPICA PARA LA
LOCALIZACIÓN DE OBJETOS**

Tesis para optar el Título De Ingeniero Electrónico, que presenta el bachiller

Jorge Antonio Rodríguez Rodríguez

ASESOR: Juan Javier Sotomayor Moriano

Lima, junio del 2009

RESUMEN

El procesamiento de imágenes aplicado al campo de control y automatización de sistemas industriales ha tenido una gran influencia en la tecnología de los robots móviles; así se han venido desarrollando diversas técnicas de visión para poder obtener nuevos parámetros cada vez más precisos sobre el entorno de trabajo del robot móvil.

Las técnicas de reconocimiento de objetos y visión estereoscópica han demostrado ser una herramienta muy útil, ya que gracias a esto, los vehículos no tripulados pueden llegar a tener una mejor interacción con su entorno, lo cual resulta en un mejor desenvolvimiento en las diferentes tareas que realizan

El presente trabajo de tesis describe la implementación de un sistema de visión estéreo aplicado al campo de los robots móviles para la localización de objetos, el cual tiene como principal características encontrar la distancia en coordenadas XYZ de un objeto en particular; para que luego esta información sea llevada al robot.

Tanto para la tarea de segmentación como del cálculo de coordenadas se utilizará a la librería OpenCV la cual nos permite el uso de los diferentes algoritmos de procesamiento de imágenes.

La etapa de visión estéreo propiamente dicha se llevará a cabo mediante la implementación de las siguientes etapas: cálculo de los parámetros de las cámaras, calibración, rectificación de imágenes y finalmente la triangulación, método por el cual se obtienen las coordenadas deseadas.

El sistema de visión estéreo desarrollado será implementado en un computador embebido y se va comunicar con el robot móvil mediante protocolo TCP/IP para el envío de las diferentes coordenadas.

FACULTAD
DE CIENCIAS
E INGENIERÍA



90
AÑOS

PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

TEMA DE TESIS PARA OPTAR EL TÍTULO DE INGENIERO ELECTRÓNICO

Título : Robot Móvil con Visión Estereoscópica para la Localización de Objetos

Índice

Introducción

1. Descripción de Robot Móvil
2. Obtención y Procesamiento de Imágenes
3. Visión Estereoscópica y Localización de Objetos
4. Pruebas y Resultados. Análisis Económico

Conclusiones

Recomendaciones

Bibliografía

Anexos

Máximo: 100 páginas



Hobmeys



FACULTAD
DE CIENCIAS
E INGENIERÍA



90
AÑOS

PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

TEMA DE TESIS PARA OPTAR EL TÍTULO DE INGENIERO ELECTRÓNICO

Título : Robot Móvil con Visión Estereoscópica para la Localización de Objetos
 Área : Control y Automatización # 561
 Asesor : Dr. Javier Sotomayor
 Alumno : Jorge Antonio Rodríguez Rodríguez
 Código : 20030261
 Fecha : 23 de Noviembre de 2007



Descripción y Objetivos

El procesamiento de imágenes aplicado al campo del control y automatización de sistemas industriales, ha permitido resolver tareas en las cuales se requiere contar con visión artificial. Para los robots móviles, se han venido desarrollando diversas técnicas de visión, a fin de que estos puedan lograr una interacción eficaz con su entorno de trabajo.

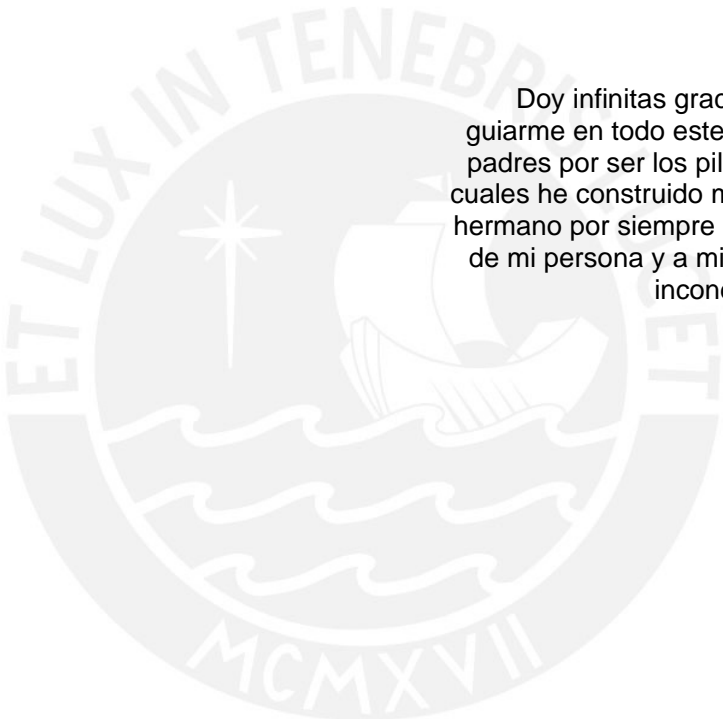
Las técnicas de reconocimiento de objetos y visión estereoscópica han demostrado ser herramientas muy útiles, ya que gracias a estas, los vehículos no tripulados pueden llegar a tener una mejor interacción con su entorno, lo cual resulta en un mejor desenvolvimiento en las diferentes tareas que realizan.

La presente tesis tiene como objetivo la implementación de un sistema de visión estereoscópica para un robot móvil. Este sistema de visión permitirá la obtención de las coordenadas de los diferentes objetos con ayuda de un software especializado.

El desarrollo del sistema de visión propuesto, incluye el estudio de las diferentes técnicas de mejoramiento y segmentación de imágenes, asimismo la elección e implementación del algoritmo para obtener la correspondencia de puntos en las imágenes y el algoritmo de triangulación para la obtención de coordenadas de los objetos. Se desarrollará también una interfaz de usuario para visualizar las pruebas realizadas con las imágenes.

PONTIFICIA UNIVERSIDAD CATOLICA DEL PERU
 SECCION ELECTRICIDAD Y ELECTRONICA

 Ing. ANDRES FLORES ESPINOZA
 Coordinador de la Especialidad de Ingeniería Electrónica

A large, faint watermark of the university logo is centered on the page, featuring the same ship, star, and Latin motto as the header logo.

Doy infinitas gracias a Dios por guiarme en todo este tiempo, a mis padres por ser los pilares sobre los cuales he construido mi destino a mi hermano por siempre sacar lo mejor de mi persona y a mi familia por su incondicional apoyo

INDICE

INTRODUCCIÓN	I
CAPÍTULO 1: DESCRIPCIÓN DE ROBOT MÓVIL	
1.1 Características de un robot móvil	1
1.2 Sistemas de navegación y detección de objetos	2
1.3 Entorno institucional	6
1.4 Recursos tecnológicos	7
1.4.1. Trabajo en linux	7
1.5 Problemas y causas	9
1.6 Objetivos	9
CAPÍTULO 2 : OBTENCIÓN Y PROCESAMIENTO DE IMÁGENES	
2.1 Tipo de imágenes	10
2.2 Propiedades de una imagen digital	11
2.2.1. Propiedades métricas y topológicas	11
2.2.2. El histograma	14
2.2.3. Percepción visual de la imagen	15
2.2.4. Ruido en imágenes	16
2.3 Procesamiento de imágenes	17
2.3.1. Etapa de acondicionamientos o pre-procesamiento	17
2.3.1.1. Transformación básica de Imágenes	17
2.3.1.2. Detección de bordes	18
2.3.1.3. Conversión a escala de grises	20
2.3.1.4. Transformaciones geométricas	21
2.3.1.5. Suavización de una imagen	22
2.4 Segmentación	23

2.4.1.	Umbralización.....	23
2.4.2.	Extracción de regiones por color.....	23
2.4.3.	Morfología	24
2.5	Trabajo con cámaras.....	27
2.5.1.	Adquisición de imágenes	27
2.5.2.	Iluminación.	29
CAPÍTULO 3: VISIÓN ESTEREOSCÓPICA Y LOCALIZACIÓN DE OBJETOS		
3.1	Visión estéreo.....	31
3.2	Parámetros Intrínsecos y Extrínsecos.....	33
3.3	Obtención de par estéreo.....	37
3.3.1.	Obtención del cuadro de la cámara	37
3.3.2.	Manejo de los parámetros de la cámara.....	38
3.4	Calibración del sistema estéreo	39
3.5	Geometría epipolar.....	40
3.6	Rectificación	45
3.6.1.	Algoritmo de bouguet.....	45
3.6.2.	Rectificación de las imágenes del par estéreo.....	46
3.7	Estimación de Coordenadas de un Punto.....	47
CAPÍTULO 4 : PRUEBAS RESULTADOS. ANALISIS ECONÓMICO		
4.1	Descripción del proyecto	50
4.2	Requerimientos:	50
4.3	Diseño de la Interfaz de usuario.....	53
4.4	Calibración del sistema estéreo	55
4.5	Rectificación	57
4.6	Obtención de coordenadas	61
4.6.1	Algoritmo de obtención de coordenadas.....	61
4.6.2	Detección y segmentación del objeto.....	62

4.6.3	Coordenadas espaciales.....	64
4.7	Comunicación con el robot móvil	68
4.8	Análisis Económico	70
	CONCLUSIONES	71
	RECOMENDACIONES	72
	BIBLIOGRAFÍA	73
	ANEXOS:	74



INTRODUCCIÓN

Hoy en día debido al avance de la tecnología, muchas veces la capacidad humana se ve ampliamente superada, tal es así que se ha visto la necesidad del desarrollo de herramientas capaces de sustituir al ser humano en tareas que este es incapaz de realizar por sí solo. La visión estereó es una de estas capacidades que se ve en la necesidad de ser replicada en sistemas autónomos para lograr emular el ser humano. Los robots móviles son las principales herramientas en donde un sistema estereó se ve en la necesidad de ser implementado, debido a los requerimientos de manipulación de objetos y control en la navegación. Para lograrlo debemos contar con algoritmos especiales de detección de objetos y así constituir el bloque de visión de un robot móvil.

El presente trabajo busca proponer la implementación de dicho sistema de visión en un robot móvil para probar su desempeño en un área de trabajo determinada así poder demostrar el gran avance que podría brindar en el futuro el uso de dichos sistemas.

El desarrollo del sistema de visión propuesto incluye el estudio del campo de aplicación para nuestro caso de los robots móviles, los diferentes algoritmos empleados en la segmentación de un objeto y por último la integración de estos algoritmos con un sistema estereó para probar la eficiencia de este último.

El trabajo de tesis presentado está dividido en 4 capítulos, en el primero de ellos veremos una aproximación al campo de robots móviles su aplicación así como la diferentes tecnologías que se aplican en ellos. El segundo capítulo, estará enfocado en el estudio de las diferentes técnicas de procesamiento de imágenes, así como su análisis. El tercer capítulo, está dedicado al entendimiento del proceso de visión estereó, y los diferentes procesos que se necesitan para elaborar el algoritmo del sistema. Finalmente en el cuarto capítulo se realizará una evaluación de los resultados del sistema de visión estereó en el robot móvil.

CAPÍTULO 1

DESCRIPCIÓN DE ROBOT MÓVIL

1.1 Características de un robot móvil

Los robots móviles son máquinas herramientas capaces de realizar tareas de forma eficiente, flexible y robusta de similar forma que los humanos. Los robots son especialmente útiles en ambientes de riesgo para los humanos como son conductos de gas, exploración de nuevos planetas, lugares hostiles, etc.

Para que un robot sea considerado móvil debe tener la característica fundamental de desplazamiento de un punto a otro mediante un sistema de locomoción.

Un robot móvil dependiendo de su aplicación debe cumplir ciertos requerimientos, como son en su transporte, rango de operación, modularidad, si son tele operados o no, etc. Su medio de transporte pueden ser mediante un sistema de orugas o de ruedas inclusive se pueden dar casos de ser guiados mediante un sistema de rieles, existen diversas formas de controlar el movimiento de un robot móvil puesto que, pueden ser completamente autónomos como es el caso de los robots enviados a la exploración de otros planetas o pueden ser manipulados de forma remota mediante un sistema inalámbrico como es nuestro caso.

Hoy en día empresas como Sony, Intel, General Motors han implementado unidades robóticas en sus líneas de producción, como resultado de esto se ha logrado tener un mayor ahorro de recursos, agilización en la producción y la mejora notable del rendimiento general reflejado en las ganancias de la empresa.

Como ejemplos notables del impacto de los robots móviles en nuestra sociedad tenemos al inspector Bot3 dedicado a la limpieza de tuberías y conductos, el robot de vigilancia conocido como Roborior; además de robots capaces de rociar agua en cultivos de forma autónomas como se puede ver en la figura 1.1.



Figura 1.1 (izquierda a derecha) Inspector Bot.3, Roborior robot de vigilancia, vehículo autónomo de Jhon Deere

1.2 Sistemas de navegación y detección de objetos

El problema de brindar información específica acerca del entorno en el cual se va desenvolver el robot móvil ha dado a paso a diversas soluciones, desde el uso de sensores infrarrojos hasta los escáneres; sin embargo el uso de un sistema u otro responde a necesidades de funcionalidad, costo y diseño.

Los sensores de menor costo son los de contacto, los cuales necesitan tener un contacto directo con el objeto para poder detectarlo. Dentro de este tipo de sensor encontramos a los micros interruptores, sensores de reconocimiento táctil, etc. Una de las principales ventajas como ya se mencionó antes es su costo, además de su fácil fabricación y manejo; sin embargo esta clase de sensores provee información limitada aparte de tener un tiempo de vida relativamente corto.(HU, y otros, 2005)

Otra clase o tipo de sensores usados en los robots móviles son los de no contacto, dentro de los más usados se encuentran los sensores infrarrojos, los cuales trabajan emitiendo luz infrarroja y midiendo la variación de frecuencia y fase de la onda de rebote; sin embargo solo nos pueden dar información en una dimensión y presencia o no de un objeto; para lograr obtener la ubicación exacta es necesario el uso de algoritmos de triangulación, con lo cual el grado de complejidad se incrementa. El uso general que se le suele dar a este tipo de sensor es para evitar colisiones con otros objetos y la obtención de una ruta segura. Su principal desventaja es la falta de

capacidad para la distinción de objetos entre sí, cosa que dificulta la labor de seguimiento de objetos (HU, y otros, 2005).

Los sensores de ultrasonido a diferencia de los anteriores si pueden brindar información de la distancia a la cual se encuentran el objeto u obstáculo mediante el uso de las propiedades de las ondas de sonido llámese efecto Doppler. Una ventaja de este sistema es la capacidad de hacer un mapeo del entorno en el cual se encuentra el robot mediante el arreglo de sensores, donde cada uno de estos cubre un campo de acción; la desventaja de este sistema viene a ser dada por los ecos falsos debido al rebote en diferentes objetos antes de llegar al receptor, otro problema se suscita cuando se tiene más de un robot con sensores ultrasónicos trabajando en un mismo entorno lo cual puede llegar a provocar varios problemas de traslape de datos.(OROSCO, y otros, 2001)

El sistema de la localización basado en los scanner láser y las señales artificiales es una técnica de ubicación absoluta prometedora en términos de funcionamiento y costo, para esta tecnología las coordenadas de las señales artificiales son pregrabadas en un mapa del ambiente las cuales van a ser detectadas por el robot móvil. Las principales desventajas de este sistema son en primer lugar que este sistema no trabaja bien cuando las señales artificiales cambian accidentalmente sus posiciones y el segundo problema se presenta cuando se escanea superficies del terreno desigual por tanto la re-calibración se debe realizar de forma continua.(HU, y otros, 2005).

Entre los instrumentos usados para determinar el posicionamiento relativo y absoluto del robot, dedicados a solucionar el problema de la navegación y conocer la posición de robot móvil en cada instante de tiempo están: los codificadores rotatorios basados en el cálculo del desplazamiento de las llantas mediante el número de giros, los sensores magnéticos como una brújula electrónica con la cual ya podemos saber el Angulo de giro del robot y con esto la dirección y trayectoria, hasta llegar elementos tan avanzados como los GPS para determinar el posicionamiento global del robot.(OROSCO, y otros, 2001).

El sistema de adquisición de datos basado en visión estéreo es aquel sistema capaz de obtener información relevante del entorno de trabajo a partir de dos imágenes del mismo y obteniendo las coordenadas de ubicación basándose en las diferencias entre dichas imágenes. La única desventaja del sistema de procesamiento de imágenes es precisamente el uso de algoritmos complejos para su elaboración; sin embargo ya existen algoritmos preestablecidos que se pueden usar y mejorar para determinadas aplicaciones.

Uno de los problemas inmediatos que surge con el tratamiento de imágenes en un sistema estéreo es el problema de correspondencia de puntos, debido a muchos factores por ejemplo distorsión por lentes, oclusión u ocultamientos de puntos etc. Para solucionar estos problemas se han dado diferentes soluciones como son la implementación de algoritmos de restricción de área, reconocimiento de patrones, lógica difusa, etc.

Una de las clasificaciones para estos algoritmos se da dependiendo de las restricciones que explota. Los llamados métodos locales son aquellos que aplican las restricciones a un determinado número de píxeles alrededor de un píxel de estudio. Son muy eficientes pero muy sensitivos a las ambigüedades locales (ejemplo regiones de oclusión o regiones con texturas uniformes), dentro de este grupo se encuentran los métodos basados en área, características y gradiente. Los métodos globales a diferencia de los locales se basan en aplicar restricciones a la línea de imagen o a la imagen completa, son menos sensitivos que los locales pero son más elaborados.(RUIZ MORENO, 2005).

Respecto a la locomoción de los robots móviles los sistemas más populares son los dispuestos por ruedas los cuales son más fáciles de montar y controlar ,la desventaja de estos sistemas vienen a estar en los caminos de forma irregular debido a que un vehículo de este tipo podrá sobrepasar los obstáculos que tengan máximo la altura de sus ruedas, una buena solución a este problema son las ruedas por cadena u orugas las cuales son mucho más resistentes y se adaptan mejor al terreno; sin embargo tienen problema de deslizamiento. Una solución más adecuada para estos problemas viene son los robots implementados con articulaciones los cuales pueden superar casi cualquier obstáculo y no adolecen del problema de deslizamiento; pero una

implementación de este tipo suele ser muy costosa y el control del grado de libertad de dichas articulaciones complica la programación.(R. GONZALES, 2002)



1.3 Entorno institucional

En nuestro país la mayoría de trabajos aun se realiza en forma manual como son las selección de objetos y el control de calidad de los productos, los cuales están sujetos básicamente fallas humanas y además en condiciones adversas aumentan la probabilidad de error; entonces es necesario la implementación de un sistema capaz de poder dar un reconocimiento fiel del entorno en donde se va a trabajar.

El uso de sistemas robotizados en la industria peruana es limitado debido al costo que implica el uso y mantenimiento del mismo; sin embargo existen empresas transnacionales como es el caso de Gloria S.A. que tienen implementado estos sistemas dentro de sus líneas de producción y para el control de calidad de su producto final; lo cual demuestra la aplicación de dichos sistemas.(Soudronic AG)

Dentro del campo de investigación de robótica y visión por computadora existen diferentes convenios entre universidades de todo el mundo para fomentar la investigación y desarrollo de estos sistemas. Una de estas redes de intercambio es RIBERO, Red Iberoamericana de Robótica, la cual está conformada por las diferentes universidades de países como Argentina, Brasil, Chile, Colombia, Perú etc.(CYTED, 2002) Otro de estos grupos son Centro de Desarrollo e Investigación en Mecatrónica (CEDIM), la Sociedad de Robótica y Automatización (RAS) etc.

