



Instituto Politécnico Nacional

CICATA - Legaria



Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada

Posgrado en Tecnología

Avanzada

“Reconocimiento de objetos a partir de visión estereoscópica “.

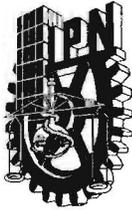
Tesis que para obtener el grado de Doctorado en Tecnología Avanzada Presenta:

M. en T. A. Alejandro Israel Barranco Gutiérrez

Directores:

Dr. José de Jesús Medel Juárez
Dr. José Antonio Calderón Arenas

México DF., Junio del 2010.



INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México D.F. siendo las 12:00 horas del día 18 del mes de Mayo del 2010 se reunieron los miembros de la Comisión Revisora de Tesis designada por el Colegio de Profesores de Estudios de Posgrado e Investigación de CICATA-IPN para examinar la tesis de titulada:

Reconocimiento de objetos a partir de visión estereoscópica

Presentada por el alumno:

Barranco

Apellido paterno

Gutiérrez

Apellido materno

Alejandro Israel

Nombre(s)

Con registro:

B	0	6	1	5	0	7
---	---	---	---	---	---	---

aspirante de:

Doctorado en Tecnología Avanzada

Después de intercambiar opiniones los miembros de la Comisión manifestaron **SU APROBACION DE LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de Tesis

Dr. José de Jesús Medel Juárez

Dr. José Antonio Calderón Arenas

Dr. Ernesto Marín Moares

Dr. Miguel Ángel Aguilar Frutis

Dr. César Eduardo Mora Ley



EL PRESIDENTE DEL COLEGIO

Dr. José Antonio Díaz Gongora



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México el día 4 del mes junio del año 2010, el que suscribe **M en T.A. Alejandro Israel Barranco Gutiérrez** alumno del Programa de **Doctorado en Tecnología Avanzada** con número de registro **B061507**, adscrito a **CICATA Unidad Legaria**, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección del **Dr. José Antonio Calderón Arenas y del Dr. José de Jesús Medel Juárez** y cede los derechos del trabajo intitulado **“Reconocimiento de objetos a partir de visión estereoscópica”**, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección **barranco_alejandro@yahoo.com.mx**. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

M en T.A. Alejandro Israel Barranco Gutiérrez

Resumen

Este trabajo de tesis presenta una innovadora solución al problema del reconocimiento de objetos 3D en imágenes, considerando la información de tres dimensiones (3D) proporcionada por un sistema de visión estereoscópico con cámaras digitales convencionales. La necesidad surge de la flexibilidad requerida por la dinámica de los sistemas en la vida real, donde los objetos no pueden ser capturados por módulos de visión por computadora que requieren el objeto a reconocer, a una distancia fija respecto de las cámaras digitales que lo capturan ó donde las dimensiones de los objetos son factor importante para realizar el reconocimiento.

En este trabajo se propone la solución con un tipo de reconocimiento de objetos en 3D basados en la visión estereoscópica y en la estadística del objeto a reconocer. Se obtuvieron resultados que abordan este tema desde una perspectiva diferente a la que comúnmente se esta haciendo alrededor del mundo (como se presenta en los antecedentes). Ya que actualmente se encuentra una división entre los investigadores que perfeccionan las técnicas de reconocimientos de patrones y quienes reconstruyen entornos en 3 Dimensiones. La metodología de esta técnica puede utilizarse para muchas aplicaciones sin modificar la estructura de esta y pueden ser perfeccionados algunos parámetros para la aplicación que se requiera utilizar. Este método descubre de manera automática los detalles para encontrar matemáticamente los rasgos de cada objeto en la imagen, esta tarea no se deja en las manos del talento de una persona, por lo tanto se comprueba la versatilidad del sistema.

Se ha reportado el funcionamiento de la metodología propuesta de reconocimiento de objetos 3D de esta tesis. En los cuales podemos observar que se logra el reconocimiento de objetos similares donde la diferencia entre los objetos radica principalmente en las dimensiones volumétricas de los cuerpos.

Abstract

This thesis presents a novel solution to the problem of 3D object recognition in images, considering the information of three-dimensional (3D) provided by a stereo vision system with conventional digital cameras. The need became from the flexibility required by the dynamics of systems in real life, where objects can not be captured by computer vision modules to do the object recognition or exist a fixed distance from digital cameras that capture it or where the dimensions of objects are important for recognition.

This paper proposes a solution with a type of 3D object recognition based on stereoscopic vision and the statistics of the object to recognize. We obtained results that address this topic from a different perspective than is commonly being done around the world (as presented in the background). Since it is now a division between researchers who perfected the techniques of pattern recognition, and those environments reconstructed in 3 dimensions. The methodology of this technique can be used for many applications without changing the structure of this and can be improved some parameters for the application presented to use. This method automatically discovers the details to find mathematically the characteristics of each object in the image; this task is not left in the hands of a person's talent, so it checks the system versatility.

It is reported the operation of the proposed methodology for 3D object recognition of this thesis. In which we can observe that achieved the recognition of similar objects where the difference between objects is mainly based on volumetric dimensions of the bodies.

Agradecimientos

A mi madre **Araceli Gutiérrez**, por todo el amor que he recibido incondicionalmente.

Gracias a Dios, por darme la salud para disfrutar de la vida y concluir esta etapa de mi vida.

Gracias todos los mexicanos, que me Han dado la oportunidad de estudiar.

Agradezco al *Dr. José de Jesús Medel Juárez* y al *Dr. José Antonio Calderón Arenas* por el inmenso apoyo que me han brindado.

A los Doctores, *Miguel Ángel Aguilar Frutis*, *José Luís Jiménez*, *Gastón Hugo Salazar Silva*, *Ana Lilia Laureano Cruces* por sus importantes enseñanzas para mi vida.

Agradezco a mis Profesores *Jean Bernard Hayet*, *Javier Flavio Viguera Gómez*, y *Kyoichi Tatsuno*, por enseñarme a ver el Mundo de una manera distinta y recibirme en el mundo de la visión por computadora.

A mi hermana *Berenice* y mi tío *Jaime* por el apoyo que siempre me han dado en el momento justo.

A mí amada Paola por sus atenciones.

A *Leticia Cardona*, *Ing. Pablo Ortiz*, por sus valiosos consejos y apoyo.

Al **CONACYT** por apoyarme en este proyecto de investigación.

Al **Instituto Politécnico Nacional** por permitirme vivir y llegar hasta estas instancias, la culminación de mis estudios de Doctorado.