Actualmente, existen además diversos concursos en donde aparte de competir y poner a prueba los conocimientos adquiridos, se pueden intercambiar conocimientos y experiencias. Uno de estos concursos es el LARC el cual se lleva a cabo cada año y tiene como objetivo reunir y poner nuevos desafíos a todos los competidores en las diferentes categorías.(Latin American Robotics Council, 2006)

1.4 Recursos tecnológicos

Para el desarrollo de este trabajo se cuenta con diferentes alternativas en relación con sistemas operativos, librerías, lenguaje de programación etc. Para el sistema de visión estéreo se va a trabajar con dos importantes bloques, en los cuales se va a soportar nuestro sistema, estos bloques son la librería OpenCV desarrollada por Intel y el sistema operativo Linux (Ubuntu 9.4).

OpenCV, una herramienta de visión y procesamiento de imágenes desarrollada por la compañía Intel, es una librería de herramientas multiplataforma puede trabajar en diversos sistema operativos como son Mac OS X, Windows y Linux. Además de estar bajo licencia BSD¹ lo cual permite que sea usada bajo propósitos comerciales y de investigación como es nuestro caso.

En el caso del sistema operativo, Ubuntu 9.4, se decidió trabajar bajo este sistema debido a su forma de trabajo de los diferentes periféricos como son en nuestro caso las cámaras USB. Su forma de montaje de estos dispositivos es a través de módulos denominados video0 y video1, lo cual permite tener un mejor acceso y direccionamiento hacia las propiedades de las cámaras sin necesidad de tener un driver instalado como es el caso del sistema operativo Windows.

1.4.1. Trabajo en linux

El programa de Visión Estéreo esta desarrollado en forma integra en el lenguaje de programación C++, el cual es uno de los lenguajes soportados por el compilador de Linux GCC. Para poder tener un mejor entorno de programación se usara la plataforma de desarrollo CodeBlocks con la cual vamos a poder tener un entorno más amigable y eficiente.

¹ La licencia BSD es la licencia de software otorgada principalmente para los sistemas BSD (Berkeley Software Distribution). Es una licencia de software libre permisiva como la licencia de OpenSSL o la MIT License. Esta licencia tiene menos restricciones en comparación con otras como la GPL estando muy cercana al dominio público. La licencia BSD al contrario que la GPL permite el uso del código fuente en software no libre.

Para poder trabajar con la biblioteca de herramientas OpenCV y en Linux vamos a tener que realizar el siguiente procedimiento.

- Tener el compilador de Linux GCC 4.0
- Contar con las librerías de reconocimiento de imágenes tipo JPG, PNG, TIFF etc.
- Soporte para estándar IEEE1394 para el trabajo con cámaras Firewire.
- Descargar y compilar la última versión probada de OpenCV.
- De darse el caso de trabajar con cámaras ImageSource® instalar el paquete de datos UNICAP.
- Para poder trabajar en Codeblocks debemos indicarle a este entorno la dirección de las librerías de la biblioteca OpenCV.

La librería OpenCV, la cual fue usada en el sistema de visión del vehículo no tripulado Stanley de la Universidad de Stanford (MEDIAWIKI), tiene un conjunto de instrucciones capaces de obtener de una forma muy eficiente tanto los parámetros del sistema estéreo como la rectificación de las imágenes, para poder luego obtener las coordenadas espaciales de nuestro objeto; por lo tanto es la herramienta ideal para trabajar en este tipo de proyectos.

En relación al hardware del sistema este se consiguió gracias al proyecto DAI 2008-0049, “Diseño E Implementación De Un Sistema Terrestre De Detección Y Clasificación De Nutrientes En Campos De Cultivos”, mediante el cual se financio la adquisición de un robot móvil profesional modelo P3-AT de active media®, además del conjunto de cámaras para la visión estéreo de la empresa The Imaging Source®, especialista en equipos para este tipo de aplicaciones.

El robot cuenta con su propio algoritmo de desplazamiento y manipulación de objetos, basado en la librería ARIA, así como el sistema de transporte en este caso mediante cuatro ruedas; por otro lado el sistema de visión estéreo será desarrollado de forma completa contando para esto de una placa gigabyte® modelo GA-EG31M-S2 y de un procesador Intel Quadcore como base para el procesamiento de imágenes.

1.5 Problemas y causas

Diversos sistemas de control de calidad aun no cuentan con mecanismos de selección y clasificación automáticos, un claro ejemplo de esto está en la agricultura peruana, la selección de los productos aun se hace a criterio humano y los estándares actuales de calidad exigen más que una simple inspección visual, exigen tamaño forma peso color textura.

El uso de sistemas de visión como sistema de navegación aun no es muy utilizado debido al costo tanto de hardware como de software que implica su desarrollo; sin embargo estos sistemas de visión son los más precisos hasta ahora; sin embargo dependen mucho del lente de la cámara y son bien inestables al cambio de iluminación del ambiente; por otro lado ya se han desarrollado algoritmos con el fin de compensar esta desventaja.

La carencia de una política de estado bien definida que favorezca a la implementación de nuevos sistemas en las compañías del estado, debido al gran costo que esto supone, prefiriendo el camino de la privatización que al desarrollo de nuevas tecnologías; todo esto sumando al poco interés del gobierno para apoyar a grupos de investigación en los diferentes temas de desarrollo hacen ver que nuestra realidad es muy diferente que la de otros países.

1.6 Objetivos

Una vez estudiado los problemas que se presentan para la implementación de dicho sistema los objetivos principales a cumplir son:

Brindar al robot un sistema de visión estereoscópica para la localización de objetos, para lograr esto es necesario cumplir con ciertas metas como es la captura eficiente del par estéreo, obtención de los parámetros intrínsecos y extrínsecos de las cámaras, rectificación de las imágenes y obtención de la profundidad.

Obtener las coordenadas XYZ del centro de un objeto de color determinado y lograr la comunicación con el robot enviando los datos de forma eficiente.

CAPÍTULO 2

OBTENCIÓN Y PROCESAMIENTO DE IMÁGENES

2.1 Tipo de imágenes

Una imagen puede ser definida por una función bidimensional $f(x,y)$ donde x e y son las coordenadas espaciales y la amplitud de f en cualquier par de coordenadas es llamada la intensidad de la imagen en dicho punto (C. GONZALES, y otros, 2004). Según la manera de representar una imagen estas pueden ser imágenes de intensidad, imágenes indexadas, imágenes binarias, RGB.

Una imagen de intensidad, es el equivalente a la conocida imagen en escala de grises y además es el tipo de imagen con el cual se trabaja más. Este tipo de imagen representada en una matriz de una dimensión donde cada elemento tiene un valor correspondiente a que tan claro u oscuro es dicho elemento denominado píxel. Hay dos maneras de representar el valor de un píxel en una imagen de intensidad: De tipo real el cual asigna un número flotante entre 0 y 1 donde 0 corresponde al valor negro y uno al valor blanco. El otro tipo es el tipo entero, en el cual se asigna un valor entre 0 y 255 para representar el brillo de un píxel, la diferencia entre ambas radica en que la tipo entero necesita un octavo de memoria menos que la tipo real; sin embargo varias funciones matemáticas solo pueden ser aplicadas al tipo real.

Una imagen binaria solo carga la imagen en una matriz en donde el píxel puede tomar solo dos valores tanto 0 o 1 donde 0 es negro y 1 es blanco. (C. GONZALES, y otros, 2004); Este tipo de imágenes es usado para realizar operaciones como segmentación, erosión y dilatación etc.

Imagen RGB es una forma de representar una imagen en tres matrices del mismo tamaño. Cada matriz corresponde a uno de los colores rojo, verde o azul y cada uno nos da una instrucción de cuanto de cada uno de estos colores contiene cada píxel, por esta razón la mayoría de las técnicas desarrolladas para imágenes monocromas

pueden ser aplicadas a imágenes a color mediante el procesamiento de sus tres componentes de forma individual (C. GONZALES, y otros, 2004)

2.2 Propiedades de una imagen digital

Una imagen digital tiene muchas propiedades como son las propiedades métricas, las propiedades topológicas y la percepción humana como criterio para la calidad de la imagen.

2.2.1. Propiedades métricas y topológicas

Como ya se ha dicho una imagen digital es representada por una matriz bidimensional cuyos elementos son números correspondientes a la cuantización de niveles en la escala de brillo.

Una de las propiedades es la distancia entre dos puntos de la imagen, dicha distancia puede ser definida de diferentes maneras como por ejemplo la distancia euclidiana, “The City Block” y la distancia “tablero de ajedrez”.

La distancia euclidiana es conocida por la clásica fórmula geométrica:

$$D_e = \sqrt{(i - h)^2 + (j - k)^2}$$

Donde:

(i, j) : Coordenadas del punto inicial

(h, k) : Coordenadas del punto final

(2.1)

La ventaja de la distancia euclidiana es que resulta algo intuitivo sin embargo el costo de cálculo debido a la raíz cuadrada es alto. La distancia entre dos puntos puede ser expresada también como el mínimo número de movimientos horizontales y verticales para poder llegar al otro punto, esta medida de la distancia es la llamada distancia “City Block” y es definida por:

$$D_{city} = |i - h| + |j - k|.$$

(2.2)

Si dejamos que solo los movimientos en diagonal sean posibles entonces obtendremos la distancia llamada tablero de ajedrez; puesto que la distancia es igual al mínimo número de movimientos que debe hacer el rey de un punto hacia otro, esta distancia está definida por la expresión:

$$D_{\text{tablero}} = \max\{|i - h|, |j - k|\}$$

(2.1)

Estas diferentes definiciones de distancias están ilustradas en la figura 2.1:

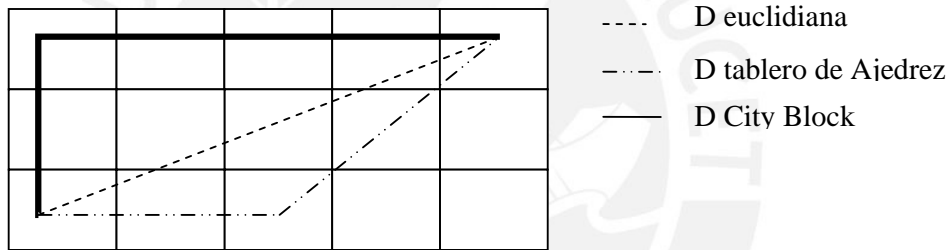


Figura 2.1 distancias métricas D euclidiana, D tablero de ajedrez y D city block

La proximidad es otro concepto importante en imágenes digitales. Si la distancia entre dos píxeles es igual a la unidad entonces dichos puntos son puntos vecinos, si no se consideran las distancias diagonales tenemos una vecindad de 4 puntos si son consideradas entonces tenemos una vecindad de 8 puntos tal como se ve en la figura 2.2.

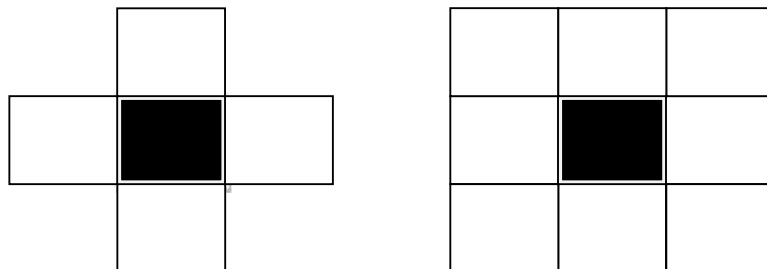


Figura 2.2 vecindad de un pixel representativo (pixel negro)

Si existe una conexión entre dos pixeles estos pixeles son denominados contiguos. Esta relación 'de ser contiguos' va permitir en muchos casos segmentar una imagen en varias regiones individuales.

Los bordes de una imagen o región mostrados en la figura 2.3, es otro concepto importante en el análisis de una imagen. En una región R los bordes de dicha región son el conjunto de pixeles dentro de la región que tienen uno o más pixeles vecinos fuera de la región R.

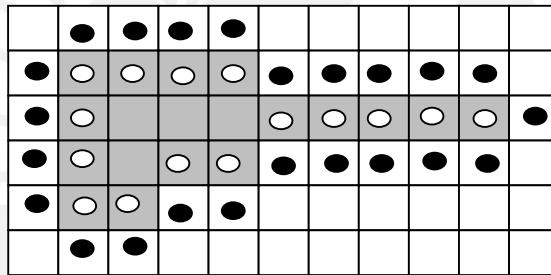


Figura 2.3 Bordes interiores y exteriores de una región, mostrados como círculos blancos y círculos negros.

Las propiedades topológicas de una imagen ya no están basadas en la distancia entre dos puntos; si imaginamos un balón de caucho con un objeto pintado en el, las propiedades topológicas del objeto son aquellas que con son invariantes al estiramiento de dicho balón, este estiramiento no cambia propiedades como continuidad del objeto y tampoco características como cantidad de agujeros en la región.

Un objeto con una forma irregular puede ser representado por el grupo de sus componentes topológicos ejemplificados en la figura2.4. La zona de la envolvente convexa que no pertenece a la Región es denominada déficit de la convexidad. Esta zona puede ser dividida en dos subzonas: lagos (pintada de gris oscuro) está

totalmente rodeada por el objeto; y las bahías (gris claro) son zonas que están continuas al borde de la envolvente convexa.

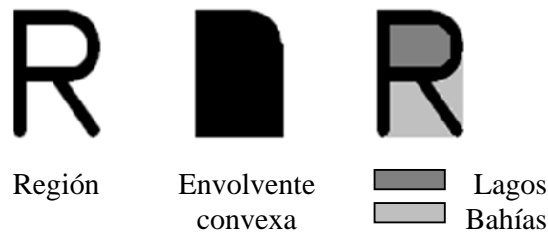


Figura 2.4 Descripción de una forma usando los componentes topológicos

La envolvente convexa, lagos y bahías son muchas veces usadas para la descripción de un objeto o también para las operaciones morfológicas. (SONKA, y otros, 2008).

2.2.2. El histograma

El histograma de brillo $h_r(z)$ de una imagen provee la frecuencia del valor del brillo en la imagen. El histograma para una imagen de escala de grises esta representado por un arreglo dimensional de L elementos. El histograma vincula la imagen y la descripción probabilística del brillo de un pixel(x,y). El histograma normalmente es mostrado como un gráfico de barras como se ve en la figura 2.5.

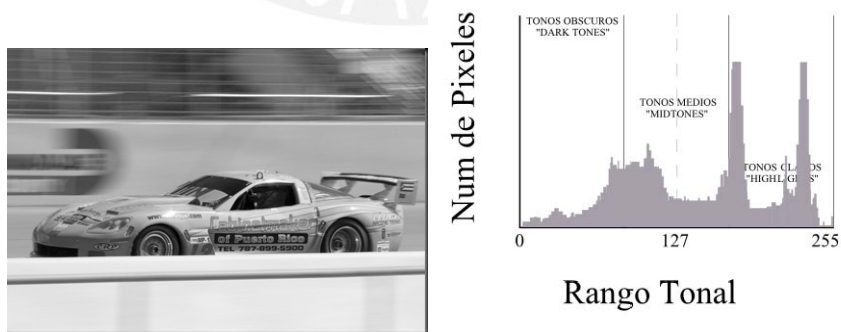


Figura 2.5 Imagen original y su histograma de brillo

El histograma es usualmente la única información global disponible que se tiene acerca de la imagen y es usado en busca condiciones de iluminación óptima para la

captura de una imagen, transformaciones en escala de grises y segmentación de imágenes.

2.2.3. Percepción visual de la imagen

Si una imagen debe ser analizada por una persona, dicha imagen debe tener información fácil de percibir para la ella; estos parámetros son el contraste, los bordes, la forma, textura, color etc.

El contraste es el cambio local del brillo y está definido como la proporción entre el promedio del brillo de un objeto y el fondo. El ojo humano es logarítmicamente sensible al brillo, esto implica que para una misma percepción, a mayor brillo requiere mayor contraste. El brillo aparente depende en gran medida del brillo de los alrededores del objeto que del objeto en sí, a este efecto se le denomina contraste condicional, la figura 2.6 nos puede ilustrar mejor este punto.

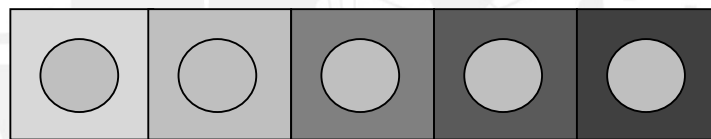


Figura 2.6 Contraste condicional. Los círculos dentro de los cuadrados tienen el mismo nivel de brillo

La habilidad de detectar detalles en una imagen, es denominada agudeza. La resolución en una imagen está delimitada por la habilidad de resolución del ojo humano; no hay ningún sentido en representar una imagen con una resolución superior a la del observador. La agudeza está relacionada también a la distancia del objeto, mientras más alejado este el objeto la agudeza disminuye.(SONKA, y otros, 2008)

2.2.4. Ruido en imágenes

Las imágenes reales casi siempre son degradadas por algunos errores aleatorios. Esta degradación es usualmente llamada ruido. El ruido puede ocurrir durante la captura de la imagen, transmisión, o procesamiento, quizás dependa o sea independiente del contenido de la imagen.

El ruido es normalmente descrito por sus características probabilísticas. El ruido ideal denominado ruido blanco, este tipo de ruido está presente en todas las frecuencias con la misma intensidad. Un caso especial de ruido blanco es el ruido Gaussiano, en donde la variable aleatoria tiene una distribución Gaussiana. (SONKA, y otros, 2008)

El ruido impulsivo o sal y pimienta se caracteriza porque los píxeles de la imagen son muy diferentes en color o intensidad a los píxeles circundantes. Generalmente, este tipo de ruido, afectará a una pequeña cantidad de píxeles de la imagen. Al ver la imagen, encontraremos puntos blancos sobre puntos negros o puntos negros sobre puntos blancos mostrados en la figura 2.7, de ahí el término sal y pimienta. (MEDIAWIKI)



Figura 2.7 Ruido sal y pimienta

2.3 Procesamiento de imágenes.

El procesamiento de imágenes es la aplicación de diferentes técnicas a una imagen con el objetivo de mejorar la calidad para la interpretación humana y tener un procesamiento de datos de forma autónoma. (MARTINSANZ, y otros, 2008)

2.3.1. Etapa de acondicionamientos o pre-procesamiento

Existen diferentes motivos por los cuales se tiene que modificar una imagen, uno de los principales es la necesidad de considerar las estructuras subyacentes, por ejemplo los bordes o esquinas. Otro de los motivos por el cual se modifica la información está ligado al mejoramiento de la imagen o a la corrección de la intensidad de la misma.

2.3.1.1. Transformación básica de Imágenes

El objetivo de cualquier transformación es preparar la imagen para un posterior análisis y su posterior interpretación.

Dentro de las diferentes operaciones básicas de transformación están, las operaciones individuales las cuales implican la generación de una nueva imagen modificando el valor del pixel en una simple localización, basándose en una regla global aplicada a cada pixel de la imagen original. Dicho proceso consta en obtener el valor del pixel de una localización dada en la imagen, modificándolo por una operación lineal o no lineal.

Existen diferentes tipos de operadores de acuerdo a su función de transformación, como es el caso de los operadores identidad el cual crea una imagen igual a la imagen de entrada, inverso o negativo que nos da una imagen de salida que es la inversa de la imagen de entrada como se ve en la figura 2.8, umbral esta clase de transformación crea una imagen de salida binaria a partir de una imagen de grises. (MARTINSANZ, y otros, 2008)



Figura 2.8 Aplicación de filtro negativo

2.3.1.2. Detección de bordes

Las operaciones de vecindad a diferencia de las operaciones individuales utilizan el mismo procedimiento excepto que el nuevo valor del pixel en la imagen de salida depende de una combinación de los valores de los pixeles en la vecindad. Gracias a las operaciones de vecindad, podemos obtener de una imagen los bordes de esta, así como también las esquinas o en todo caso mejorar el contraste (MARTINSANZ, y otros, 2008)

La detección de bordes es una colección de métodos utilizados para encontrar cambios en la intensidad; la idea es encontrar pixeles cuyo brillo cambia abruptamente.

Un borde o contorno es una propiedad unida a un pixel y es calculada por el comportamiento de cambio de la imagen en la vecindad del pixel representado por dos componentes, magnitud y dirección. La magnitud de dicho vector es igual a la magnitud del gradiente y su dirección esta desplazada -90° con respecto a la dirección del gradiente.

El gradiente de una imagen en el punto (x, y) se define como un vector bidimensional dado por la siguiente ecuación:

$$\Delta f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

(2.4)

Y la magnitud de este vector es:

$$|\Delta f| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$$

2.2

Con una dirección:

$$\theta = \arctan\left(\frac{\frac{\partial f}{\partial y}}{\frac{\partial f}{\partial x}}\right)$$

2.3

Para aproximar digitalmente las derivadas de primer o segundo orden de una imagen se trabajan con mascarar en donde a cada pixel de la vecindad se le agrega un peso, y el resultado de la sumatoria de esos pesos nos da el valor de la derivada. En la siguiente imagen tenemos las diferentes mascarar para cada tipo de operador mostradas en la figura 2.9. (SONKA, y otros, 2008)

Z1	Z2	Z3
Z4	Z5	Z6
Z7	Z8	Z9

Vecindad de la imagen

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
1	0	1

Operador Sobel

$$G_x = (z7 + 2z8 + z9) - (z1 + 2z2 + z3)$$

$$G_y = (z3 + 2z6 + z9) - (z1 + 2z4 + z7)$$

-1	-1	-1
0	0	0
1	1	1

-1	0	1
-1	0	1
1	0	1

Operador Prewitt

$$G_x = (z7 + z8 + z9) - (z1 + z2 + z3)$$

$$G_y = (z3 + z6 + z9) - (z1 + z4 + z7)$$

Figura 2.9 Algunas de las mascararas utilizadas por los operadores

El resultado de aplicar los diferentes operadores a una imagen lo podemos observar en la figura 2.10.



Figura 2.10 Resultado de aplicar el operador Sobel a una imagen

La detección de contornos es usada generalmente para el análisis de una imagen para determinar los límites de una región con brillo homogéneo.

2.3.1.3. Conversión a escala de grises

La conversión a una imagen de escala de grises es otra de las transformaciones más comunes en el procesamiento de imágenes. Una imagen digital tiene un número limitado de niveles de grises, es por eso que las transformaciones de escala de grises son fáciles de realizar tanto en hardware como en software.

Las transformaciones en escala de grises son usadas mayormente cuando una imagen al ser mejorado su contraste puede ser interpretada de forma más fácil. La técnica de ecualización por histograma es una transformación de la imagen cuyo objetivo o principal es crear una imagen con igual distribución de niveles de brillo.

La ecualización por histograma mejora el contraste para valores de brillo cercanos al nivel máximo del histograma y disminuye contraste para los valores cerca del mínimo valor del histograma. El resultado de esta operación se puede ver en la figura 2.11. (SONKA, y otros, 2008)

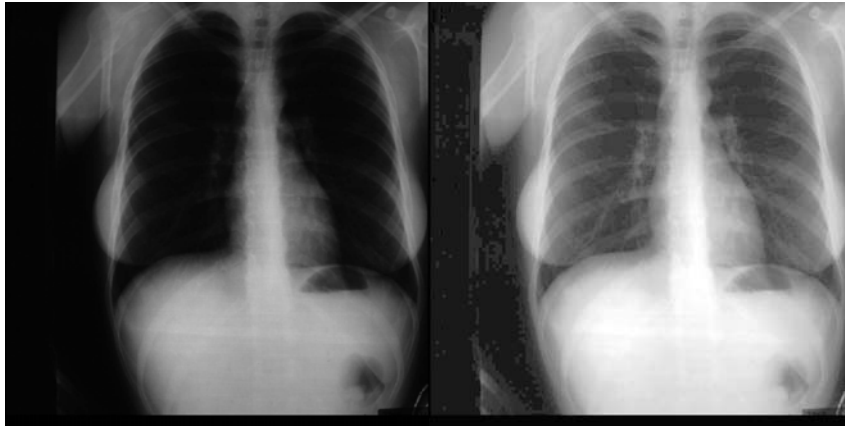


Figura 2.11 Como resultado de la ecualización por histograma los detalles de la placa son más notorios

2.3.1.4. Transformaciones geométricas

Este tipo de transformaciones son muy usadas en gráficos computarizados y en análisis de imágenes. Estas transformaciones permiten eliminar la distorsión geométrica que se da en la captura de una imagen. Un ejemplo de esto es su aplicación para emparejar dos imágenes de un mismo objeto tomada con diferentes cámaras.

La transformación geométrica es una función T vector que traslada un píxel (x, y) hacia una posición (x', y') . T está definida por dos ecuaciones en el eje x y en el eje y . la figura 2.12 ilustra como toda la región de una imagen es transformada punto a punto.

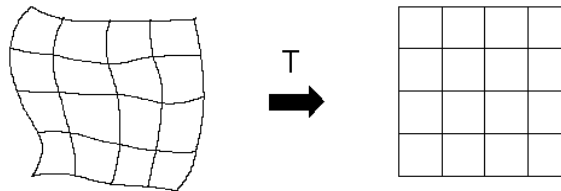


Figura 2.12 Transformación geométrica en un plano

La transformación geométrica consiste en dos etapas básicas. La primera de ellas es la transformación de las coordenadas del píxel, en donde las coordenadas de los píxeles en la imagen de entrada son trasladadas a su nueva posición en la imagen de salida, donde las nuevas coordenadas de salida deben tener valores reales continuos.

La segunda etapa es la interpolación del brillo de varios puntos en la vecindad.
(SONKA, y otros, 2008)

2.3.1.5. Suavización de una imagen

La suavización de una imagen está referida al conjunto de métodos cuyo uso está dirigido a la eliminación del ruido en una imagen como se puede ver en la figura 2.13.



Figura 2.13 (izquierda) imagen afectada por ruido (derecha) imagen filtrada

El promediado es una técnica directa en el dominio espacial. Dada una imagen G se obtiene una imagen F cuya intensidad para cada punto resulta ser el promediado de los valores de intensidad de los píxeles de G incluidos en el entorno de vecindad. Un caso particular del promediado es el suavizado Gaussiano, en el cual el núcleo es una gaussiana 2-D, de media 0 y desviación estándar σ de la forma.(MARTINSANZ, y otros, 2008 págs. 86,87)

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

2.4

Donde x e y son las coordenadas de las imágenes y σ es la desviación estándar. La desviación estándar es el único parámetro del filtro Gaussiano y es proporcional al tamaño de la vecindad en donde se va aplicar el filtro. Los píxeles más distantes del centro del operador tienen una menor influencia. En la figura 2.14 podemos ver una máscara de 7x7 de un filtro Gaussiano.(SONKA, y otros, 2008)

0	1	4	5	4	1	0
1	9	23	29	23	9	1
4	23	61	78	61	23	4
5	29	78	100	78	29	5
4	23	61	78	61	23	4
1	9	23	29	23	9	1
0	1	4	5	4	1	0

Figura 2.14 mascara de convolución gaussiana de 7x7

2.4 Segmentación

La segmentación consiste en subdividir una imagen en sus regiones u objetos. El nivel o precisión con la cual la imagen se va dividir depende del problema a resolver; entonces la segmentación debe detenerse cuando el objeto de búsqueda ha sido separado por completo. (C. GONZALES, y otros, 2004)

2.4.1. Umbralización

El más fácil de los procesos de segmentación es la Umbralización. Los objetos o regiones tienen como característica la reflexión o absorción de luz en sus superficies; la constante de brillo o umbral puede determinarse para segmentar objetos y el fondo.

La Umbralización tiene un coste computacional bajo y rápido, puede ser fácilmente realizada en aplicaciones en tiempo real usando hardware especializado.

2.4.2. Extracción de regiones por color

El uso del color en el procesamiento de imágenes está motivado por dos factores principales, a menudo simplifica la identificación y extracción de objetos de una escena. Para nuestro caso usaremos la segmentación de un objeto donde predomine el color rojo por encima de otros colores. En el caso de la figura 2.15 presentamos la segmentación del color verde. (MARTINSANZ, y otros, 2008)



Figura 2.15 Segmentación de color Verde

Una de las técnicas más simples para lograr la segmentación por color es trabajar con las tres componentes de una imagen RGB; para la segmentación de un color en particular por ejemplo el color rojo en la componente roja de la imagen el valor del color rojo puro es 255 mientras este mismo color en las otras capas tiene un valor mucho menor, dependiendo de las características de captura de imagen. El resultado de esto se puede ver en la figura 2.16.

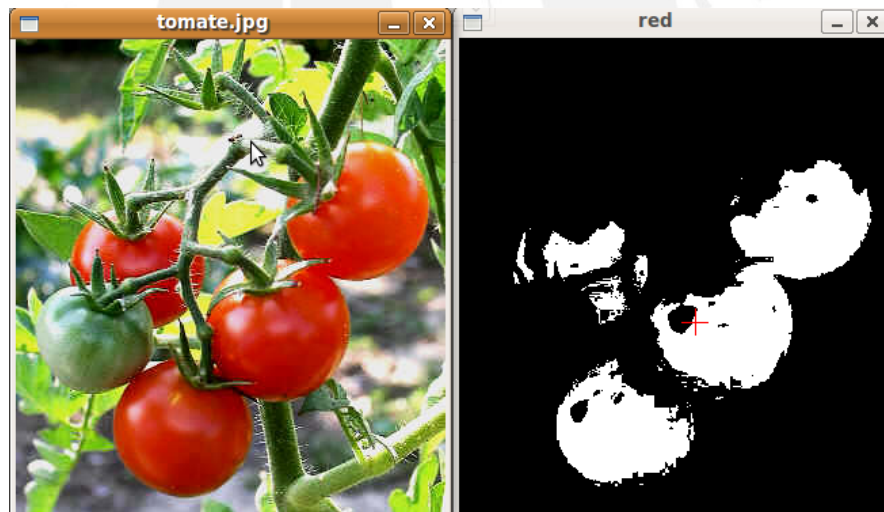


Figura 2.16 Segmentación de los tomates por color

2.4.3. Morfología

Las operaciones de morfología matemática están orientadas al tratamiento de imágenes binarias con operaciones matemáticas basadas en la teoría de conjuntos. Estas operaciones permiten extraer componentes de la imagen que son útiles en la representación y descripción de la forma de las regiones, fronteras y esqueletos,

además permite obtener características relevantes de los objetos de la imagen como forma y tamaño. Las operaciones más conocidas son la dilatación y erosión.

La dilatación es una operación que hace “crecer” los objetos de una imagen binaria esta modificación está determinada por el llamado elemento estructurante el cual actúa como un patrón o guía para determinar el “crecimiento”.

La operación morfológica de dilatación combina dos grupos de elementos con la operación suma de vectores, la figura 2.17 nos ilustra un ejemplo de dilatación.

$$\begin{aligned}
 X &= \{(1,0), (1,1), (1,2), (2,2), (0,3), (0,4)\} \\
 B &= \{(0,0), (1,0)\} \\
 X \oplus B &= \{(1,0), (1,1), (1,2), (2,2), (0,3), (0,4), (2,0), (2,1), (2,2), (3,2), (1,3), (1,4)\}
 \end{aligned}$$

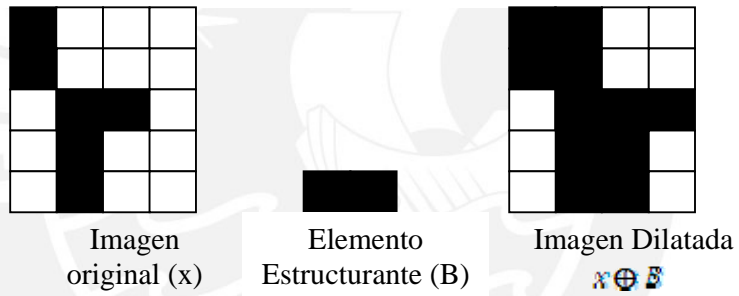


Figura 2.17 Operación de Dilatación

La operación de erosión al contrario de la dilatación permite contraer la imagen; pero al igual que la dilatación también cuenta con un elemento estructurante; sin embargo en este caso la operación de vectores es la sustracción. Podemos ver el resultado de dichas operaciones en la figura 2.18.

$$\begin{aligned}
 X &= \{(1,0), (1,1), (1,2), (0,3), (2,3), (3,3), (1,4)\} \\
 B &= \{(0,0), (1,0)\} \\
 X - B &= \{(0,3), (1,3), (2,3)\}
 \end{aligned}$$

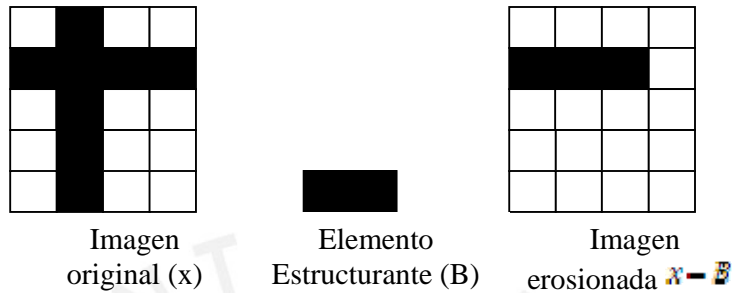


Figura 2.18 Operación de Erosión

La erosión y dilatación no son operaciones inversas esto quiere decir que si primero se dilata una imagen y luego se la erosiona el resultado no es la imagen original.

Si la operación de erosión es seguida por la dilatación, el conjunto de operaciones se denomina operación de apertura. De modo inverso si la operación de erosión es seguida por la operación de dilatación, se denomina operación de cierre.

Las operaciones de apertura y cierre sirven para eliminar detalles inservibles o elementos más pequeños que el elemento estructurante. Se puede observar el resultado de la combinación de estas operaciones en la figura 2.19.



Figura 2.19 resultado de las operaciones de dilatación y erosión

2.5 Trabajo con cámaras

Existen diferentes tipos de cámaras dependiendo de factores como resolución de imágenes, si son digitales o analógicas y del tipo de imagen que entregan, tenemos una amplia gama de productos en el mercado; sin embargo todas entregan formatos de imagen estándar. Y dependiendo del tipo de iluminación que reciba el objeto se resaltarán o no ciertos detalles.

2.5.1. Adquisición de imágenes

Para la adquisición de imagen digital se necesitan tres componentes de hardware, una cámara, un grabador de cuadros, y una computadora en donde se hará el procesamiento.

Toda cámara tiene en su interior un arreglo de $N \times M$ fotosensores cada uno de ellos sensible a la intensidad de luz. Cada sensor (Figura 2.20) puede ser tomado como una minúscula caja negra que convierte la energía luminosa en voltaje; entonces si se graba un video, la entrada de la cámara sería la luz proveniente del exterior y la salida una señal eléctrica continua. Esta señal continua es llevada a un Digitalizador de vídeo (grabador de cuadros) donde se digitaliza la imagen en un arreglo de $N \times M$ valores enteros guardados en un buffer finalmente se lleva a la computadora para su posterior procesamiento. (TRUCCO, y otros, 1998).

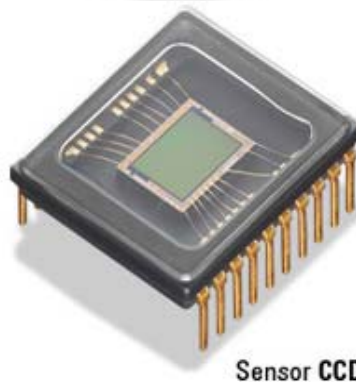


Figura 2.20 Sensor de cámara CCD

En la actualidad existen dos tipos de sensores semiconductores fotoresistivos los llamados CCD y los CMOS (LUKAC, 2009), ambas tecnologías fueron desarrolladas en los años 1960's y 1970's. En un sensor tipo CCD cada carga del pixel es transferida hacia un nodo de salida para ser convertida a voltaje y enviada como una salida análoga donde un circuito externo se encarga de la digitalización como se puede ver en el diagrama de bloques de la figura 2.21. A diferencia de los sensores tipo CCD en un sensor CMOS cada pixel tiene su propio convertidor de carga a voltaje y generalmente cada sensor incluye su amplificador, corrección de ruido y circuitos de digitalización entonces en este caso la salida de un sensor CMOS ya es digital.

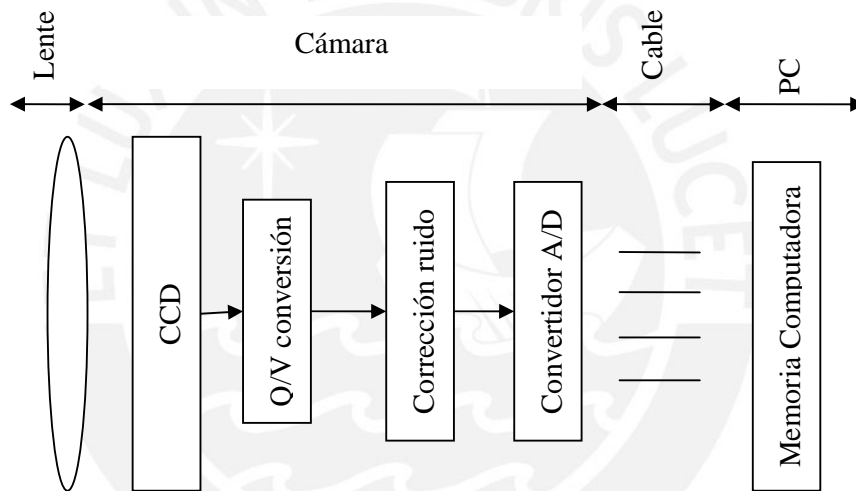


Figura 2.21 diagrama bloques cámara CCD

En una cámara a color los sensores fotosensibles son monocromáticos, entonces para lograr una imagen a color se han usado tres estrategias. La primera de ellas es capturar tres diferentes imágenes con filtros de color (LUKAC, 2009); sin embargo esta técnica solo es utilizada en laboratorios de medición ya que resulta muy impráctica para cualquier aplicación que involucre movimiento. La segunda estrategia es separar la luz que entra a la cámara en sus componentes usando un prisma. La tercera estrategia está referida al uso de arreglos de filtros en cámaras de un solo chip, el arreglo de filtros de color más usado es el filtro Bayer el cual se basa en la característica del ojo humano el cual es más sensitivo al color verde que a los demás colores (rojo y azul). (SONKA, y otros, 2008)

2.5.2. Iluminación.

La imagen obtenida por las cámaras difiere mucho dependiendo de la iluminación con la cual se obtuvo, ya sea con iluminación ambiental o iluminación artificial. Entonces la imagen adquirida es la suma de luces y sombras creadas por la incidencia o no de luz, tal y como se aprecia en la figura 2.22.