Gracias

Alejandro Israel Barranco Gutiérrez

Índice

Capítulo	Capitulado	Página
CAPITULO 1. Introducción	1.1 Objetivo Científico	11
	1.2 Marco de referencia	12
	1.3 Enfoque de la tesis	12
		14
CAPITULO 2. Antecedentes	2.2 Cámaras digitales	16
	2.3 El dispositivo CCD como elemento básico en la captura de imágenes	17
	2.4 La fotografía digital a color como una matriz de 3 dimensiones	18
	2.5 Modelo de cámara fotográfica pinhole	19
	2.6 Transformación lineal de coordenadas del espacio a coordenadas de una fotografía	21
	2.7 Parámetros extrínsecos	21
	2.8 Parámetros intrínsecos	23
	2.9 Métodos de calibración	25
	2.9.1 Método de Fougeras	26
	2.9.2 Método de Zhang	28
	2.10 Reconocimiento de patrones	30
2.11 Componentes de un SRAO	34	
CAPITULO 3. Estado Del Arte	3.1 Calibración estereoscópica	35
	3.2 Aproximación de un Punto 3D en base a 2 fotografías	38
	3.3 Reconocimiento de objetos a partir del reconocimiento de patrones	40
	3.4 Reconocimiento de objetos en 2D por componentes principales	42
CAPITULO 4. Reconocimiento de Objetos En 3D	4.1 ¿Por qué hacer reconocimiento de objetos en 3D?	43
	4.2 ¿Cuál es la aportación de la presente tesis?	47
	4.3 Experimentación	47
	4.3.1 Aproximación de un Punto 3D en base a 2 fotografías usando visión estereoscópica	50
	4.3.2 La calibración de las cámaras	50
	4.3.3 Obtención de imágenes y conversión a escala de grises	51
	4.3.4 Filtrado y Umbralado	53
	4.3.5 Detección de bordes	53
	4.3.6 Etiquetado y separación de objetos en la imagen	54
4.3.7 Prerreconocimiento de objetos sin dimensionalidad	55	
4.3.8 Medición automática de dimensiones	56	
CAPITULO 5. Conclusiones		58
Apéndice A: Flexible Camera Calibration By Viewing a Plane From Unknown Orientations, Zhang.		61
Apéndice B: Reconocimiento automático de objetos en imágenes estereoscópicas por medio de características dimensionales		

Lista de Figuras

Pie de figura	Página
Figura 1. <i>Ilustra un mini robot que navega de forma autónoma e identifica los obstáculos a su alrededor con la técnica de visión estereoscópica.</i>	11
Figura 2. <i>X es un punto en el espacio (x,y,z), C son los centros de referencia de cada una de las cámaras que están captando el punto X . Las líneas punteadas azul y rojo son todas las posibles soluciones al punto que observa cada una de las cámaras, la línea verde el vector de posición que localiza a cada una de las cámaras respecto de otra.</i>	13
Figura 3. <i>En la figura se muestran diferentes formas como rectángulos, un triángulo y la sombra de una cara que se pregunta: ¿Qué objetos tengo frente a mí? Es lo que se pregunta una computadora antes de reconocer las formas en una imagen.</i>	14
Figura 4. <i>Imagen digital de una habitación tomada por una cámara digital para ser analizada.</i>	18
Figura 5. <i>Sistema CCD para capturar imágenes digitales, matriz de sensores y circuito acondicionador de señales.</i>	19
Figura 6. <i>La imagen digital a color del tipo RGB se representa por medio de 3 matrices de dimensiones $M \times N$ donde cada matriz representa la cantidad de color rojo, verde o azul de la imagen.</i>	20
Figura 7. <i>El principio del modelo de cámara PinHole invierte la imagen de los objetos externos en el interior del sistema.</i>	21
Figura 8. <i>Sistema de referencia del modelo PinHole montado en el plano donde se encuentra el orificio (PinHole) del sistema.</i>	22
Figura 9. <i>Una fotografía que muestra personas en distintas posiciones del espacio.</i>	23
Figura 10. <i>La transformación Lineal modela correspondencias entre el sistema de coordenadas del objeto en el espacio y sus correspondencias en el sistema de coordenadas de la cámara.</i>	24
Figura 11. <i>Rotación (α, β, θ) de un punto alrededor de los ejes X, Y, Z respectivamente.</i>	25
Figura 12. <i>La geometría del píxel influye en la manera de modelar las imágenes</i>	27

<i>digitales, las dimensiones de f_x y f_y; así como el ángulo s, también llamado sesgo.</i>	
Figura 13. <i>Aquí se muestra el patrón para calibrar en el método de Zhang, con diferentes orientaciones y traslaciones para una mejor aproximación.</i>	30
Figura 14. <i>A la izquierda una imagen con algunos de los objetos conocidos que se enumeran a la derecha de la presente figura.</i>	34
Figura 15. <i>Diagrama a bloques de un SRAO típico con flujos de información.</i>	35
Figura 16. <i>Para obtener un mejor rango de trabajo se aumenta el contraste de una imagen, mejorándola para procesamiento posterior.</i>	36
Figura 17. <i>El segmentador se encarga de extraer sólo los objetos que interesan al análisis que se está desarrollando en cada aplicación.</i>	36
Figura 18. <i>Se muestran dos sistemas de referencia en azul claro y azul oscuro, donde el primero localiza al segundo y los puntos expresados a partir del sistema de referencia azul oscuro pueden expresarse desde el sistema de referencia azul claro.</i>	39
Figura 19. <i>(a) Dos cámaras separadas 70 mm., a una distancia infinita la disparidad estéreo es cero. (b) Cámaras acopladas para uso en visión estereoscópica.</i>	39
Figura 20. <i>Cámaras calibradas, donde se conoce la transformación rígida del sistema de coordenadas de la cámara izquierda a la cámara derecha.</i>	40
Figura 21. <i>Diagrama a bloques de un sistema de reconocimiento de patrones en general donde el sensor puede ser de audio, imágenes, señales electromagnéticas y otros.</i>	42
Figura 22. <i>Ejemplo de la recta que minimiza las distancias ortogonales de los puntos a ella.</i>	44
Figura 23. <i>Sistema de control de calidad automático, trabaja por visión, compara las dimensiones de la pieza a revisar con la pieza modelo.</i>	48
Figura 24. <i>Diagrama de flujo de la metodología propuesta para el reconocimiento de objetos en tres dimensiones.</i>	49
Figura 25. <i>Se muestra una escena tomada de 2 diferentes puntos de vista. El punto de interés es la esquina superior derecha de la vista frontal del libro.</i>	50
Figura 26. <i>En la parte superior se muestran las imágenes de las fotografías en</i>	53

<i>escala de grises filtradas del tipo mediana de 10 x 10, en la parte inferior las imágenes binarizadas por el método de Otsu.</i>	
Figura 27. <i>Las imágenes de la parte inferior muestran el resultado de la detección de bordes de las imágenes de la parte superior por el método de sobel.</i>	54
Figura 28. <i>Elementos de las imágenes binarizadas presentados en imágenes independientes para ser analizados.</i>	55
Figura 29. <i>Se detecta el punto superior de la esfera y el punto inferior de la misma para medir su diámetro.</i>	58

Lista de Tablas

Pie de tabla	Página
Tabla 1. Principales aplicaciones de la visión por computadora.	17
Tabla 2. <i>Los resultados (parámetros intrínsecos) de calibrar la cámara izquierda del sistema en formato Heikkilä.</i>	51
Tabla 3. <i>Los resultados (parámetros intrínsecos) de calibrar la cámara derecha del sistema en formato Heikkilä.</i>	51
Tabla 4. <i>Resultados de calibración estéreo, donde se indican los 3 ángulos de rotación en el vector de traslación respecto de la cámara izquierda.</i>	52
Tabla 5. <i>Valores de los 7 invariantes de Hu para las 3 esferas contenidas en la imagen de la cámara derecha.</i>	57
Tabla 6. <i>Los límites para la clasificación de esferas tomados de forma experimental solo seleccionando máximos y mínimos y redondeando.</i>	57
Tabla 7. <i>Se presentan las coordenadas (en píxeles) de los dos <u>puntos superiores</u> de la esfera en la cámara izquierda (x_{left_1}) y derecha (x_{right_1}), así como sus coordenadas en el espacio respecto a la cámara izquierda (Xc_1_left).</i>	58
Tabla 8. <i>Se presentan las coordenadas (en píxeles) de los dos <u>puntos inferiores</u> de la esfera en la cámara izquierda (x_{left_1}) y derecha (x_{right_1}), así como sus coordenadas en el espacio respecto a la cámara derecha.</i>	59