2.3.2.1. Iluminación Frontal

La cámara se sitúa mirando al objeto en la misma dirección que la luz (Bajo ángulo de incidencia) y recibe la luz reflejada por el objeto. Para esto se usan fuentes de luz puntuales, se usa para superficies planas; sin embargo tienen problemas de sombras y deslumbramientos.

2.3.2.2. Iluminación Lateral

La luz incide de forma lateral con un ángulo de incidencia. Se puede generar mediante iluminadores puntuales o mediante paneles, se usa para resaltar detalles laterales de los objetos o para mostrar aspectos de la estructura superficial. Tienen problemas de sombra.

2.3.2.3. Iluminación Difusa

Generada por diferentes puntos de luz en todas direcciones, se puede usar para todo tipo de objetos mates o brillantes. Presenta dificultad de ser usada en espacios reducidos.

2.3.2.4. Iluminación Coaxial

La iluminación pasa por el mismo eje del objetivo. La luz llega de manera perpendicular a la superficie del objeto creando fuertes contrastes; ello es debido al retorno total de los rayos provenientes de superficies lisas y de los rayos que las superficies inclinadas reflejan. (M. VALIENTE, y otros, 2010)

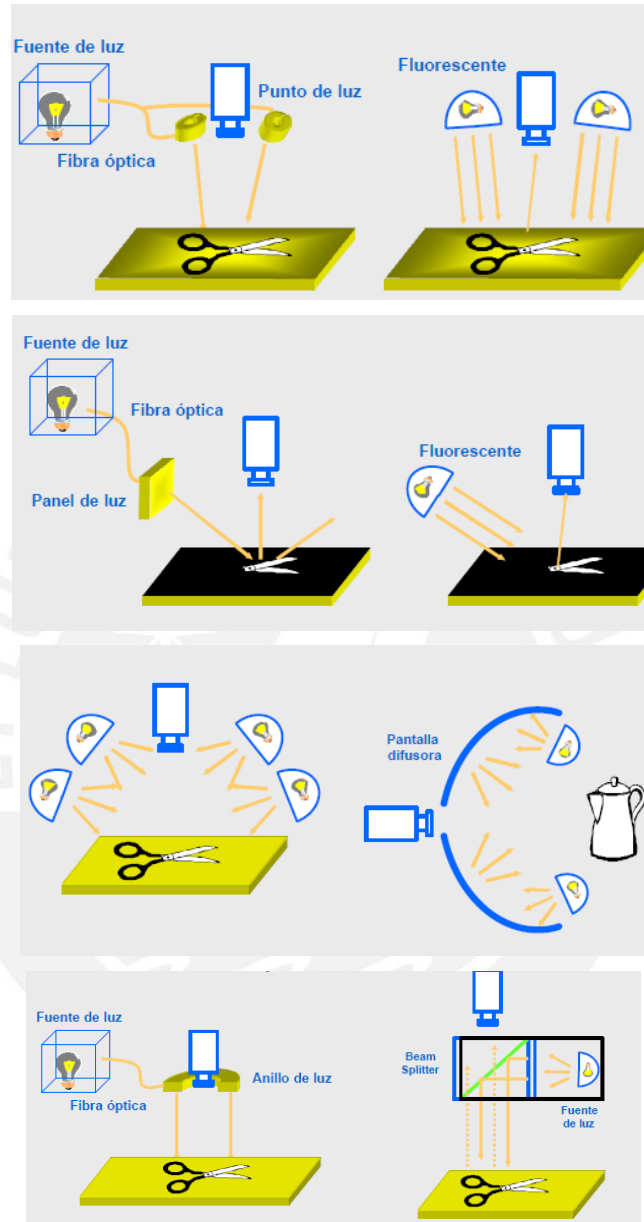


Figura 2.22 Iluminación Frontal, Lateral, Difusa y Coaxial

CAPÍTULO 3

VISIÓN ESTEREOSCÓPICA Y LOCALIZACIÓN DE OBJETOS

3.1 Visión estéreo.

La visión estéreo está referida a la habilidad de inferir información de una estructura determinada en el espacio a partir de 2 o más imágenes tomadas desde diferentes puntos de vista. (TRUCCO, y otros, 1998).

Como ejemplo en la figura 3.1, tenemos al punto P en el espacio y a sus respectivas proyecciones x^l y x^r en las cámaras izquierda y derecha respectivamente las cuales han sido obtenidas por la intersección del plano de las imágenes con los vectores P_l y P_r .

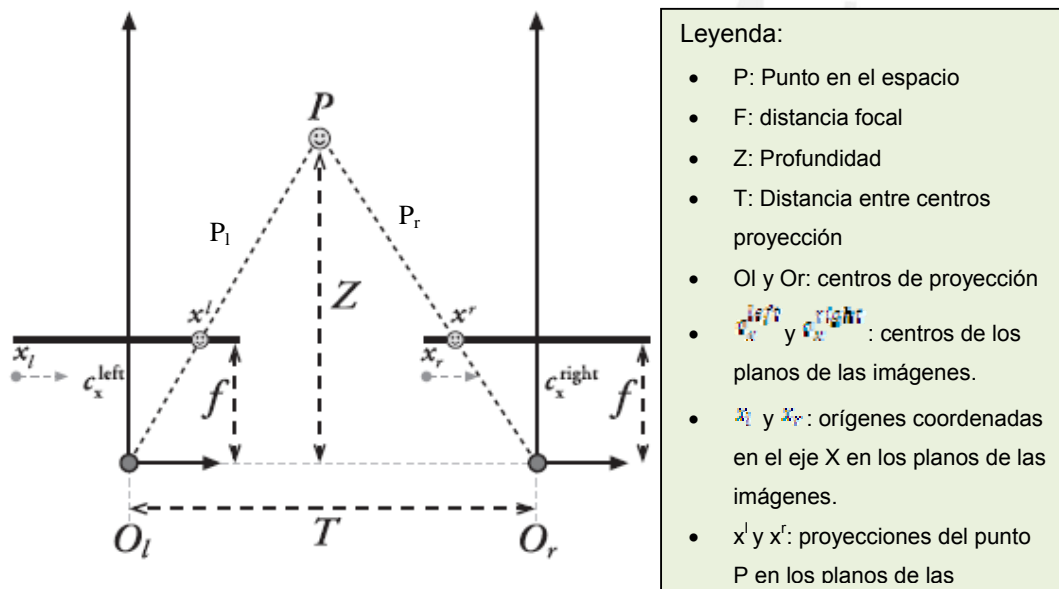


Figura 3.1 Vista superior de un sistema de visión estéreo

Por semejanza de triángulos (P,O_l,O_r) y (P,x^l,x^r) tenemos

$$\frac{T - (x^l - c_x^{left}) - (c_x^{right} - x^r)}{Z - f} = \frac{T}{Z}$$

(3.1)

En este punto las cámaras son tomadas como ideales entonces

$$c_x^{\text{left}} = c_x^{\text{right}}$$

(3.2)

Finalmente se tiene:

$$\frac{T - (x^l - x^r)}{Z - f} = \frac{T}{Z}$$

(3.3)

A la diferencia de los puntos de proyección $x^l - x^r$ se le denomina disparidad “d” (TRUCCO, y otros, 1998), por lo tanto despejando Z de la ecuación anterior tenemos:

$$Z = f \left(\frac{T}{d} \right)$$

(3.4)

Donde $d = x^l - x^r$ es denominada disparidad.

De la figura 3.2; donde se muestra el gráfico distancia versus disparidad, se deduce la primera conclusión importante, mientras más alejado se encuentre un objeto de las cámaras la disparidad tiende a disminuir, esto es debido a que mientras más alejado se encuentre de las cámaras, más cerca está del su vector principal. Este comportamiento se manifiesta en la realidad al acercar un objeto a la vista del observador generando el efecto de “visión doble”.

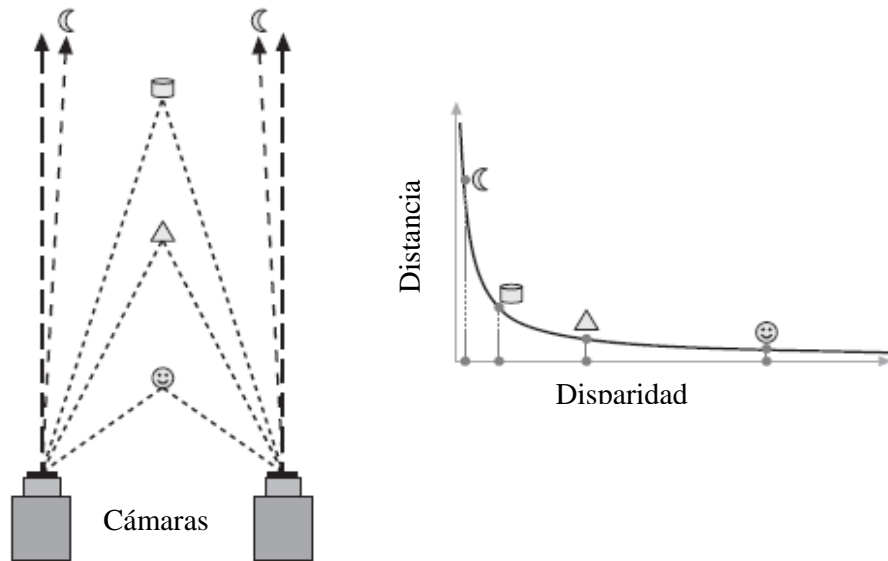


Figura 3.2 Disparidad versus distancia

3.2 Parámetros Intrínsecos y Extrínsecos

La relación entre el punto en el espacio P con coordenadas espaciales (X, Y, Z) y con las coordenadas (x,y) del punto p en el plano de la imagen se le denomina relación de transformación. Cuando se trabaja con este tipo de transformaciones entre diferentes planos proyectivos es mejor hacerlo con coordenadas homogéneas, en nuestro caso el plano de la imagen es nuestro plano de proyección con una restricción w . Esta relación entre el punto en el espacio y su proyección se representa en la ecuación 3.5

$$p = MP$$

(BRADSKI, y otros, 2008)

(3.5)

Donde:

- $p = \begin{bmatrix} x \\ y \\ w \end{bmatrix}$: Punto de proyección en el plano de la imagen con coordenadas (x, y, w)

- $M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$: Matriz de parámetros intrínsecos donde:

- (f_x, f_y, f_z) Son coordenadas de la distancia focal
- (c_x, c_y) Son coordenadas del centro de la imagen

- $P = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$: Punto en el espacio con coordenadas (X, Y, Z)

De la ecuación anterior, tenemos a la matriz M la cual contiene todos los parámetros intrínsecos de la cámara, distancia focal y centro de imagen, por este motivo se le conoce como matriz intrínseca.

Los parámetros extrínsecos describen la posición y orientación relativa de las dos cámaras, entonces tenemos en primer lugar la matriz de traslación (T) la cual nos indica el desplazamiento de una cámara respecto de la otra y la matriz de rotación (R) la cual nos dice la orientación relativa de las cámaras. Ambas variables se relacionan en la siguiente ecuación:

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

(TRUCCO, y otros, 1998)

(3.6)

Donde:

- P_r : Proyección del punto P en la cámara de la derecha
- P_l : Proyección del punto P en la cámara de la izquierda
- R : Matriz de Rotación del sistema

- $T = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$: Matriz de Traslación del sistema

- T_x, T_y, T_z Son las componentes del vector traslación

Debido a que las cámaras que vamos a usar no son ideales debemos introducir los llamados coeficientes de distorsión, los más importantes son los radiales y tangenciales.

El coeficiente de distorsión radial está referido a la distorsión generada por el lente circular de la cámara que genera una proyección errónea de la imagen en el plano como se muestra en la figura 3.3.

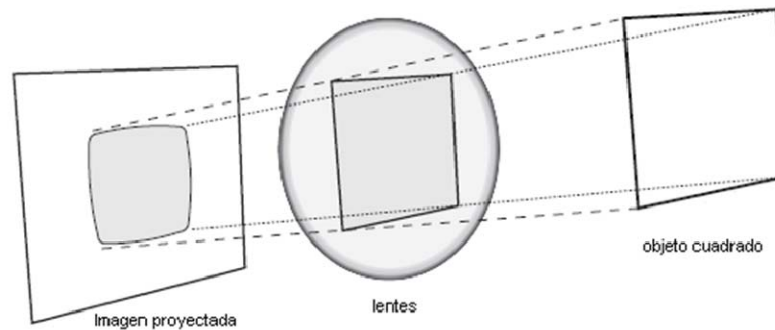


Figura 3.3 distorsión radial

Dado que mientras más alejada este la imagen proyectada mayor es la distorsión, esto se aproxima a una serie de Taylor que depende de la distancia radial como se ve en las ecuaciones.

$$\begin{aligned}
 X_{\text{correcto}} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\
 Y_{\text{correcto}} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)
 \end{aligned}$$

(BRADSKI, y otros, 2008)

(3.7)

Otro de los factores de distorsión está dado por la distorsión tangencial (figura 3.4) que es el resultado de defectos de manufactura; como resultado el sensor CMOS no está paralelo al lente

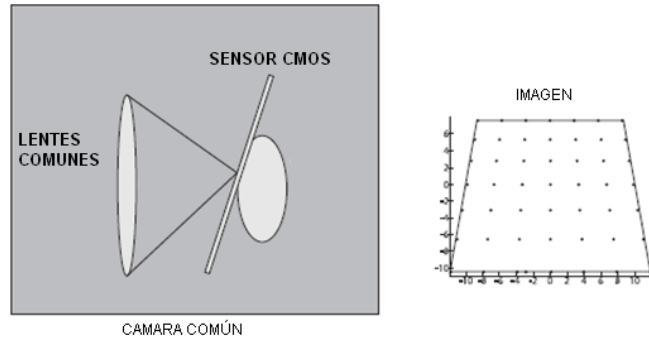


Figura 3.4 distorsión tangencial

Para corregir este problema se tiene la siguiente expresión casi similar a la anterior pero agrega dos parámetros p_1 y p_2 derivados del modelo "Plumb Bob" de Ingeniería Fotométrica (BRADSKI, y otros, 2008)

$$x_{correcto} = x + [2p_1y + p_2(r^2 + 2x^2)]$$

$$y_{correcto} = y + [2p_2x + p_1(r^2 + 2y^2)]$$

(BRADSKI, y otros, 2008)

(3.8)

3.3 Obtención de par estéreo

El par estéreo consiste en el par de imágenes obtenida del sistema de visión estéreo, es entonces el punto de partida del sistema puesto que, si se quiere tener un buen modelamiento del sistema entonces se hace necesario una buena captura y puesta a punto de las cámaras, de forma independiente en sus parámetros de enfoque, contraste, brillo y de forma conjunta en distancia entre cámaras e igual nivel de captura. Para lograr lo anterior es recomendable tener en cuenta:

- Las cámaras deben tener la misma resolución de captura
- Deben ser de preferencia del mismo tipo de fabricación para que la distancia focal sea la misma así como el punto principal.
- Necesidad de crear un soporte para garantizar que la posición de las cámaras sea siempre la misma una respecto de la otra.
- Nivelar el nivel de brillo y contraste para una mejor captura.

3.3.1. Obtención del cuadro de la cámara

Uno de los primeros pasos que debemos cumplir para empezar a realizar el programa es la captura de un cuadro de la cámara para su análisis, esto se logra gracias a una serie de comandos de la librería OpenCV.

En primer lugar debemos acceder a la cámara, para esto se usa la estructura `CvCapture` la cual nos brinda una interface con la cámara USB, luego como se usan dos cámaras deberemos referenciarlas para esto usamos la función `cvCaptureFromCAM(i)`; y finalmente para lograr capturar el cuadro debemos usar la función `cvQueryFrame()` la cual nos devuelve el cuadro de la cámara en el formato predeterminado ya sea `bpm` o `jpg`.

Un ejemplo de código para la referencia y captura de un cuadro de una cámara sería:

```
CvCapture* capture, *capture1 = 0;  
capture = cvCaptureFromCAM(i);  
capture1 = cvCaptureFromCAM(i+1);
```

Otro punto muy importante es la velocidad de captura, esto dependerá del tipo de cámara y de la velocidad de procesamiento del programa desarrollado. (BRADSKI, y otros, 2008)

3.3.2. Manejo de los parámetros de la cámara.

Para un mejor manejo de las imágenes, las cámaras USB tienen muchos parámetros configurables como son el brillo, contraste, gama etc. Todo esto con el único objetivo de lograr una mejor nitidez en la captura de las imágenes, esto ayudará de gran manera al momento de calibrar el sistema con ayuda de un patrón que no es más que un tablero de ajedrez.

OpenCV, también nos brinda un soporte en este aspecto pudiendo nosotros modificar diversos parámetros adicionales a los anteriormente mencionados como son la velocidad de cuadro, resolución de cámara etc.

Para la calibración del sistema, así como la rectificación de las imágenes nos es más fácil trabajar con imágenes en formato escala de grises debido a que, no presenta mucha dificultad a diferencia del formato RGB el cual cuenta con 3 capas cada imagen. Muchas de las cámaras USB actuales tienen como formato predeterminado el RGB o formato color, entonces para esto OpenCV nos brinda una herramienta llamada `cvCvtColor` la cual es una función capaz de convertir una imagen en otro formato incluido el que nosotros necesitamos.

3.4 Calibración del sistema estéreo

La calibración de cualquier sistema es la etapa más importante, debido a que, mediante esta etapa o procedimiento se va a obtener los parámetros de dicho sistema. En el caso de un sistema estéreo lo que se va a obtener son los parámetros de las cámaras así como la información del entorno en el cual se va a desarrollar el robot móvil.

Para un sistema estéreo su calibración pasa por encontrar la relación geometría entre las dos cámaras, esto depende de encontrar tanto la matriz de rotación como la matriz de traslación. Con esta información podremos relacionar ambas cámaras en un mismo entorno, tomando siempre como referencia la imagen de la izquierda, la matriz de rotación se interpreta como la rotación de la cámara derecha con respecto a la izquierda y la matriz de traslación está dada por el desplazamiento de la cámara derecha hacia la izquierda como lo ilustra la figura 3.5. (BRADSKI, y otros, 2008)

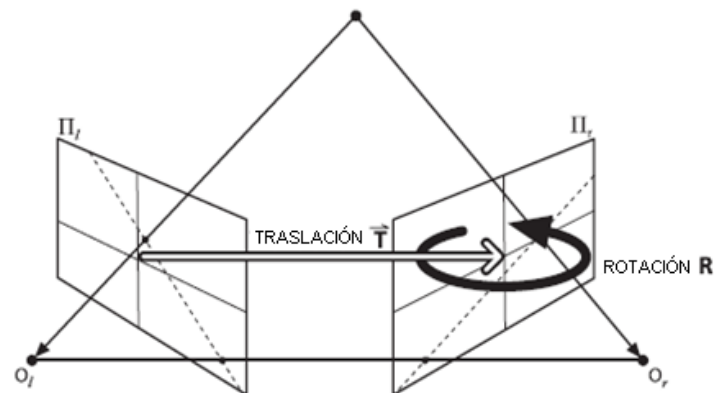


Figura 3.5 Matriz De traslación y rotación

El procedimiento por el cual vamos a llegar a la calibración de nuestro sistema estéreo será mediante el reconocimiento de patrones. Para esto se tomarán diferentes muestras de pares estéreos con la imagen patrón (tablero de ajedrez) de la cual recogeremos la información de la ubicación de las esquinas y la ingresaremos en la función `cvStereoCalibrate` la cual tiene la siguiente estructura:


```
bool cvStereoCalibrate(const CvMat* objectPoints, const CvMat* imagePoints1,  
const CvMat* imagePoints2, const CvMat* npoints, CvMat* cameraMatrix1, CvMat* distCoeffs1,  
CvMat* cameraMatrix2, CvMat* distCoeffs2, CvSize imageSize, CvMat* R, CvMat* T, CvMat*  
E, CvMat* F, )
```

El primer dato de entrada *objectPoints* es una matriz de $N \times 3$ en donde se encuentran las coordenadas físicas del tablero donde la coordenada z es igual a 0 debido a que se toma como referencia el plano del tablero y donde N es el número de esquinas que contiene nuestro tablero multiplicado por el número de cuadros tomados para la calibración. Los datos de entrada *imagePoints1* y *2* son las coordenadas de las esquinas del tablero tanto de la cámara1 como de la cámara2 y por último *npoints* es el número de puntos encontrados por cada cuadro.