1

Introducción

Un robot debe cumplir con las siguientes leyes

1. *Un robot no puede dañar a un ser humano ni, por inacción, permitir que éste sea dañado.*
2. *Un robot debe obedecer las órdenes dadas por los seres humanos excepto cuando estas órdenes entren en conflicto con la Primera Ley.*
3. *Un robot debe proteger su propia existencia hasta donde esta protección no entre en conflicto con la Primera o la Segunda Ley.*

Isaac Asimov y John W. Campbell



En esta tesis se presentan los resultados de la investigación realizada para el Programa de Doctorado en Tecnología Avanzada en el área de control e instrumentación, del CICATA – Legaria del Instituto Politécnico Nacional (IPN). La investigación tiene por objetivo resolver el problema del reconocimiento de formas a partir de la visión estereoscópica por computadora. La Figura 1., ilustra ampliamente el ejemplo de un mini robot que navega de forma autónoma y utiliza la técnica de visión estereoscópica para percibir su entorno.

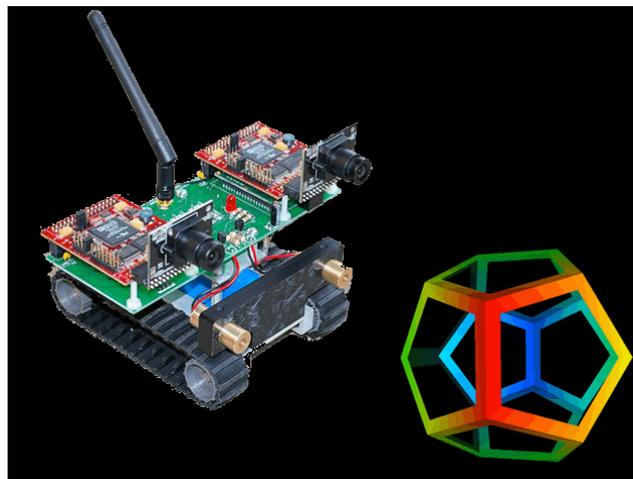


Figura 1. *Ilustra un mini robot que navega de forma autónoma e identifica los obstáculos a su alrededor con la técnica de visión estereoscópica.*

1.1 Objetivo Científico

Demostrar que se puede realizar el reconocimiento de objetos con la misma silueta pero de diferentes dimensiones automáticamente, utilizando cámaras digitales convencionales de bajo costo. El problema resuelve otros problemas derivados como: falsificación de identidad (billetes, rostros, huellas y formas en general que se detectan por medio de imágenes), en el caso de la robótica, se pueden calcular los puntos finales de una trayectoria planeada para el movimiento de articulaciones, que es para lo que se ha enfocado este trabajo.

1.2 Marco de referencia

Se obtuvieron resultados que abordan este tema desde una perspectiva diferente a la que comúnmente se está haciendo alrededor del mundo (como se presenta en los antecedentes). Ya que actualmente se encuentra una división entre los investigadores que perfeccionan las técnicas de reconocimientos de patrones y quienes reconstruyen entornos en 3 Dimensiones.

La visión por computadora es una disciplina que tiene por objetivo hacer que un sistema automático entienda su entorno a través de imágenes del mismo, esto con la finalidad de que sistemas artificiales tomen decisiones basados en la información del entorno obtenida a partir de fotografías ó videos digitales. La visión estereoscópica está enfocada a analizar dos imágenes de la misma escena tomadas al mismo tiempo, con la finalidad de obtener diferentes puntos de vista de los objetos en la escena y en consecuencia más información de los objetos de análisis. Específicamente lo que comúnmente se busca con esta metodología es, el conocer “la profundidad en la fotografía”. Otro de los aspectos importantes que nos aporta es la capacidad de medir las dimensiones de los objetos presentes en la escena, siempre y cuando dichos objetos se encuentren dentro ambas imágenes del sistema y las cámaras se encuentren calibradas.

La calibración de cámara fotográfica o comúnmente llamada calibración de cámara es simplemente obtener los parámetros de la misma, necesarios para conocer de forma automática las dimensiones de los elementos presentes en escenas tomadas con la cámara calibrada por medio de la computadora. Los parámetros antes mencionados y las dimensiones de los objetos en la fotografía se relacionan matemáticamente para obtener las dimensiones reales de los elementos a estudiar en la escena. La calibración del sistema estereoscópico se basa en la calibración individual de cámaras que lo conforman (parámetros de calibración), el conocimiento del vector de traslación que une a ambos sistemas de referencia (el vector que une el centro de referencia de la cámara izquierda con el centro de referencia de la cámara de la derecha) y los ángulos de rotación entre los ejes de ambos sistemas. Básicamente el problema en visión estereoscópica es: encontrar un punto en el espacio de 3 dimensiones a partir de 2 fotografías distintas del punto a estimar, todo con

respecto a un sistema de referencia. Esto normalmente se resuelve por aproximación geométrica, comúnmente llamada triangulación, como se muestra en la Figura 2.

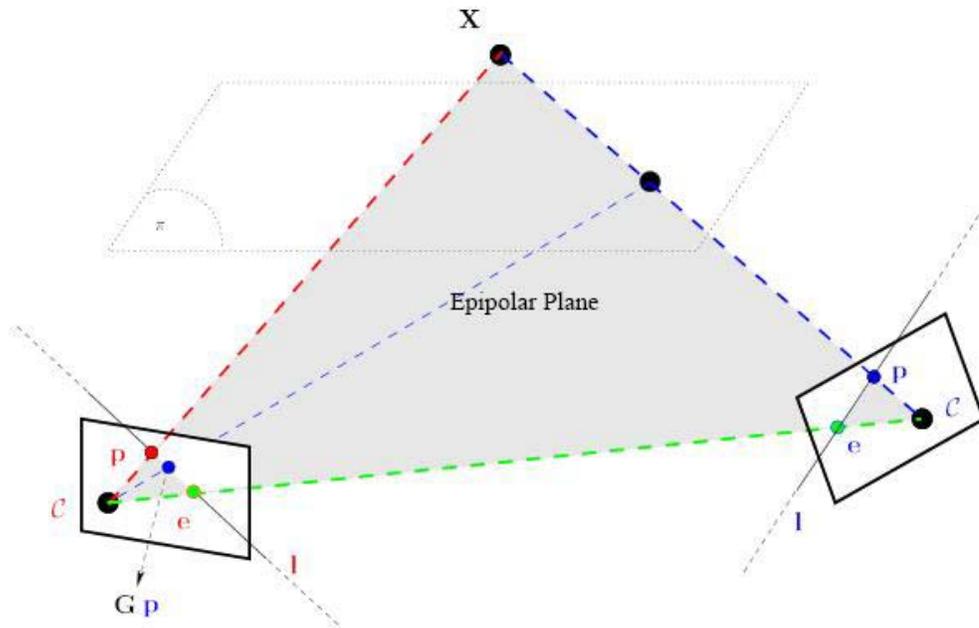


Figura 2. *X es un punto en el espacio (x,y,z) , C son los centros de referencia de cada una de las cámaras que están captando el punto X . Las líneas punteadas azul y rojo son todas las posibles soluciones al punto que observa cada una de las cámaras, la línea verde el vector de posición que localiza a cada una de las cámaras respecto de otra.*

Como se observa, la calibración de cámaras fotográficas digitales se refiere al conocimiento de variables que tienen que ver con la configuración de dichos sistemas y no con ajustarlo para que logre cumplir con ciertas restricciones. Este fenómeno nos dota de flexibilidad para resolver el problema con cualquier cámara fotográfica digital que logre expresar de manera detallada el fenómeno a analizar dentro de la escena y que sus parámetros sean estables en cierto grado.

Otro tema fundamental en este trabajo es el reconocimiento de objetos también llamado reconocimiento de patrones que clasifica objetos presentes en las imágenes de una escena, a partir de información estadística de los objetos presentes en la fotografía. Para lograr analizar la imagen se separan a cada objeto presentes en la imagen de los demás, a este proceso se le llama segmentación y etiquetado de objetos, después a cada objeto se le compara con la estadística de objetos predefinidos y se observa con cual se tiene menos diferencia, el par que resulte tener la menor serán asociados. Con esto se logra clasificar a los objetos de acuerdo a su forma. A este proceso se le llama clasificación. En la figura 3., se encuentran diferentes

objetos que como personas podemos clasificar fácilmente, como un triángulo o un rectángulo; esto es lo que se busca hacer con una máquina que clasifique automáticamente.

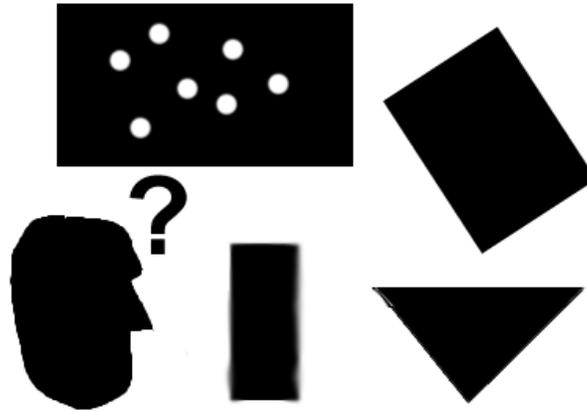


Figura 3. En la figura se muestran diferentes formas como rectángulos, un triángulo y la sombra de una cara que se pregunta: ¿Qué objetos tengo frente a mí? Es lo que se pregunta una computadora antes de reconocer las formas en una imagen.

Aportes del reconocimiento de formas a partir de la visión estereoscópica por computadora:

- ♣ Se tiene la información necesaria para determinar las dimensiones del objeto de estudio
- ♣ Se pueden eliminar de la escena elementos que no son necesarios en el análisis del objeto, con solo discriminar puntos que sean asociados a cierta región del espacio en 3 dimensiones.
- ♣ Se pueden utilizar herramientas estadísticas de reconocimiento de objetos en 2 dimensiones.

Además de las ventajas que cada una de las disciplinas (Visión estereoscópica y reconocimiento de patrones) aportan a la solución de problemas en general; que están referenciadas en los antecedentes.

1.3 Enfoque de la tesis

Se optó por trabajar en el reconocimiento de objetos 3D por medio de un sistema estereoscópico (que describe puntos en el espacio); ya que el reconocimiento de objetos a partir de imágenes monoculares está limitado por: la oclusión entre objetos en la imagen, el

desconocimiento de las dimensiones reales del objeto capturado en la imagen y/o la forma en tres dimensiones (3D) de los objetos. Otro de los problemas que pretende resolver la Visión por Computadora (VC) es la reconstrucción tridimensional de los objetos, que se encuentran en una escena. Este problema también es resuelto en esta tesis; con mejores resultados que los que presentan en la visión monocular. Este problema puede definirse como: La recuperación digital de la estructura espacial de una escena, sin el conocimiento previo de sus dimensiones, localización, número y medidas de los objetos que están contenidos en el espacio de análisis [1].

Capítulo

2

“El hombre encuentra a Dios detrás de cada puerta que la ciencia logra abrir”

Albert Einstein

Antecedentes



Una versión amplia de la historia de la visión por computadora se encuentra en [1] y de forma resumida en [2]. Uno de los motivos más importantes que inspiraron a trabajar en esta área de investigación a miles de científicos y a nosotros; es que la visión en las personas, es uno de los mecanismos sensoriales de percepción más importantes en el ser humano, aunque evidentemente no es exclusivo [3]. Además, muchas de las decisiones que tomamos, se basan en nuestra percepción visual [4][5][6]. Para comenzar a hablar sobre esta disciplina, comencemos hablando de la clasificación de las principales áreas de estudio de esta, como se muestra continuación:

- Reconocimiento de Patrones (Pattern Recognition).
- Reconstrucción tridimensional y fotogrametría (3D Reconstruction).
- Detección de objetos en movimiento (Tracking).
- Reconocimiento de objetos tridimensionales (3D Object Recognition).

En visión por computadora, el reconocimiento de patrones básicamente se encarga de identificar objetos presentes en las imágenes por medio de cálculos estadísticos [7]. Por otro lado la reconstrucción tridimensional y fotogrametría captura imágenes del entorno y reconstruye a cada uno de los puntos presentes en las imágenes de 2 dimensiones a 3 dimensiones, también mide las dimensiones de los objetos presentes en la imagen [8]. Por su parte el “tracking” se encarga de detectar objetos de la manera más veloz posible dentro de videos, aunque sabemos que un video está constituido por una secuencia de imágenes tomadas a intervalos de tiempos comúnmente constantes [9][10]. El reconocimiento de objetos tridimensionales es un área que esta comenzando a surgir debido a las necesidades de reconocimiento sofisticadas, pues busca ir más allá del reconocimiento de patrones en 2

dimensiones y combina a las otras 3 disciplinas para lograr identificar a objetos en 3 dimensiones, que en 2 dimensiones no podrían identificarse diferencias.

La visión por computadora ha cobrado mucha importancia en el Mundo Tecnológico y Científico en los últimos años, debido a la gran cantidad de aplicaciones y preguntas pendientes por resolver, y por supuesto; también por la gran cantidad de soluciones a problemas de automatización que se han logrado resolver a través de sus técnicas [10][11][12]. Es por esta razón la visión por computadora es estudiada por muchos profesionales de diferentes campos del conocimiento. En la Tabla 1., se enlistan algunas de las principales aplicaciones de esta disciplina en diferentes campos.