Esta función nos da como parámetros de salida la distancia focal y centro de cámara contenidas en las matrices *cameraMatrix1* y *cameraMatrix2*, además de los coeficientes de distorsión tanto radial como tangencial contenidos en la matriz *distCoeffs*. Y por último y más importante las matrices de rotación, translación, esencial y fundamental del sistema. (BRADSKI, y otros, 2008)

3.5 Geometría epipolar

La geometría epipolar busca encontrar una relación geométrica entre los puntos de una imagen y los puntos de su similar estéreo. Para esto se basa en el modelo geométrico de las cámaras en donde el plano formado por el punto en el espacio P y los centros de las imágenes O_1 y O_2 denominado plano epipolar interseca a los planos de las imágenes en dos rectas denominadas líneas epipolares, logrando así una restricción en nuestro rango de búsqueda de los puntos correspondientes al de solo una recta. El modelo geométrico se puede entender mejor en figura 3.6.

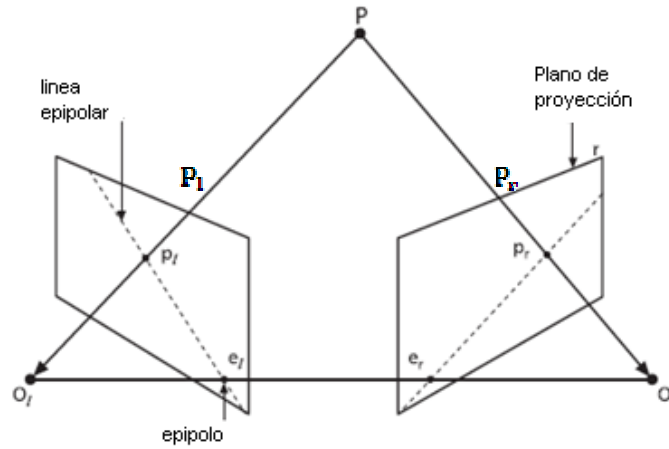


Figura 3.6 Líneas epipolares

Para poder obtener nuestras líneas epipolares, las cuales resultan de la intersección entre el plano epipolar y el plano de proyección de las imágenes, tenemos que obtener las matrices fundamental y esencial de nuestro sistema estéreo.

3.51. Matriz esencial y fundamental.

La matriz E se denomina la matriz esencial y establece una unión natural entre la restricción epipolar y los parámetros extrínsecos del sistema estéreo, a través de la matriz esencial se establece una relación entre los puntos de una imagen y las rectas epipolares en la otra imagen. (CAPILLA, 2002).

De forma matemática (por propiedad de vectores perpendiculares) tenemos que:

$$(\mathbf{P}_l - \mathbf{T})^T (\mathbf{T} \times \mathbf{P}_r) = 0$$

(BRADSKI, y otros, 2008)

(3.9)

Despejando \mathbf{P}_l de la ecuación 3.6 y reemplazando en 3.9 tenemos

$$(\mathbf{R}^T \mathbf{P}_r)^T (\mathbf{T} \times \mathbf{P}_l) = 0$$

(3.10)

Reescribiendo el producto cruz $(\mathbf{T} \times \mathbf{P}_1)$ tenemos:

$$\mathbf{T} \times \mathbf{P}_1 = \mathbf{S} \mathbf{P}_1 \quad (3.11)$$

Donde S está dada por la siguiente expresión.

$$\mathbf{S} = \begin{bmatrix} 0 & -T_z & -T_y \\ T_z & 0 & -T_x \\ -T_y & -T_x & 0 \end{bmatrix}$$

En Donde:

T_x, T_y, T_z : Coordenadas del vector traslación T.

Reemplazando la equivalencia 3.11 en la ecuación 3.10 se obtiene

$$(\mathbf{P}_R)^T \mathbf{R} \mathbf{S} \mathbf{P}_1 = \mathbf{0} \quad (3.12)$$

Donde la matriz Esencial “E” queda definida como el producto:

$$\mathbf{E} = \mathbf{R} \mathbf{S} \quad (\text{BRADSKI, y otros, 2008}) \quad (3.13)$$

Como se observa la matriz esencial contiene la información que relaciona geoméricamente una cámara con la otra; sin embargo para nuestro caso nos interesa tener esa información en pixeles para poder relacionar ambas cámaras en términos de estos. Entonces de las ecuaciones 3.12 y 3.13 se tiene:

$$\mathbf{P}_R^T \mathbf{E} \mathbf{P}_1 = \mathbf{0} \quad (3.14)$$

En la figura 3.7 se observa la relación de proyección de una cámara. De donde se deduce:

$$\frac{\mathbf{p}}{\mathbf{f}} = \frac{\mathbf{P}}{\mathbf{Z}} \quad (3.15)$$

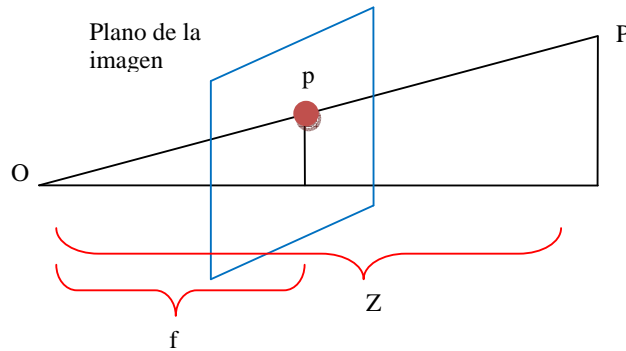


Figura 3.7 Relación de proyección de una cámara

Entonces de las ecuaciones 3.14 y 3.15 se llega a la siguiente expresión:

$$\mathbf{p}_r^T \mathbf{E} \mathbf{p}_l = 0 \tag{3.16}$$

Ahora se introduce en la relación los parámetros intrínsecos de la cámara. Recordemos que $\mathbf{p} = \mathbf{M}\mathbf{P}$ donde \mathbf{p} es la proyección del punto \mathbf{P} en la imagen; sin embargo la matriz \mathbf{M} también expresa la relación entre los vectores \mathbf{p}_l y \mathbf{p}_r (en términos generales \mathbf{p}) con sus respectivas proyecciones en unidades de pixeles de tal forma que:

$$\mathbf{p} = \mathbf{M}^{-1} \mathbf{q} \tag{3.17}$$

(BRADSKI, y otros, 2008)

Donde:

\mathbf{q} : Representa al vector \mathbf{p} pero en unidades de pixeles

Reemplazando 3.17 en 3.16:

$$\mathbf{q}_r^T (\mathbf{M}_r^{-1})^T \mathbf{E} \mathbf{M}_l^{-1} \mathbf{q}_l = 0 \tag{3.18}$$

Entonces se define la matriz fundamental \mathbf{F} como:

$$\mathbf{F} = (\mathbf{M}_r^{-1})^T \mathbf{E} \mathbf{M}_l^{-1} \tag{3.19}$$

De tal forma que:

$$q_1^T F q_2 = 0 . \quad (3.20)$$

Como se observa la matriz fundamental F nos relaciona los puntos proyección del un punto en el espacio en una cámara con los puntos de proyección del mismo punto en la otra cámara.

Para encontrar la matriz fundamental existen varios algoritmos como son:

Value of method	Number of points	Algorithm
CV_FM_7POINT	$N = 7$	7-point algorithm
CV_FM_8POINT	$N \geq 8$	8-point algorithm
CV_FM_RANSAC	$N \geq 8$	RANSAC algorithm
CV_FM_LMEDS	$N \geq 8$	LMedS algorithm

Para nuestro caso usaremos el algoritmo RANSAC² debido a que es un algoritmo robusto y además permite reconocer y remover los llamados “outliers” que son variaciones extremas en una muestra o conjunto. Para poder obtener la matriz fundamental se usará la función cvFindFundamentalMat, la cual recibe como datos de entrada los puntos encontrados por el reconocimiento del patrón y el tipo de algoritmo a usar en este caso será RANSAC y como dato de salida nos brinda la matriz fundamental.

Una vez obtenida la matriz fundamental podremos dibujar las líneas epipolares. En el caso de usar la función cvStereoCalibrate vista anteriormente ya tiene implementada la función cvFindFundamentalMat ahorrando el trabajo llamarla nuevamente.

² Es una abreviatura para el “consenso de la muestra escogida al azar”. Es método iterativo para estimar parámetros de un modelo matemático de un sistema de los datos observados que contiene afloramientos. El algoritmo primero fue publicado por Fischler y Bolles en 1981.

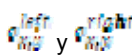
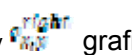
3.6 Rectificación

Por diversos factores ya sea de fabricación o de posición los planos de las imágenes no son coplanares, entonces debemos re proyectar esas imágenes para que las filas de las imágenes estén perfectamente alineadas.

Existen diferentes algoritmos para este procedimiento; sin embargo en nuestra librería solo están implementados dos de esos algoritmos de rectificación: el algoritmo de Hartley y el algoritmo de Bouguet. Se usará el segundo de ellos puesto que toma en consideración no solo la matriz fundamental sino también la matriz de rotación.

3.6.1. Algoritmo de bouguet

El algoritmo desarrollado por Jean-Yves Bouguet trata que los cambios en las imágenes re-proyectadas no sean demasiados con respecto a las imágenes originales. Este algoritmo nos devuelve como resultados en primer lugar la matriz de rotación del sistema rectificado y las matrices de re-proyección con las cuales se aplicarán las transformaciones geométricas al par estéreo.

El algoritmo realiza una rotación de ambas imágenes, la mitad de la rotación de la matriz R obtenida por la calibración estéreo; con esto lo que se obtiene es que ambas imágenes estén rotadas y puestas en un plano coplanar. Para tener una alineación horizontal de tal manera que las líneas epipolares sean horizontales se hace una segunda rotación teniendo como eje principal el epipolo y como referencia los centros de las imágenes  y  graficados en la figura 3.1.

La biblioteca opencv lidia con esto en la función `cvStereoRectify` en donde como parámetros de entrada están la matriz de parámetros intrínsecos de ambas cámaras más los coeficientes de distorsión y las matrices de rotación y translación (R T) obtenidas de la calibración estéreo, la declaración de esta función es como sigue a continuación:

```
void cvStereoRectify( const CvMat* cameraMatrix1, const CvMat* cameraMatrix2, const CvMat*  
distCoeffs1, const CvMat* distCoeffs2, CvSize imageSize, const CvMat* R, const CvMat* T,
```

```
CvMat* RI, CvMat* Rr, CvMat* PI, CvMat* Pr, CvMat* Q=0, int  
flags=CV_CALIB_ZERO_DISPARIITY);
```

Como parámetros de salida esta función nos entrega la nueva matriz de rotación y las matrices de parámetros intrínsecos con los nuevos valores, contenidos en las matrices PI y Pr ya sean los parámetros de la cámara izquierda o derecha respectivamente.

3.6.2. Rectificación de las imágenes del par estéreo.

El proceso de rectificación del sistema, consiste en llevar a un plano paralelo los pares estéreo además de eliminar o disminuir los efectos de distorsión tanto radial como tangencial; con el objetivo de reducir los cálculos matemáticos en la obtención de las coordenadas.

Una vez que tenemos las matrices M_{rec} . (Matriz 3x3 de los parámetros rectificadas cuyos elementos están en las 3 columnas de las matrices PI y Pr .) y R_r (matriz de rotación rectificadas), podemos entonces obtener el mapa de rectificación para esto se usa la función `cvInitUndistortRectifyMap` la cual nos devuelve como resultado un arreglo de coordenadas en donde nos indica las nuevas coordenadas para los píxeles de la imagen original. Una vez obtenido el mapa de coordenadas se procede a rectificar las imágenes con la función `cvRemap` con la cual se obtiene finalmente lo que se busca que son las imágenes rectificadas.

El proceso de rectificación entonces está dado por varias etapas como se muestra en la figura 3.8. Etapa (a) de captura del par estéreo, (b) se eliminan los coeficientes de distorsión tanto tangencial como radial, (c) se lleva ambas imágenes a sus proyecciones en el plano horizontal y finalmente en (d) se recupera la información deseada.

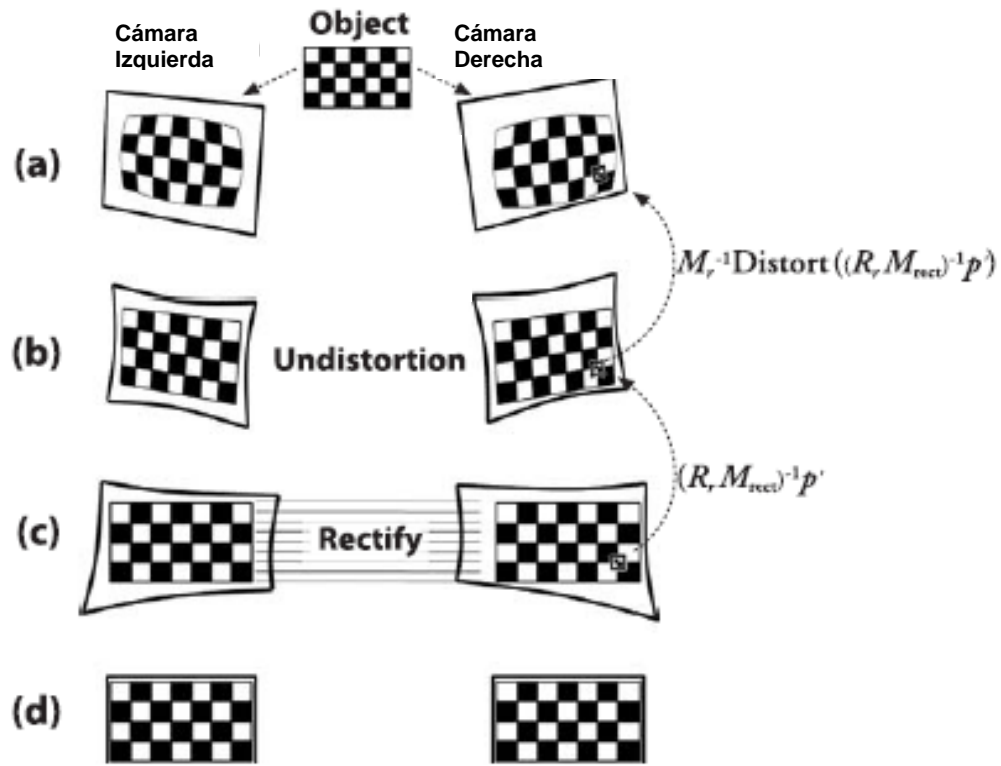


Figura 3.8 Proceso de rectificación de una imagen:

3.7 Estimación de Coordenadas de un Punto

Como hemos visto anteriormente para estimar la distancia de un punto partimos de la siguiente ecuación:

$$Z = f \left(\frac{T}{d} \right) \tag{3.21}$$

En donde d es la disparidad; es decir, la diferencia de las coordenadas de los puntos correspondientes en ambas imágenes; sin embargo esta ecuación es solo válida cuando los centros de las imágenes son los mismos que el centro del lente lo cual cambian las referencias de las imágenes, entonces agregamos un factor de corrección en la definición de disparidad y como resultado se tiene la siguiente expresión.

$$d = x^l - x^r - (c^l - c^r)$$

(BRADSKI, y otros, 2008)

$$(3.22)$$

Teniendo como ecuación final, reemplazando 3.14 en 3.4:

$$Z = f \left(\frac{T}{x^l - x^r - (c^l - c^r)} \right)$$

$$(3.23)$$

Las variables f, T, c^l y c^r serán tomadas de la matriz de parámetros intrínsecos, M , de la cámara izquierda y derecha según corresponda.

Para encontrar las coordenadas X e Y de nuestro sistema tomaremos para nuestro caso como punto de origen el centro de la cámara derecha (puede ser la izquierda es indiferente) como se muestra en la Figura 3.9.

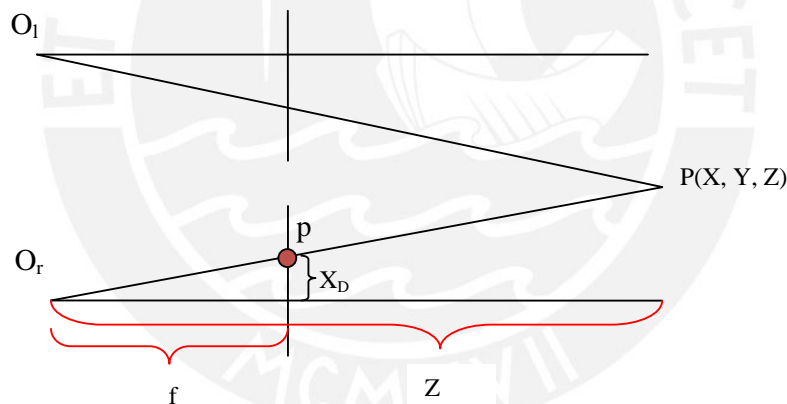


Figura 3.9 vista superior de sistema estéreo

De la figura 3.9 podemos deducir que:

$$X = Z \left(\frac{x_D}{f} \right)$$

$$(3.24)$$

Donde:

x_D : Distancia en pixeles del punto p en la imagen con respecto al centro de la misma en el eje x.

f : distancia focal también en pixeles³.

De la misma forma para la coordenada Y tenemos la ecuación.

$$Y = z \left(\frac{y_d}{f} \right)$$

(3.25)

Donde:

y_d : Distancia en pixeles del punto p en la imagen con respecto al centro de la misma en el eje Y.



³ La distancia focal en unidades de pixeles ya está dada como tal por las funciones de calibración y rectificación.

CAPÍTULO 4

PRUEBAS RESULTADOS. ANALISIS ECONÓMICO

4.1 Descripción del proyecto

El presente trabajo busca desarrollar un algoritmo de visión estereoscópica capaz de obtener las coordenadas espaciales de un objeto, esto quiere decir las componentes X, Y, Z de la ubicación del mismo, y luego transmitir estas coordenadas a un robot móvil para su posterior utilización.

4.2 Requerimientos:

Para el desarrollo de este sistema se tienen que considerar varios factores como son: el sistema operativo, el tipo de cámaras, las características de la comunicación etc.