Campo de aplicación	Ejemplos
Robótica	Construcción de mapas de navegación, medición de distancias, evasión de obstáculos, etc.
Medicina	Identificación de enfermedades en la piel, análisis de la sangre.
Astronomía	Identificación de estrellas, aproximación del tamaño de objetos celestes.
Control de calidad	Control de calidad de piezas mecánicas (medidas).
Cartografía	Reconstrucción en 3D de terrenos.
Seguridad	Identificación de rostros, identificación de billetes falsos.

Tabla 1. Principales aplicaciones de la visión por computadora.

2.2 Cámaras digitales

La tecnología actual utilizada para desarrollar visión por computadora esta basada en las cámaras digitales, debido a que éstas tienen las propiedades de una fuente de información digital: como relación señal a ruido constante y conocida, así como una infraestructura tecnológica avanzada [13][14]. Este sistema de obtención de imágenes es el que se utiliza en robots y también el que se pretende usar en la presente tesis debido a que trabajamos en sistemas mecatrónicos. Estos capturan videos que se traducen en frames o fotografías para

posteriormente analizarlas con un microprocesador o microcontrolador para ayudar en la toma de decisiones sobre rutas a seguir, elementos a perforar, pintar, martillar y otras tareas solicitadas a un sistema robótico. Una fotografía le sirve para observar el entorno y analizarlo a través de un programa de cómputo [15]. De tal manera que este puede decidir los movimientos a realizar para evadir obstáculos, mover objetos, construir mapas, reconocer objetos, etc.

En la Figura 4. Se presenta la imagen de una habitación; que un sistema de fotografía digital presentaría a un sistema computacional de análisis de imágenes para un robot, para navegar por la habitación sin problemas de obstáculos. Al tratar de moverse dentro de esta habitación el sistema necesita conocer las dimensiones de objetos y del lugar dónde se encuentra, la posición y características dimensionales de los obstáculos (una cama, una computadora, un enchufe de energía eléctrica, etc.).



Figura 4. Imagen digital de una habitación tomada por una cámara digital para ser analizada.

2.3 El dispositivo CCD como elemento básico en la captura de imágenes

Para saber cómo trabajar con imágenes que provienen de una cámara digital es necesario conocer la técnica utilizada en las fotografías digitales utilizadas. Este proceso, se lleva a cabo por medio de un dispositivo llamado CCD (del inglés *Charge-Coupled Device*, "dispositivo de cargas (eléctricas) interconectadas"). El dispositivo CCD es un circuito integrado que contiene un número determinado de condensadores acoplados en forma de matriz. Dichos

condensadores son sensores de luz de una determinada frecuencia: las frecuencias del color rojo, verde y azul las cuales constituyen el nombrado esquema RGB.

El CCD está compuesto por una matriz de $n \times m$ número de sensores de luz que translucen la energía luminosa a movimiento de electrones para provocar una corriente eléctrica. En la Figura 5., se muestra la configuración de un sistema CCD: la matriz de sensores conectados a un circuito acondicionador de la señal y después a un circuito integrado que muestrea (lee la señal que proviene de los sensores) cada cierto tiempo T la señal que envía cada sensor de luz del CCD para después almacenar la información en una memoria temporal. Por último la información de la memoria es enviada al software que se encargará de guardar, visualizar ó editar la imagen capturada.

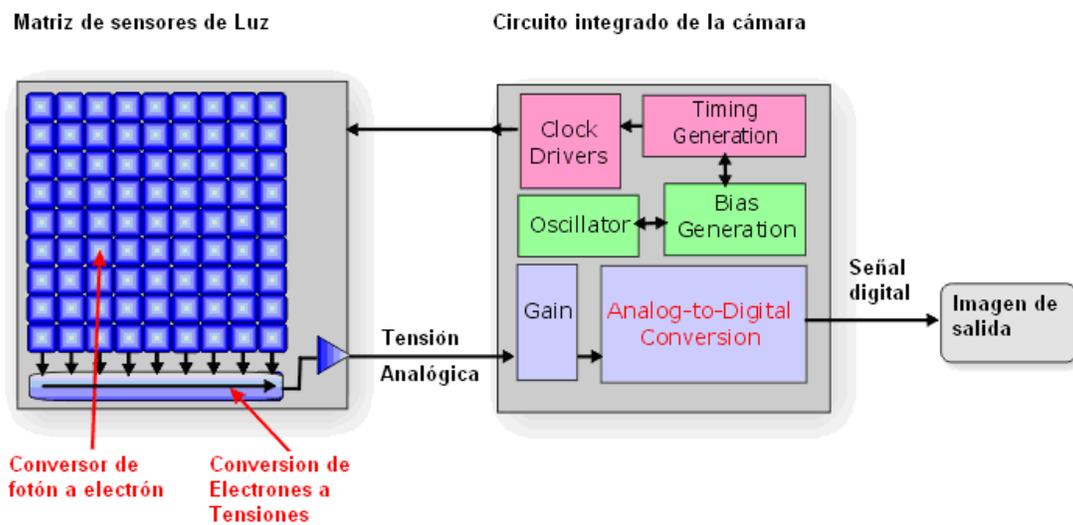


Figura 5. Sistema CCD para capturar imágenes digitales, matriz de sensores y circuito acondicionador de señales.

2.4 La fotografía digital a color como una matriz de 3 dimensiones.

Cuando el CCD entrega a un software una fotografía digital M por N dimensiones, esta viene comúnmente en una matriz tridimensional de $M \times N \times 3$, para el modelo de color RGB[]. Donde M es la cantidad de píxeles a lo ancho de la imagen, N la cantidad a lo largo y 3 debido a que son 3 matrices de $M \times N$. Para un píxel la descripción RGB (del inglés Red, Green, Blue; "rojo, verde, azul") de un color hace referencia a la composición del color en términos de la intensidad de los colores primarios con que se forma: el rojo, el verde y el azul. Es un modelo de color basado en la síntesis aditiva, con el que es posible representar un color mediante la mezcla por adición de los tres colores luz primarios. es frecuente que cada color primario se

codifique con un byte (8 bits). Así, de manera usual, la intensidad de cada una de las componentes se mide según una escala que va del 0 al 255. Por lo tanto, el rojo se obtiene con (255,0,0), el verde con (0,255,0) y el azul con (0,0,255), obteniendo, en cada caso un color resultante monocromático. La ausencia de color —lo que nosotros conocemos como color negro— se obtiene cuando las tres componentes son 0, (0, 0, 0). La combinación de dos colores a nivel 255 con un tercero en nivel 0 da lugar a tres colores intermedios. De esta forma el amarillo es (255,255,0), el cyan (0,255,255) y el magenta (255,0,255). Obviamente, el color blanco se forma con los tres colores primarios a su máximo nivel (255, 255, 255). En la Figura 6., se muestra una imagen RGB como una matriz de $M \times N$ y su descomposición en el modelo RGB [16][17].

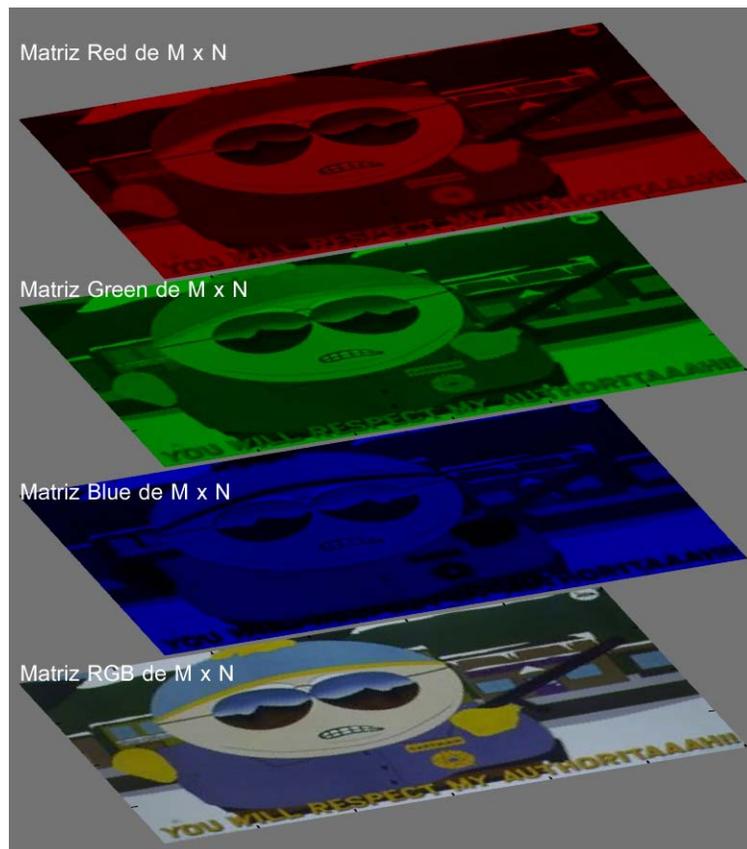


Figura 6. La imagen digital a color del tipo RGB se representa por medio de 3 matrices de dimensiones $M \times N$ donde cada matriz representa la cantidad de color rojo, verde o azul de la imagen.

2.5 Modelo de cámara fotográfica pinhole.

El modelo de cámara PinHole es el más utilizado para modelar la formación de imágenes de una cámara digital, éste no se ocupa en estudiar físicamente el sistema de lentes de la cámara, ni la naturaleza física de los transductores que conforman el CCD. Sin embargo, es muy utilizado debido a las ventajas que proporciona: las ecuaciones no son tan complicadas comparadas con el modelo físico de la cámara y proporciona una precisión suficiente para que se pueda aplicar en la solución de problemas de robótica y otras disciplinas[18].

El modelo PinHole consiste en modelar una cámara como una caja que cuenta con un pequeño orificio en un plano, sólo ese orificio deja pasar rayos de luz provenientes de algún objeto que emite luz (ya sea por refracción, reflexión o por ser una fuente luminosa). En la Figura 7., se ilustra el principio del modelo de cámara PinHole.

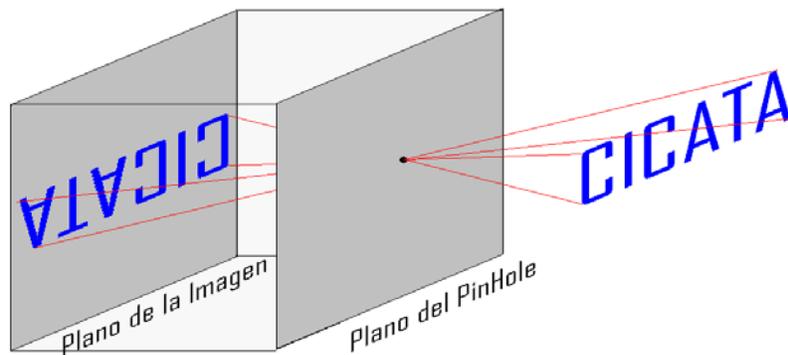


Figura 7. El principio del modelo de cámara PinHole invierte la imagen de los objetos externos en el interior del sistema.

El modelo matemático del modelado de cámaras PinHole se basa en el esquema de la Figura 8. El sistema de referencia O , tiene fijado su origen en el Orificio (PinHole) de la cámara mostrada en la figura 3. Todos los rayos que provienen del exterior de la cámara pasan por la coordenada $(0, 0, 0)$ del sistema de referencia O .

Un punto P de un objeto es proyectado en el plano Π' y es nombrado como P' . El centro del Plano Π' tiene coordenadas C' ; f es la distancia entre el plano XY del sistema de referencia O y el plano Π' ; también llamada distancia focal.

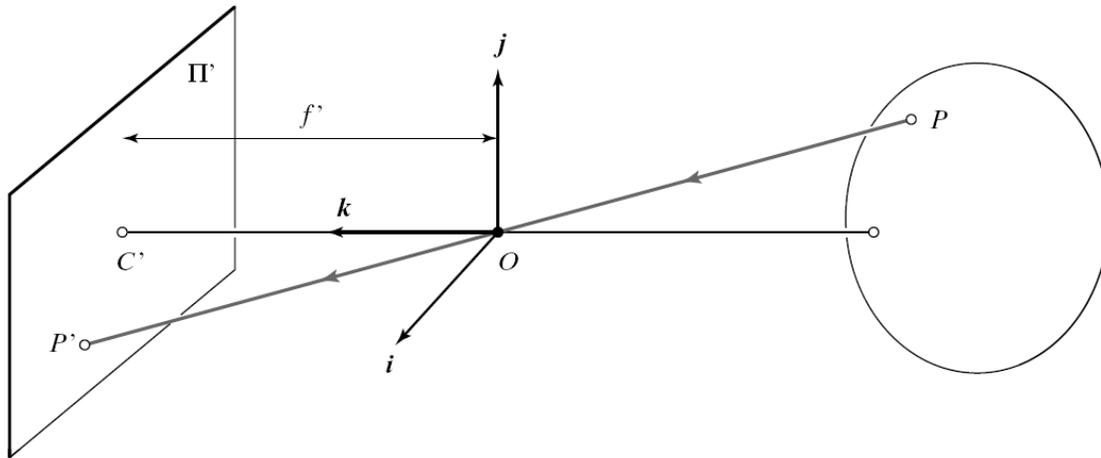


Figura 8. Sistema de referencia del modelo PinHole montado en el plano donde se encuentra el orificio (PinHole) del sistema.

Este tipo de modelado es llamado proyección en perspectiva, debido a que el punto P del objeto se proyecta en el plano Π' solo con un escalamiento λ que depende directamente de la distancia focal f' y de la distancia entre el objeto y el plano del PinHole descrito en el conjunto de igualdades 1.1.

Considerando que el punto P del objeto tiene coordenadas (x, y, z) y el punto P' tiene coordenadas (x', y', z') , también que el rayo de luz PO y OP' son colineales.

$$\begin{cases} x' = \lambda x \\ y' = \lambda y \\ z' = \lambda z \end{cases} \quad \text{debido a que } f' \text{ tiene la misma dirección de } \mathbf{k} \quad \lambda = \frac{x'}{x} = \frac{y'}{y} = \frac{f'}{z}$$

Entonces

$$\begin{cases} x' = f' \frac{x}{z} \\ y' = f' \frac{y}{z} \end{cases} \quad (1)$$

2.6 Transformación lineal de coordenadas del espacio a coordenadas de una fotografía

Cuando vemos una fotografía, nosotros como seres humanos, podemos interpretar algunas distancias en 3 dimensiones de la foto. Por ejemplo en la fotografía de la Figura 9., podemos saber que el muro se encuentra a 6 cm. aproximadamente del pie derecho de Zapata (la persona del sombrero y escopeta en la fotografía). Eso lo sabemos por que la foto fue tomada de un mundo en 3 dimensiones (3D), pero ahora se encuentra en 2 dimensiones (2D).