4.2.1. Sistema operativo

El sistema operativo a elegir tiene que tener diferentes características de funcionamiento y manejo de recursos como son:

- Brindar fácil acceso al hardware, en nuestro caso a las cámaras
- No debe implicar un costo elevado de compra o de derechos.
- Tener soporte en internet sobre diferentes problemas que se puedan presentar
- Poder manejar enlaces de red mediante protocolo TCP/IP

Como resultado del análisis, el sistema operativo escogido es una distribución de Linux (Ubuntu 9.04) capaz de manejar y brindar acceso a los periféricos como son las cámaras web, además de brindar un mejor manejo de recursos que otros sistemas operativos.

4.2.2. Cámaras:

- Se necesitan dos cámaras para poder capturar el par estéreo.
- Que sean del mismo fabricante para poder tener los mismos parámetros intrínsecos en ambas cámaras y así evitar problemas de calibración y programación.

- Como la programación se va a realizar en un sistema operativo libre (Linux) las cámaras deben ser compatibles con este sistema es decir se tiene que poder acceder a estas mediante el sistema operativo.
- La resolución de las cámaras debe ser buena para facilitar la tarea de segmentación.

Por estas razones y sobre todo por el tema de compatibilidad con el sistema operativo se decidió probar el sistema con dos tipos de cámaras:

- camaras creative® modelo Cam Notebook Pro (VF0250)
- imagingsource® modelo DFK 31BU03.H.

4.2.3. Placa madre

La placa madre donde se van a cargar los programas y conectar las cámaras debe contar con las siguientes características:

- Poder trabajar a velocidades de procesamiento capaces de poder soportar el envío de datos de dos cámaras de forma casi simultánea; es decir poder procesar datos a una velocidad de procesamiento solo para el uso de las cámaras superior a 47.18 mbps a esto tenemos que sumarle el procesamiento de imágenes, el enlace de comunicación, etc.
- El tamaño también es importante debido que la placa madre no debe ser demasiado grande por razones de practicidad y de montaje.
- Debe tener tarjeta de video integrada para evitar la compra de componentes extras; así como incrementar el tamaño de la placa.
- Debido a que la aplicación es dedicada a un solo propósito, el uso de la memoria RAM será de 1 gb.

Por las razones expuestas anteriormente se vio conveniente elegir una placa que se encuentra en el mercado local y cumple todos los requerimientos anteriormente

mostrados, dicha placa es la gigabyte modelo GA-EG31M-S2. La cual es de tamaño pequeño y capaz de manejar puertos USB, enlace de red, etc.

4.2.4. Comunicación

Para la transmisión de datos existen diferentes modos de transmisión como son: transmisión serial, paralela etc.; sin embargo en este caso se van a transmitir imágenes, para esto se necesita un medio de transmisión capaz de manejar velocidades superiores a los 50mbps. El medio que nos permite cumplir con este requerimiento es el enlace de red mediante el protocolo TCP/IP, debido a que puede manejar velocidades superiores a la indicada.

Para realizar este tipo de enlace entre dos computadoras (sistema de visión estéreo y robot móvil) se hace mediante un cable cruzado cuya disposición de conexiones de puede observar en la figura 4.1

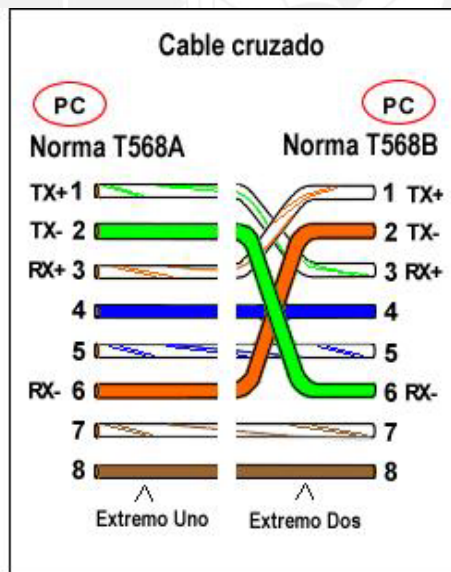


Figura 4.1 Conexiones cable cruzado para un enlace de red entre dos computadoras

4.3 Diseño de la Interfaz de usuario

El diseño de la interfaz de usuario está basado en la librería multiplataforma GTK. La cual es usada principalmente en los entornos GNOME, XFCE y ROX aunque también se puede usar en el escritorio de Windows, MacOS y otros.

Lo primero que debe hacer el programa es poder mostrar las imágenes de las cámaras, para esto se crearán 2 ventanas de entorno gráfico en las cuales se mostrarán los cuadros capturados. Figura 4.2

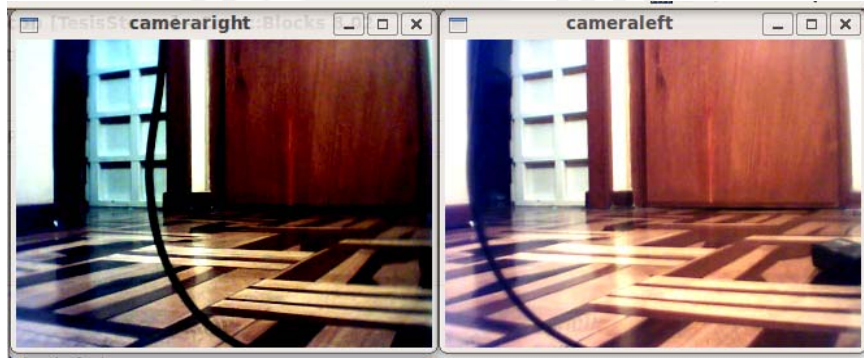


Figura 4.2 Captura de imágenes y muestra de imágenes de ambas cámaras en la interfaz de usuario

Ahora debemos tener en cuenta cuales son los parámetros que el usuario va manejar y los resultados que se deben mostrar. En primer lugar los parámetros que se va manejar son el brillo y contraste de ambas cámaras tanto de la izquierda como la derecha por separado.

Los valores de brillo y contraste de la imagen van a variar en un rango de 0 y 200 para ambas imágenes; siendo el valor 100 el valor por defecto. La forma de ingresar estos datos será mediante una barra de desplazamiento, debido a que es uno de los medios más usados para el ingreso de datos y además de permitir una visualización del rango del parámetro. Figura 4.3

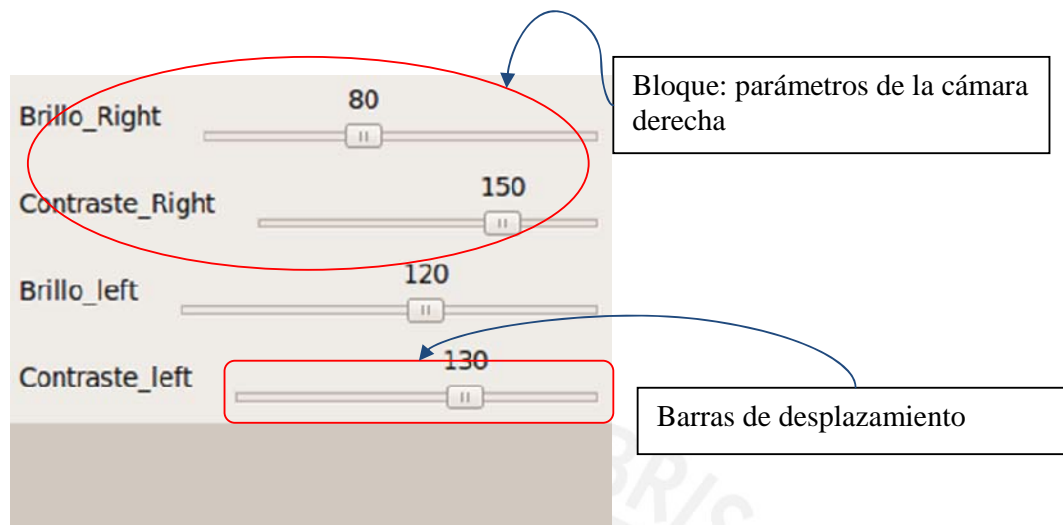


Figura 4.3 Interfaz de usuario gráfica de variación de brillo y contraste de las cámaras

Una vez que los parámetros de brillo y contraste se han establecido, llega la etapa de segmentación por color en donde tenemos que variar los límites de cada color para lograr segmentar el objeto deseado. Como las imágenes a procesar son de tipo entero es decir cada pixel tiene 3 valores (RGB) y cada uno de ellos entre el rango de 0 a 255, entonces el rango de entrada también debe variar en ese rango. Figura 4.4

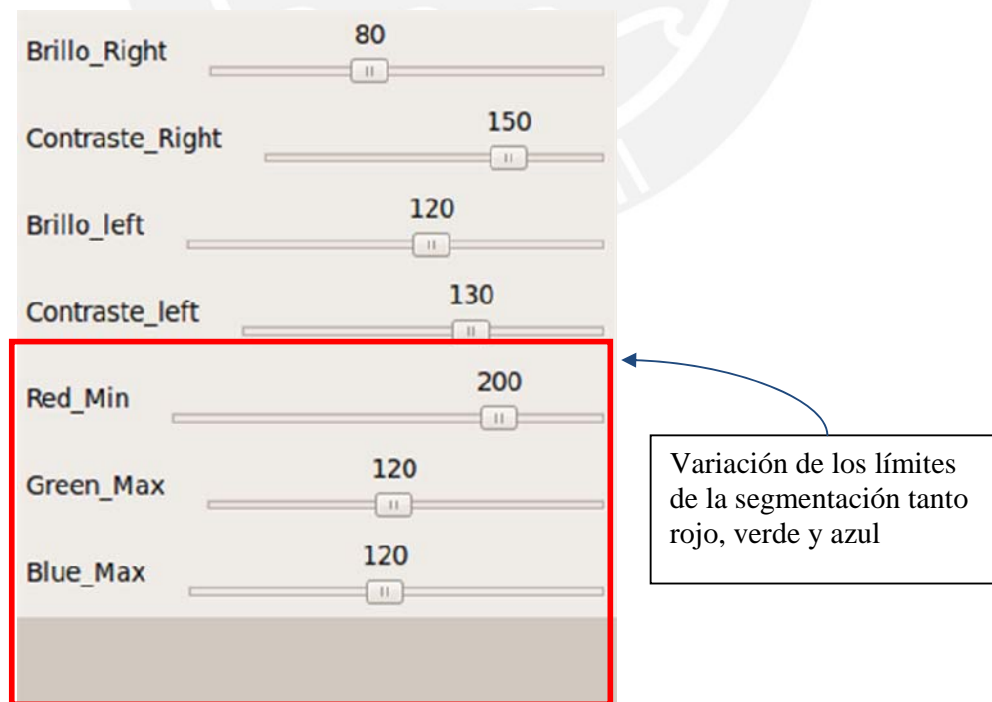


Figura 4.4 Interfaz de segmentación

4.4 Calibración del sistema estéreo

Para calibrar el sistema estéreo, primero se tiene que obtener los parámetros extrínsecos e intrínsecos del mismo. Este trabajo se va centrar en dos sistemas particulares, el primero de ellos basado en las cámaras creative® modelo Cam Notebook Pro (VF0250) y el segundo basado en las cámaras de la compañía imagingsource® modelo DFK 31BU03.H.

El algoritmo implementado sigue la misma secuencia que el diagrama de flujo de la figura 4.5. Donde podemos apreciar los diferentes procesos y funciones estudiados en el capítulo 3.

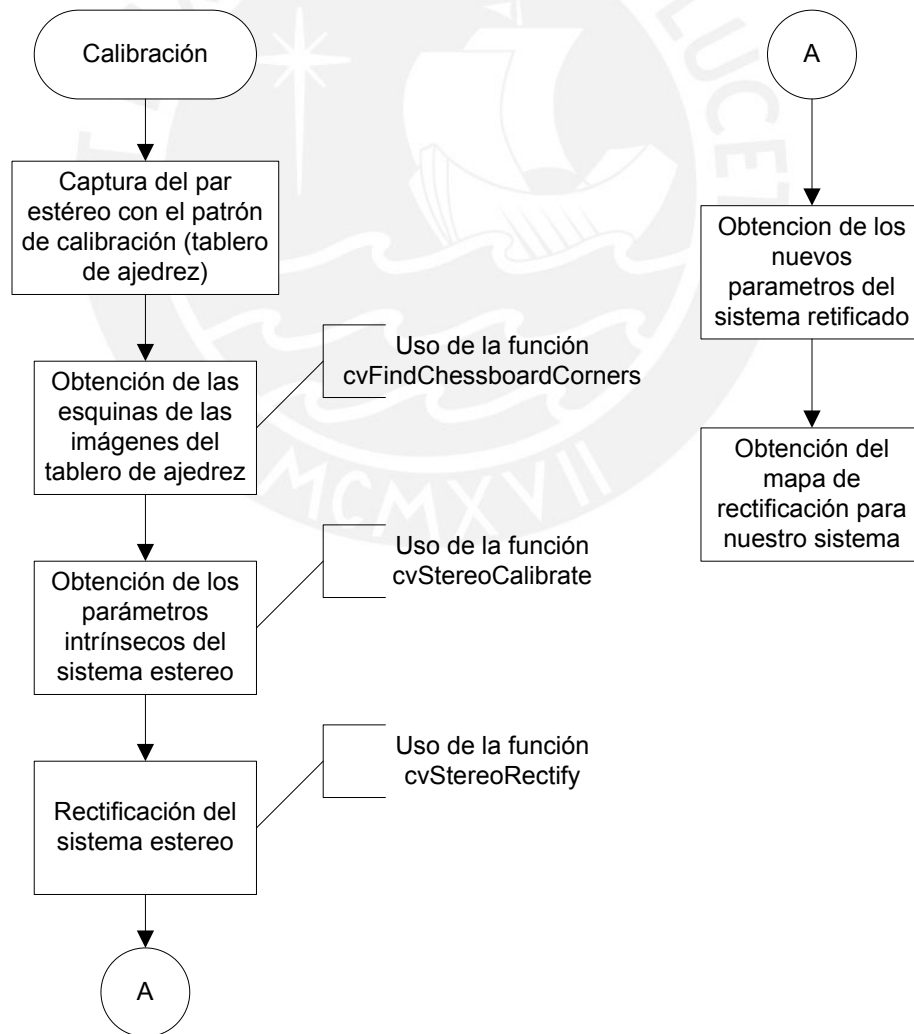


Figura 4.5 Diagrama de flujo del programa de calibración

Como resultado de la calibración, los parámetros intrínsecos (matriz M ecuación 3.5) en unidades de pixeles obtenidos por el programa se muestran en la tabla 4.1

Cam Notebook Pro (VF0250)	
Matriz 1	$M1 = \begin{bmatrix} 421.793142 & 0.0 & 132.118813 \\ 0.0 & 421.793142 & 119.060923 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$
Matriz 2	$M2 = \begin{bmatrix} 421.793142 & 0.0 & 132.902382 \\ 0.0 & 421.793142 & 120.898814 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$

DFK 31BU03.H	
Matriz 1	$M1 = \begin{bmatrix} 2370.663214 & 0.0 & 263.380473 \\ 0.0 & 2370.663214 & 103.196466 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$
Matriz 2	$M2 = \begin{bmatrix} 2370.663214 & 0.0 & 263.364336 \\ 0.0 & 2370.663214 & 103.196466 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$

Tabla 4.1 Resultados del modelamiento de las cámaras

Donde; como primera observación las distancias focales son la mismas, esto es importante porque permite una mejor calibración del sistema estéreo. La diferencia sin embargo radica en el centro de la imagen. Como se aprecia no necesariamente el centro de la imagen coincide con el centro del eje focal. Otro de los parámetros intrínsecos de las cámaras son las matrices de distorsión de los sistemas, en nuestro caso las matrices son los mostrados en la tabla 4.2.

Cam Notebook Pro (VF0250) DFK 31BU03.H	
$d1 = \begin{bmatrix} 0.821678 \\ 1.080914 \\ 0.0 \\ 0.0 \\ 4.796282 \end{bmatrix}$	$d1 = \begin{bmatrix} 0.240294 \\ 17.746269 \\ 0.0 \\ 0.0 \\ 0.026848 \end{bmatrix}$
$d2 = \begin{bmatrix} 0.439913 \\ 1.433240 \\ 0.0 \\ 0.0 \\ 4.203315 \end{bmatrix}$	$d2 = \begin{bmatrix} -0.326490 \\ 18.344711 \\ 0.0 \\ 0.0 \\ 639.796385 \end{bmatrix}$

Tabla 4.2 Matrices de distorsión

Los parámetros extrínsecos de los sistemas estéreo que son la matriz de rotación y traslación son los mostrados en la tabla 4.3.

Sistema estéreo Creative			Sistema estéreo Imaging Source		
$T = \begin{bmatrix} -3.969763 \\ 0.028066 \\ -0.069999 \end{bmatrix}$			$T = \begin{bmatrix} -4.683302 \\ -0.068748 \\ -1.733343 \end{bmatrix}$		
$R = \begin{bmatrix} 0.999923 & 0.009978 & 0.007319 \\ -0.009928 & 0.999927 & -0.00691 \\ -0.007388 & 0.006841 & 0.999949 \end{bmatrix}$			$R = \begin{bmatrix} 0.999983 & -0.003426 & -0.00044 \\ 0.003427 & 0.999979 & 0.003623 \\ 0.000423 & -0.003626 & 0.999993 \end{bmatrix}$		

Tabla 4.3 Parámetros extrínsecos del sistema estéreo

4.5 Rectificación

Una vez obtenido el modelo del sistema se procede a rectificar las imágenes; sin embargo como la posición de las imágenes en el plano espacial cambian debido al proceso de rectificación, entonces se debe recalcular los parámetros intrínsecos y extrínsecos del sistema rectificado. Para esto se va usar la función cvStereoRectify detallada anteriormente en el capítulo 3.

Como se puede apreciar en la tabla 4.4. Los sistemas tienen una nueva distancia focal y además las coordenadas del punto central tienen la misma componente en el eje y.

Sistema estéreo Creative						
Camara1	M_{1rect}	$\begin{bmatrix} 380.076974 & 0.0 & 137.826988 \\ 0.0 & 380.076974 & 120.036796 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$				
		Camara2	M_{2rect}	$\begin{bmatrix} 380.076974 & 0.0 & 145.262936 \\ 0.0 & 380.076974 & 120.036796 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$		

Sistema estéreo Imaging Source						
Camara1	M_{1rect}	$\begin{bmatrix} 2564.892251 & 0.0 & -677.857910 \\ 0.0 & 2675.765522 & 108.146100 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$				
		Camara2	M_{2rect}	$\begin{bmatrix} 2564.892251 & 0.0 & -678.816040 \\ 0.0 & 2675.765522 & 108.146100 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$		

Parámetros extrínsecos					
Sistema estéreo Creative		$R_{rect} = \begin{bmatrix} 0.999804 & 0.005327 & 0.019075 \\ -0.005292 & 0.999960 & -0.003452 \\ -0.019098 & 0.003350 & 0.999812 \end{bmatrix}$			
Sistema estéreo Imaging Source		$R_{rect} = \begin{bmatrix} 0.938 & 0.007413 & 0.346997 \\ -0.007874 & 0.999969 & 0.000080 \\ -0.346547 & -0.002654 & 0.938029 \end{bmatrix}$			

Tabla 4.4 Resultados del proceso de rectificación

Con el nuevo sistema se obtiene el **mapa de rectificación** y así se procede a rectificar las imágenes. Para probar si la rectificación se llevó a cabo de forma satisfactoria, se dibujan las llamadas líneas epipolares, que resultan de la intersección del plano formado por los focos de las cámaras con el punto en el espacio y las imágenes. Y como se observa en las figuras 4.6 y 4.7 resultan ser paralelas al eje y, lo cual prueba que la rectificación se realizó de forma satisfactoria.

Sistema Estéreo Creative:

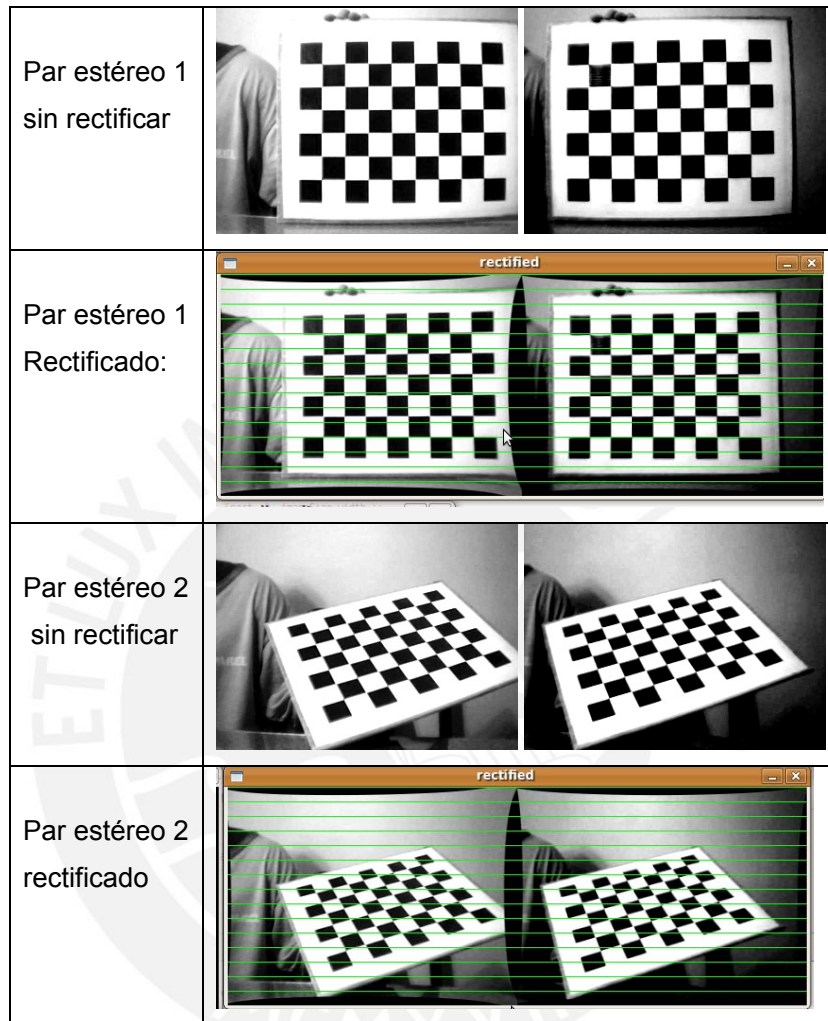


Figura 4.6 Rectificación de imágenes y líneas epipolares (verde)

Sistema estéreo Imaging Source

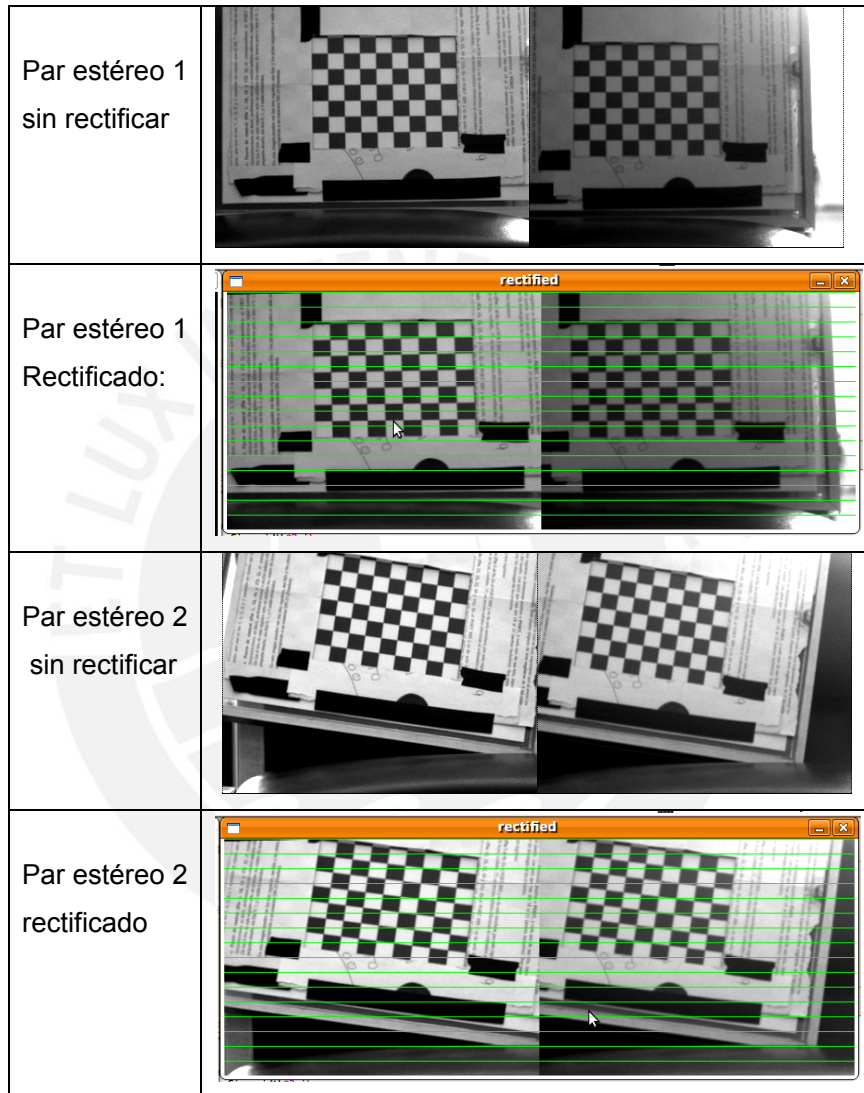


Figura 4.7 Rectificación de la cámaras Imaging Source®

El mapa de rectificación entregado por la función `cvStereoRectify` va a ser posteriormente utilizado en la rutina de obtención de las coordenadas, puesto que la rectificación para las imágenes capturadas será realizada de forma continua.

4.6 Obtención de coordenadas

4.6.1 Algoritmo de obtención de coordenadas

El algoritmo realizado consta de diversas etapas dentro de las cuales tenemos el reconocimiento de las cámaras, rectificación del par estéreo, correspondencia y triangulación. Todas estas etapas previamente estudiadas en el capítulo 3 están presentes en el programa desarrollado que obedece al siguiente diagrama de flujo mostrado en la Figura 4.8

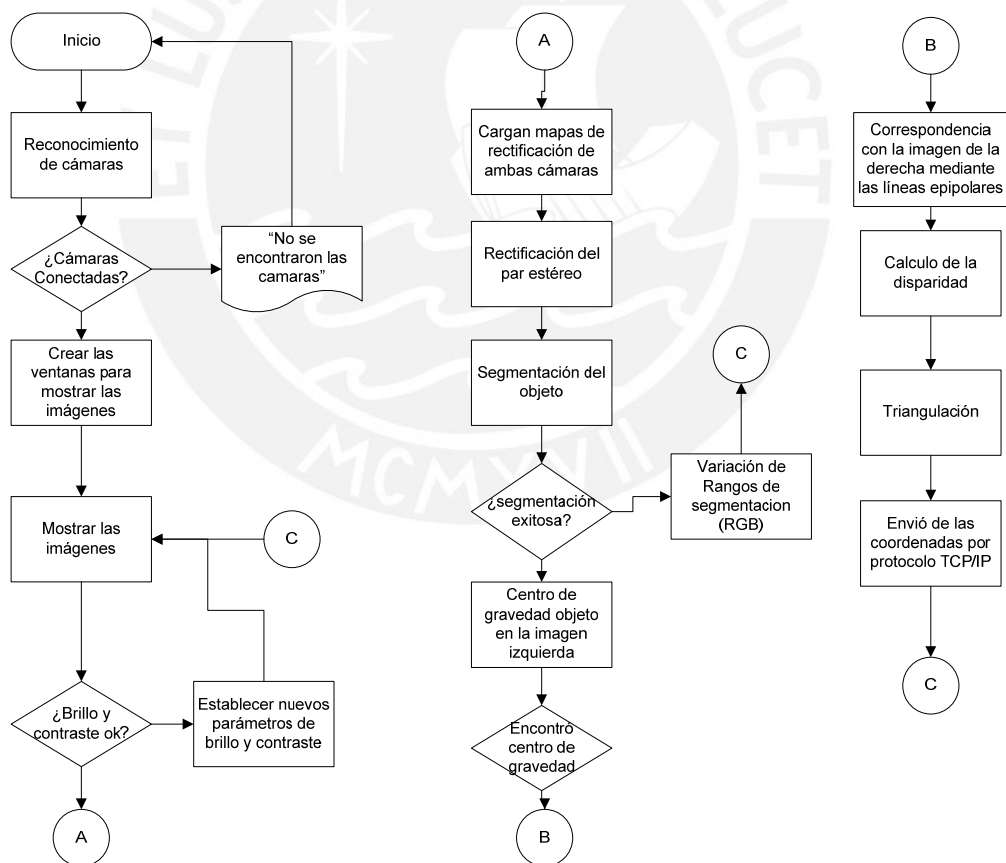
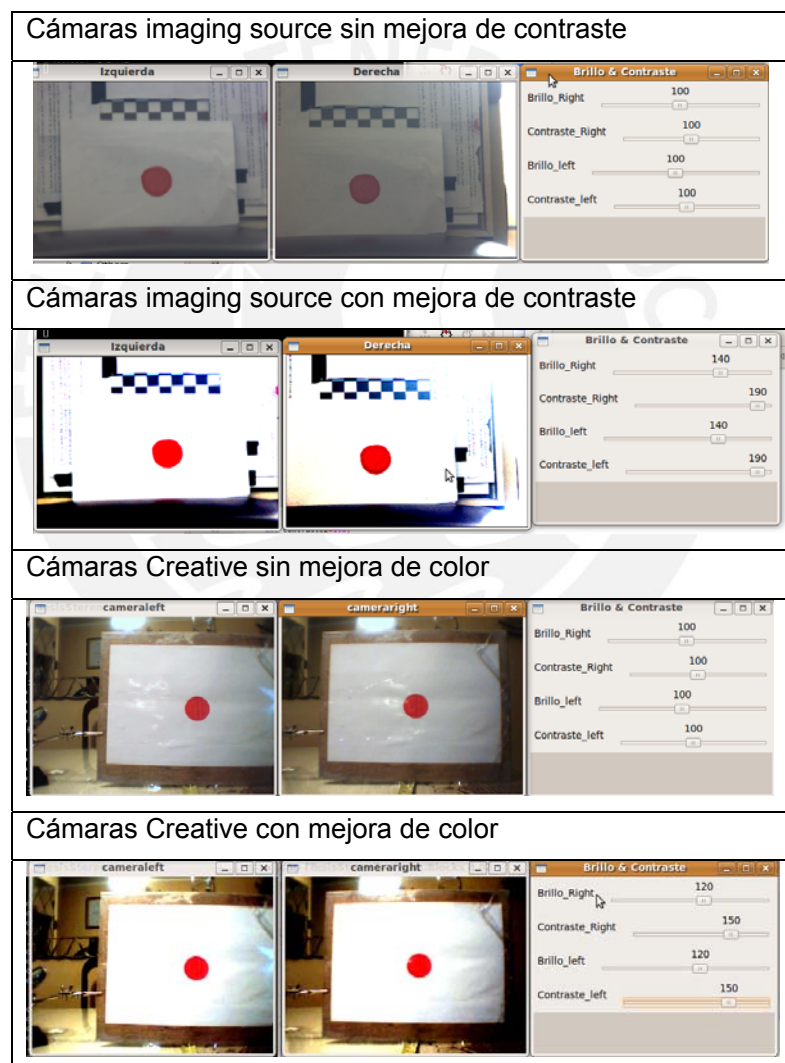


Figura 4.8 Diagrama de flujo del programa de obtención de las coordenadas

4.6.2 Detección y segmentación del objeto

El objeto a detectar será un círculo de color rojo, para tal caso se varían los valores de brillo y contraste de la imagen para mejorar su detección. Mediante la librería OpenCV se usará el código desarrollado por Werner D. Streidt,⁴ cuyos resultados los vemos en la figura 4.9.



⁴ Código fue desarrollado aplicando técnicas de ingeniería inversa para aproximarse a los algoritmos de programas como photoshop o coreldraw en el manejo del brillo y contraste <http://visca.com/ffactory/archives/5-99/msg00021.html>

Figura 4.9 Realce color Rojo con mejora de contraste

De la figura 4.9 se observa una mejor detección del objeto gracias a la modificación de los parámetros del brillo y contraste. Para el caso del sistema imaging source los valores de brillo y contraste adecuados son 140 y 190 respectivamente, esto quiere decir que se ha incrementado en un **40% y 90%** tanto el brillo y el contraste de las imágenes. En el caso del sistema basado en las cámaras creative los valores de brillo y contraste son 120 y 150 con lo que se deduce un incremento tanto de brillo y de contraste de un **20% y 40%** respectivamente.

Una vez detectado el objeto se procede a su segmentación este proceso se gráfica de forma secuencial en la figura 4.10 donde se puede apreciar los procesos de umbralización y segmentación propiamente dicha estudiados en el capítulo 2. Y el resultado de dicha segmentación se puede ver en las figuras 4.11 y 4.12

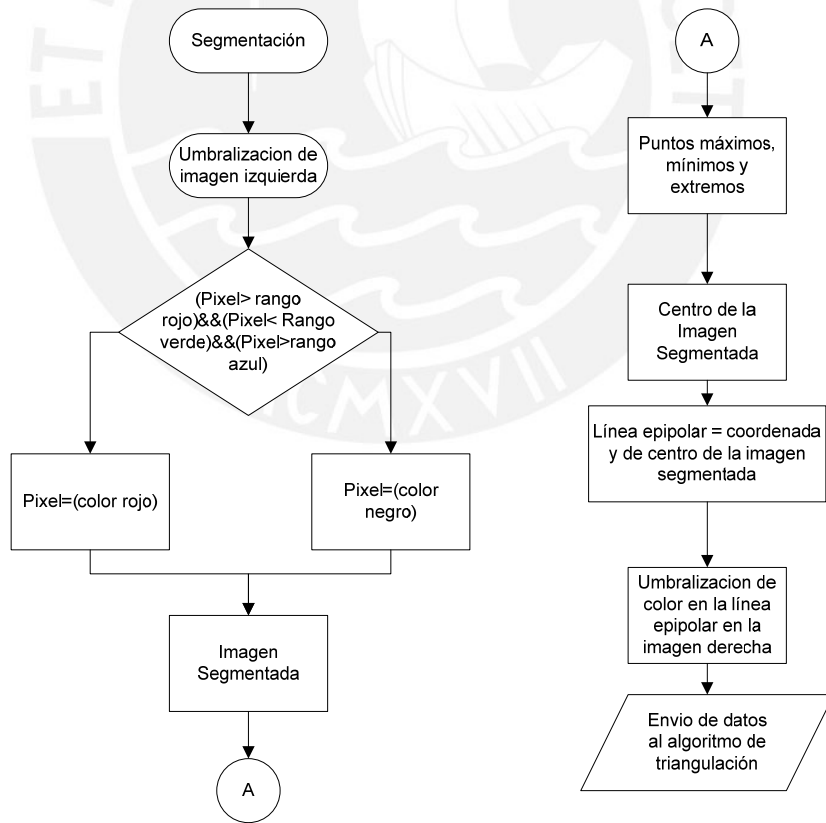


Figura 4.10 algoritmo de segmentación

4.6.3 Coordenadas espaciales

Una vez segmentado el objeto vamos a encontrar las coordenadas de su centro aplicando la triangulación cuya base matemática se ha estudiado en el capítulo 3, obteniendo el resultado que se puede apreciar en la figura 4.11.

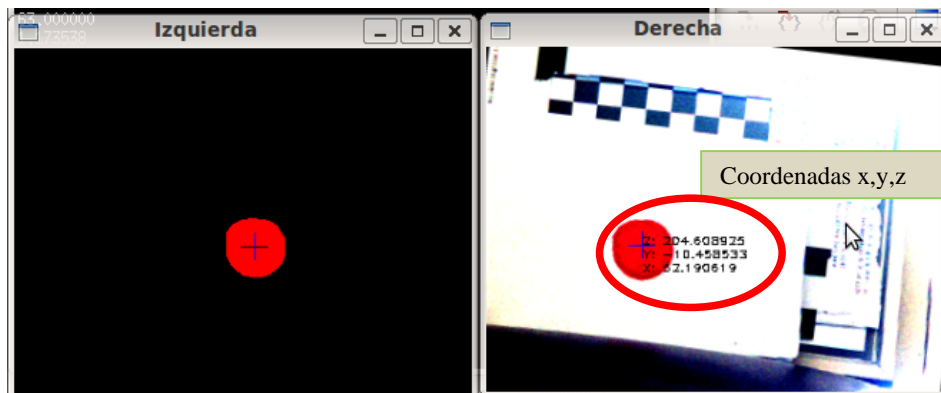
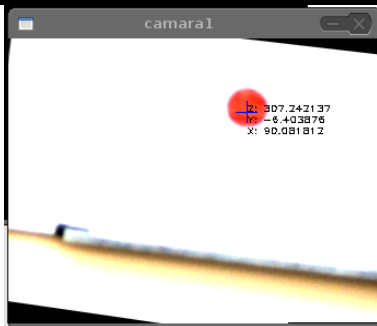
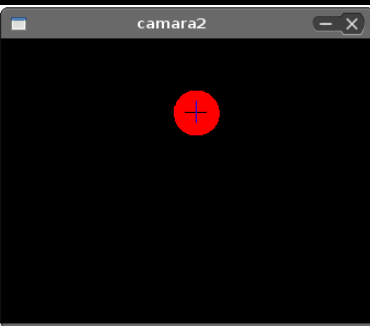


Figura 4.11 Sistema visión estéreo

Para probar el sistema se tomó como punto de referencia una medida fija, que en este caso fue 3 metros de distancia entre el objeto y las cámaras y se tomó la medida en diferentes posiciones del objeto a una misma distancia. Figura 4.12

Posiciones		Distancia
		Posicion superior
		307.24cm

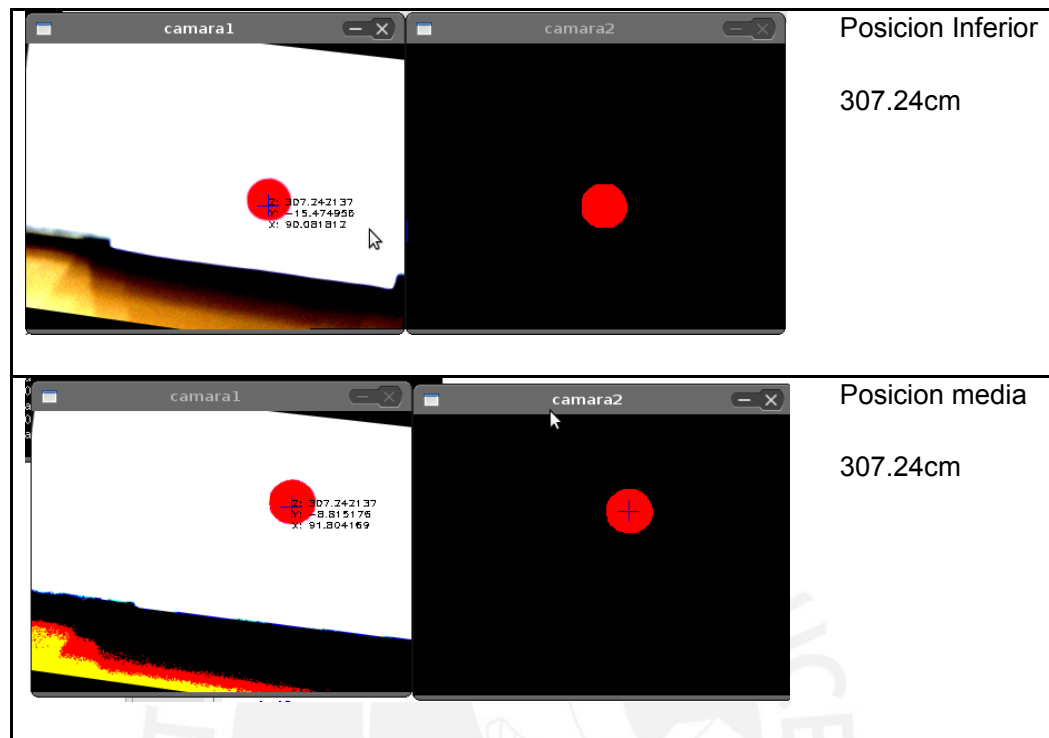


Figura 4.12 tomas a una misma profundidad con diferente ubicación

De los resultados obtenidos, se puede apreciar que para diferentes posiciones del objeto a una misma profundidad, la medida entregada por nuestro sistema no tiene variación alguna, esto quiere decir que el sistema logra calcular la profundidad de forma satisfactoria para diferentes posiciones.