Figura 9. Una fotografía que muestra personas en distintas posiciones del espacio.

La distancia que estimamos entre el pie derecho de Zapata y el muro próximo (6 cm.), se explica gracias a que hicimos una correspondencia de puntos en el espacio con los puntos de la fotografía, así como también debido a que tenemos la referencia de que los pies de una persona, generalmente tienen un ancho de 5 a 6 cm.

A la correspondencia entre un punto del espacio y el mismo punto de la fotografía se le puede modelar con una transformación lineal. Existe una transformación rígida de 3D a 2D, que es muy utilizada para analizar este fenómeno propuesta por Zhang [13], la transformación propuesta por este autor es la que se muestra continuación (2).

$$\lambda \tilde{m} = A[R \ t]\tilde{M} \quad (2)$$

donde:

$\tilde{M} = [x, y, z, 1]$: es el punto en coordenadas Homogéneas en 3D

$\tilde{m} = [u, v, 1]$: es el punto en coordenadas Homogeneas en 2D

λ : es un escalar

$[R \ t]$: es la matriz de parametros extrínsecos

A : es la matriz de parámetros Intrínsecos

$$A = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

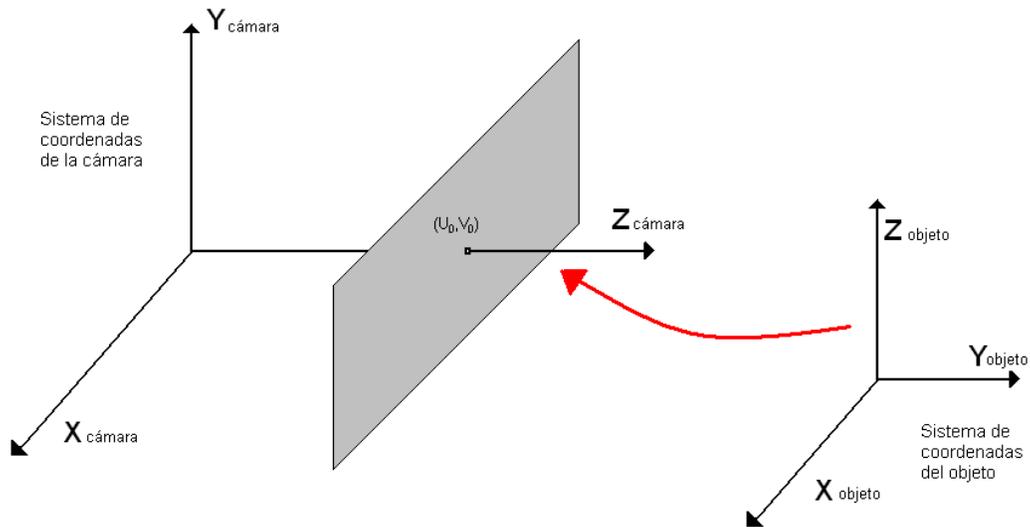


Figura 10. La transformación Lineal modela correspondencias entre el sistema de coordenadas del objeto en el espacio y sus correspondencias en el sistema de coordenadas de la cámara.

El punto en coordenadas homogéneas \tilde{M} identifica un punto 3D del objeto en el espacio este punto corresponde al punto \tilde{m} de la fotografía, en coordenadas de la cámara (2D), como se

muestra en la Figura 10. La ecuación (2) modela esta correspondencia. Esta ecuación utiliza 2 matrices muy importantes; la matriz **A** llamada matriz de parámetros propios de la cámara y la matriz **[R t]** de parámetros externos a la cámara, específicamente la posición de la cámara con respecto al objeto.

2.7 Parámetros extrínsecos

[R t] es la matriz de parámetros extrínsecos ó matriz de parámetros externos a la cámara. **R** es la matriz de rotación que alinea en la misma dirección y sentido los ejes **X**, **Y** y **Z**, de los dos sistemas de coordenadas. El vector **t** de la ecuación (4) es el vector de traslación que traslada el origen del sistema de coordenadas del objeto en el espacio al sistema de coordenadas de la cámara.

$$t = [t_x, t_y, t_z]^T \quad (4)$$

La rotación de ejes se da gracias a 3 matrices de rotación, la rotación alrededor del eje X que esta especificada con un ángulo α , la rotación alrededor del eje Y especificada por el ángulo β y la rotación alrededor del eje Z por θ , tal y como se muestra en la Figura 11.

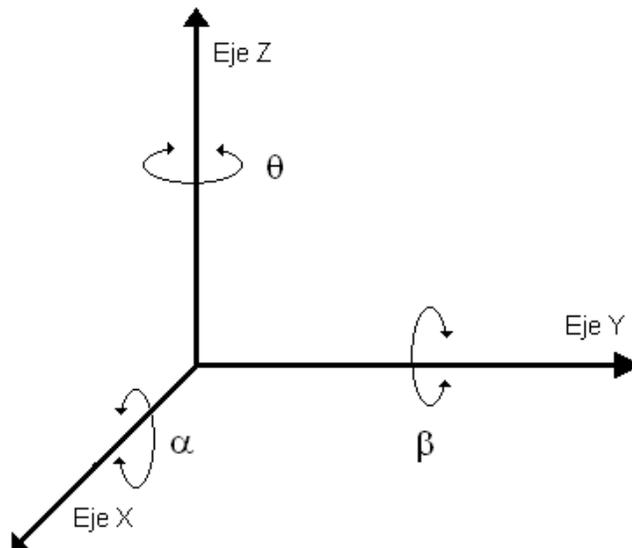


Figura 11. Rotación (α, β, θ) de un punto alrededor de los ejes X, Y, Z respectivamente.

Las matrices de rotación alrededor de cada uno de los ejes del sistema de coordenadas X, Y y Z se muestran a continuación respectivamente.

$$R_{\alpha} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \operatorname{sen} \alpha \\ 0 & -\operatorname{sen} \alpha & \cos \alpha \end{bmatrix} \quad R_{\beta} = \begin{bmatrix} \cos \beta & 0 & -\operatorname{sen} \beta \\ 0 & 1 & 0 \\ \operatorname{sen} \beta & 0 & \cos \beta \end{bmatrix} \quad R_{\theta} = \begin{bmatrix} \cos \theta & \operatorname{sen} \theta & 0 \\ \operatorname{sen} \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

En general al aplicarle a un punto una rotación alrededor de los 3 ejes X, Y y Z con ángulos α , β y θ respectivamente, la matriz de rotación resultante, es la que se muestra a continuación:

$$R = R_{\alpha} R_{\beta} R_{\theta} = \begin{bmatrix} \cos \beta \cos \theta & \operatorname{sen} \alpha \operatorname{sen} \beta \cos \theta + \cos \alpha \operatorname{sen} \theta & -\cos \alpha \operatorname{sen} \beta \cos \theta + \operatorname{sen} \alpha \operatorname{sen} \theta \\ -\cos \beta \operatorname{sen} \theta & \operatorname{sen} \alpha \operatorname{sen} \beta \operatorname{sen} \theta + \cos \alpha \cos \theta & \cos \alpha \operatorname{sen} \beta \operatorname{sen} \theta + \operatorname{sen} \alpha \cos \theta \\ \operatorname{sen} \beta & -\operatorname{sen} \alpha \cos \beta & \cos \alpha \cos \beta \end{bmatrix} \quad (6)$$