Como hemos visto también en el capítulo 3, la distancia está en función inversa a la disparidad, esto se comprueba al graficar (Figura 4.13) las distancias obtenidas con su correspondiente disparidad; presentando el comportamiento esperado de un sistema estéreo descrito en el capítulo 3 en la figura (3.2).

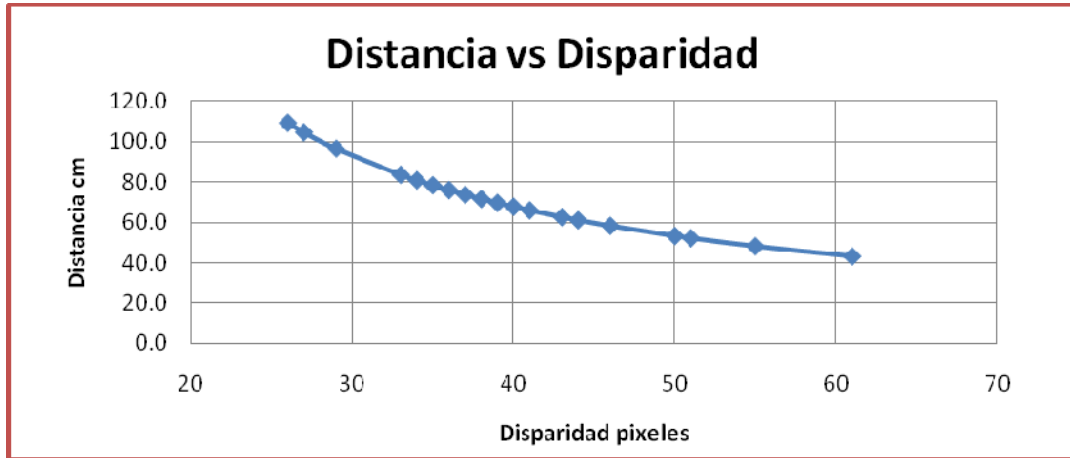


Figura 4.13 Gráfica distancia (del objeto a la cámara) versus disparidad

De la figura 4.13 se observa que la variación de la distancia con respecto a la disparidad de las imágenes no varía en forma lineal sino en forma exponencial.

Como conclusión principal se tiene que la resolución de nuestro sistema es mucho mejor para los objetos que están en un rango de posición entre 30 y 40 cm, puesto que la diferencia entre disparidades para distancias mayores no varían mucho.

Como siguiente paso se pone a prueba la precisión del sistema contrastando la distancia real a diferentes profundidades con la distancia obtenida mediante nuestro sistema estos datos están expresados en la tabla 4.5.

Medida Real(cm)	Calculada(cm)	Error(cm)
27	27,1	0,1
28	28,06	0,06
29	29,04	0,04
30	30,46	0,46
31	31,22	0,22
32	32,44	0,44
33	33,31	0,31
34	34,22	0,22
35	34,7	-0,3
36	36,21	0,21
37	37,3	0,3
38	38,45	0,45
39	39,05	0,05
40	40,97	0,97
41	41,66	0,66

42	42,37	0,37
43	43,1	0,1
45	45,4	0,4
46	46,3	0,3
47	47,1	0,1
48	48,09	0,09
50	50	0
51	51,04	0,04
53	53,22	0,22

Tabla 4.5 Toma de datos de las diferentes medidas en el espacio

De los datos obtenidos observamos que nuestro sistema tiene un error máximo de 0.97cm.; además también se concluye que para distancias menores de 35 cm la exactitud de nuestro sistema mejora notablemente en comparación a otras medidas en donde el error es mayor.

Para poder apreciar mejor los resultados obtenidos, se muestra en la figura 4.14 tanto la medida real como la calculada por el sistema de visión estéreo. En donde se aprecia mejor para que rango de distancias nuestro sistema muestra un mejor desempeño.

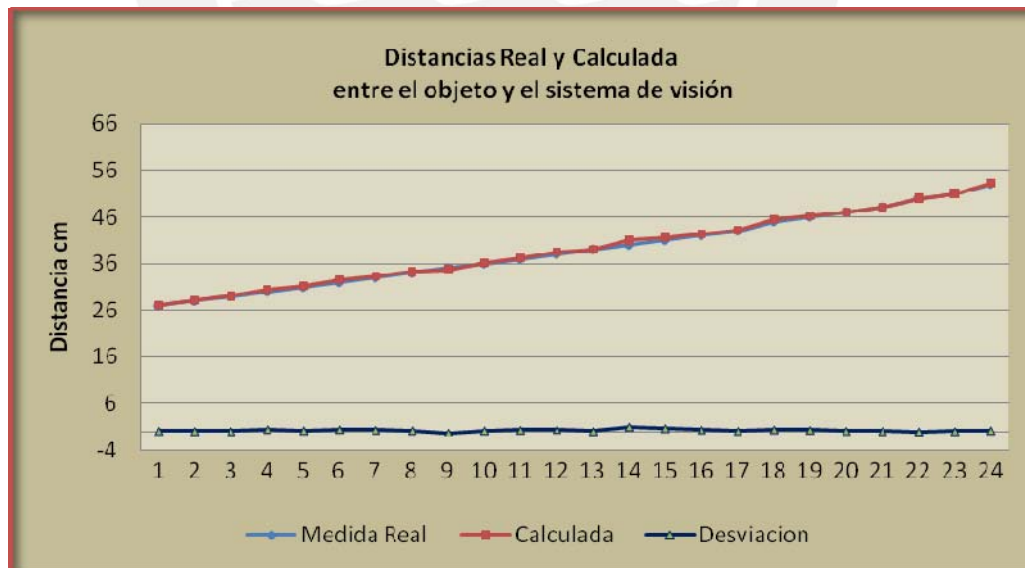


Figura 4.14 Evaluación de los resultados obtenidos al medir la distancia del objeto mediante el sistema de visión estéreo y comparándolos con las medidas reales

Calculando la desviación estándar de los datos anteriores vemos que tiene un valor de **0.25**, esto se interpreta como la variación esperada o en nuestro caso el error esperado del sistema.

4.7 Comunicación con el robot móvil

Para probar el sistema se pone a prueba el algoritmo desarrollado en un robot móvil, ya que para este fin fue desarrollado este trabajo. El robot móvil seleccionado fue el pioneer P-AT el cual se adquirió gracias al proyecto DAI 2008-0049, “Diseño E Implementación De Un Sistema Terrestre De Detección Y Clasificación De Nutrientes En Campos De Cultivos”. Figura 4.15

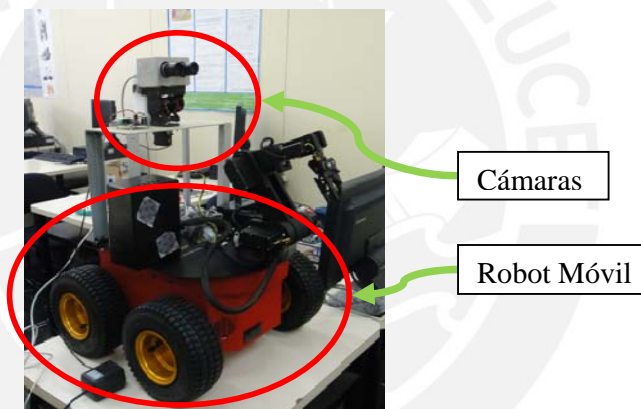


Figura 4.15 Robot móvil Pioneer P-AT con el sistema de visión estéreo implementado

Una vez obtenidas las coordenadas del centro del objeto se envían al robot vía comunicación de red (Figura 4.16) y a su vez también se envían las imágenes del par estéreo rectificado para el monitoreo de la aplicación.

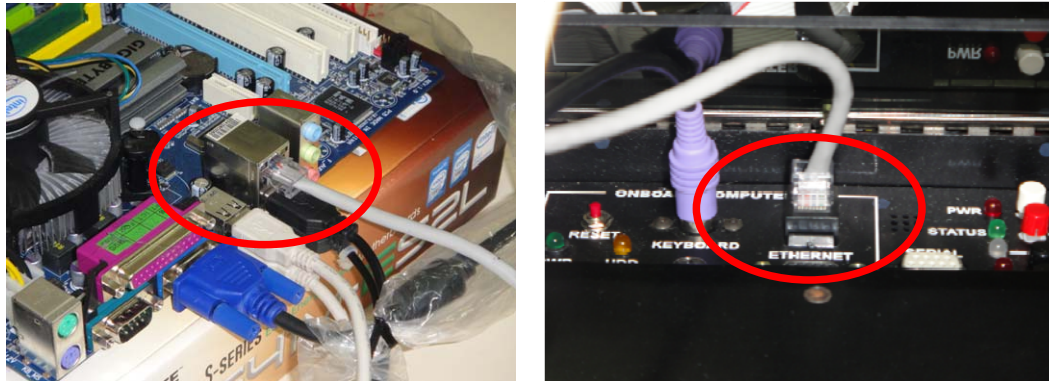


Figura 4.16 Conexión de enlace de red entre el sistema de visión y la computadora del robot móvil

Para la configuración del enlace de red cada una de las computadoras debe ser configurada con una dirección IP diferente, para el caso del servidor esta dirección será la **192.168.1.1** y para la computadora cliente será **192.168.1.2**. Un ejemplo de esta configuración se puede apreciar en la figura 4.17.

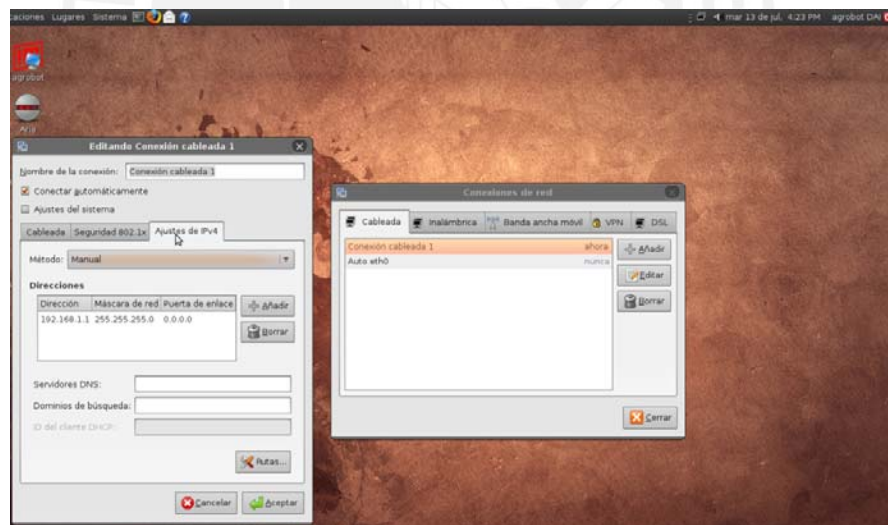


Figura 4.17 configuración de la dirección IP del servidor

Para verificar que la conexión se ha realizado con éxito en la ventana de conexiones de redes verificaremos si la nuestra red está conectada (Debe mostrar “ahora” para indicar cuál de las redes está siendo utilizada).Figura 4.18

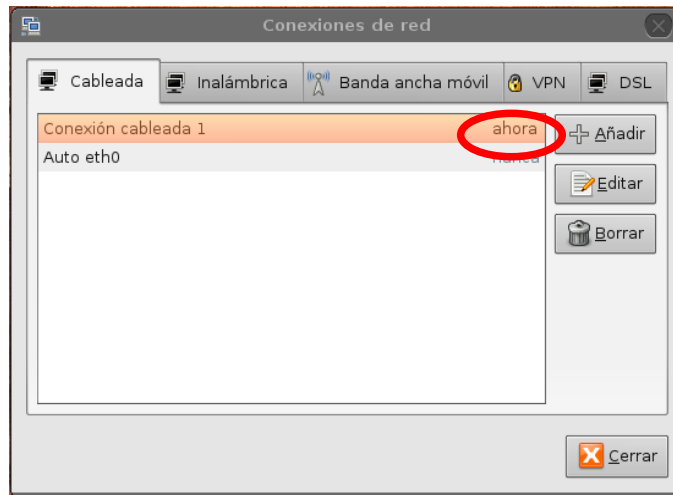


Figura 4.18 Verificación de la conexión de red

4.8 Análisis Económico

El presupuesto de los recursos tecnológicos usados tanto para hardware como para software es:

Artículo	Cantidad	Precio unidad (\$)	Total (\$)
S.O. Linux Ubuntu 9.04	1	0.00	0.00
Programa CodeBLoocks	1	0.00	0.00
Librería OpenCV 2.0	1	0.00	0.00
Cámaras Imaging source modelo DFK 31BU03.H	2	650.00	1300.00
Camaras Creative modelo Cam Notebook Pro (VF0250)	2	30.00	60.00
Placa madre modelo GA-EG31M-S2	1	70.00	70.00
Disco duro 80gb modelo Seagate Momentus 5400.3	1	60.00	60.00

Tabla 4.6 Presupuesto utilizado para el desarrollo del proyecto.

CONCLUSIONES

- 1 Se ha desarrollado un sistema de visión estereoscópica para el robot móvil pioneer P3-AT, capaz de calcular distancias de hasta 2m con un error máximo de 0.91 cm.
- 2 Se ha logrado la segmentación de un objeto de manera exitosa, fue necesario para contar con un fondo controlado y de contraste con el objeto de estudio.
- 3 Las etapas de calibración del sistema así como la rectificación del mismo fueron llevadas a cabo de forma satisfactoria mediante el uso de la librería OpenCV.
- 4 Se ha desarrollado una interfaz de usuario mediante la cual se puede observar la identificación de los principales resultados; así como la manipulación de algunos parámetros como es el caso de brillo, contraste y umbral.
- 5 Se logró la comunicación con el robot móvil modelo Pioneer P3-AT de la corporación ActivMedia®, mediante protocolo TCP/IP logrando así una transmisión de datos (imágenes y coordenadas) eficiente.

RECOMENDACIONES

1. El presente trabajo se puede ampliar para el modelamiento en 3D de objetos ya que se trata de hallar la correspondencia y profundidad de no solo un punto sino de varios.
2. El algoritmo de segmentación de la imagen no es muy complejo puesto que el costo de procesamiento tuvo que ser bajo para tener la respuesta esperada en el robot; sin embargo se recomienda probar técnicas avanzadas de reconociendo de objetos en sistemas más robustos.
3. Se propone la implementación de este algoritmo bajo un sistema en tiempo real, para aplicaciones que requieran una velocidad de procesamiento mucho mayor.
4. Se recomienda el desarrollo de algoritmos de segmentación adaptativos para evitar la influencia del cambio de luz en el ambiente.

BIBLIOGRAFÍA

- BRADSKI Gary y KAEHLER Adrian. 2008
Learning OpenCV. Sebastopol, CA: O'Reilly Media, Inc,
- C. GONZALES Rafael, E. WOODS Richard y L. EDDINS Steven. 2004
Digital Image Processing using Matlab. New Jersey : Pearson Prentice Hall,
- CAPILLA Nicolás Pérez de la Blanca. 2002
Algoritmos de Estimación de la Geometría de Múltiples Vistas. Montevideo
- CYTED Programa Ibero Americano Para el Desarrollo, Subprograma VII. 2002
Santiago de Chile
- HU Housheng y Q. GAN John. 2005.
Sensor and Data Fusion Algorithms in Mobile Robots. Colchester : Departamen of
computer science university of Essex
- IEE-RAS. 2006
Consejo Latinoamericano de Robótica [En línea] [consultado junio de 2010]
<http://ewh.ieee.org/reg/9/robotica/indexsp.htm>.
- LUKAC Rastislav. 2009
Single-Sensor Imagin Methods and Applications for Digital Cameras. New York : Taylor
& Francis Group
- M. VALIENTE Jose, ANDREU Gabriela y RODAS Ángel. 2010
Vision por Computador- Iluminacion. Valencia
- MARTINSANZ Gonzalo y DE LA CRUZ GARCIA Jesus . 2008
Vision por Computador Imágenes digitales y aplicaciones. Mexico: alfaomega
- OROZCO Duque y ANDRÉS Felipe. 2001
Sensores, Métodos y técnicas para la navegación de robots móviles. Medellín:
Universidad Pontifica Bolivariana
- R. GONZALES Victor. 2002
Robotica [En línea][consultado 7 de 11 de 2009]
http://cfievalladolid2.net/tecno/cyr_01/robotica/.
- RUIZ MORENO Carlos, 2005
Visión estereoscopica técnicas de correspondencia. Madrid : Universidad Politecnica
de Madrid
- SONKA Milan, HLAVAC Vaclav y BOYLE Roger. 2008
Image Processing, Analysis, and Machine Vision. Toronto: Thomson
- Soudronic AG. 2010

Soudronic [Enlínea].consultado[2010/6/18]
<http://www.soudronic.com/canbody/index.html>.

TRUCCO Emanuele y VERRI Alessandro. 1998
Introductory techniques for 3-d Computer Vision. New Jersey : Prentice-Hall

Wikipedia, 2010
la enciclopedia libre [En línea]. consultado [2010/4/20]
<http://es.wikipedia.org/wiki/Wikipedia:Portada>.

ANEXOS:

Se adjunta un CD conteniendo el código desarrollado además de diversos archivos y hojas de datos en formato PDF que refuerzan lo planteado en este trabajo de tesis.