Para observar la matriz de rotación de forma simplificada generalmente la encontramos en la bibliografía con la siguiente notación:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad \text{entonces} \quad [Rt] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \quad (7)$$

2.8 Parámetros intrínsecos

Los parámetros intrínsecos están expresados en la matriz A de la ecuación (2). Resulta conveniente aclarar qué significa cada uno de estos: una cámara digital está constituida por

una matriz de sensores de luz que tienen formas específicas generalmente cuadradas y que afectan en la manera en que se forman las imágenes en fotografías, éstos se representan por píxeles. Los coeficientes f_x y f_y nos proporcionan información acerca de las dimensiones del píxel. El parámetro s es llamado *skew* y representa la inclinación que tiene el píxel, generalmente la inclinación se considera muy pequeña ó cero (en este trabajo se considerará como 0), esta representación se ilustra en la Figura 12.

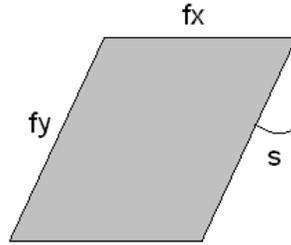


Figura 12. La geometría del píxel influye en la manera de modelar las imágenes digitales, las dimensiones de f_x y f_y ; así como el ángulo s , también llamado sesgo.

Entonces para encontrar el punto de la fotografía que corresponde con el punto del espacio fotografiado se utiliza la ecuación (2) pero ésta puede ser simplificada para utilizar sólo una matriz en lugar de utilizar dos. Este paso se muestra en la ecuación (8).

$$\begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (8)$$

Al realizar el producto matricial de la matriz de parámetros intrínsecos con la matriz de parámetros extrínsecos obtenemos:

$$\begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (9)$$

2.9 Métodos de calibración

2.9.1 Método de Fougeras

La calibración de una cámara por este método consiste en conocer los parámetros intrínsecos (4 parámetros si se considera el skew como cero) y los parámetros extrínsecos de la cámara (6 parámetros; las rotaciones y desplazamientos respecto a un sistema de referencia). Partiendo de la ecuación (9) podemos observar que al dar un punto $[X, Y, Z, 1]^T$ del espacio podemos mapearlo en el plano de la cámara con coordenadas $(\lambda u, \lambda v, \lambda)$; si consideramos a λ sólo para tomarlo en cuenta después; entonces obtenemos el siguiente conjunto de ecuaciones:

$$u = \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \quad (10)$$

$$v = \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \quad (11)$$

Estas expresiones se utilizan en el caso de que se conoce un punto en el espacio y su correspondencia en la imagen. La finalidad de escribir las ecuaciones (10) y (11) es encontrar los valores de m_{ij} ; que es el proceso conocido como calibración de una cámara.

Por cada punto expresado en las ecuaciones (10) y (11) contamos con 2 ecuaciones; por lo tanto para obtener el valor de los 12 términos m_{ij} es necesario tener por lo menos 6 correspondencias de puntos y resolver el sistema de ecuaciones. Las ecuaciones (10) y (11) quedan rescritas en (12) y (13) para el i -ésimo conjunto de correspondencias de puntos.

$$m_{11}X_i + m_{12}Y_i + m_{13}Z_i + m_{14} - (m_{31}u_iX_i + m_{32}u_iY_i + m_{33}u_iZ_i) = u_i m_{34} \quad (12)$$

$$m_{21}X_i + m_{22}Y_i + m_{23}Z_i + m_{24} - (m_{31}v_iX_i + m_{32}v_iY_i + m_{33}v_iZ_i) = v_i m_{34} \quad (13)$$

Podemos re-escribirlo también en forma matricial:

$$\begin{bmatrix}
& & & & : & : & : \\
X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -u_i X_i & -u_i Y_i & -u_i Z \\
0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -v_i X_i & -v_i Y_i & -v_i Z \\
& & & & : & : & :
\end{bmatrix}
\begin{bmatrix}
m_{11} \\
m_{12} \\
m_{13} \\
m_{14} \\
m_{21} \\
m_{22} \\
m_{23} \\
m_{24} \\
m_{31} \\
m_{32} \\
m_{33}
\end{bmatrix}
=
\begin{bmatrix}
: \\
u_i m_{34} \\
v_i m_{34} \\
:
\end{bmatrix}
\quad (14)$$

Fougeras propone resolver (14) haciendo $m_{34}=1$, ya que m_{34} será no nula, debido a que el sistema que estudiamos no estudia objetos que se encuentran en el lente de una cámara y m_{34} es simplemente la componente Z del vector de traslación entre el sistema coordenado del mundo exterior y el sistema de referencia de la cámara. La desventaja con este método es que el resultado queda escalado y es una incógnita. El método se describe a continuación:

Si $m_{34}=1$,

$$\begin{bmatrix}
& & & & : & : & : \\
X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -u_i X_i & -u_i Y_i & -u_i Z \\
0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -v_i X_i & -v_i Y_i & -v_i Z \\
& & & & : & : & :
\end{bmatrix}
\begin{bmatrix}
m_{11} \\
m_{12} \\
m_{13} \\
m_{14} \\
m_{21} \\
m_{22} \\
m_{23} \\
m_{24} \\
m_{31} \\
m_{32} \\
m_{33}
\end{bmatrix}
=
\begin{bmatrix}
: \\
u_i \\
v_i \\
:
\end{bmatrix}
\quad (15)$$

Esto lo podemos resolver por el método de mínimos cuadrados.

$$KM = U \quad (16)$$

$$M = (K^T K)^{-1} U \quad (17)$$

2.9.2 Método de Zhang

Zhang [13] propone una técnica de calibración basada en la observación de una plantilla estrictamente plana (tablero de juego de ajedrez) desde varias posiciones. La ventaja de este método de calibración es que permite obtener los parámetros de la cámara fácilmente a partir de una sistema de referencia expresado en un plano cuadrículado plano, en la cual no es necesario conocer la posiciones de los puntos de interés, ni tampoco es necesario conocer las posiciones de la cámara desde donde se han tomado las imágenes de la misma. La cámara se puede mover simplemente con la mano. Esto hace que sea una técnica muy flexible ya que los métodos descritos hasta ahora necesitan de una preparación exhaustiva de la escena ya que es necesario que los puntos no formen un plano y también es necesario conocer las posiciones de los mismos.



Figura 13. Aquí se muestra el patrón para calibrar en el método de Zhang, con diferentes orientaciones y traslaciones para una mejor aproximación.

El modelo de la cámara que utiliza es:

$$s\tilde{m} = A[R \ t]\tilde{M} \quad (18)$$

Donde:

$\tilde{M} = [x, y, z, 1]$: es el punto en coordenadas Homogéneas en 3D

$\tilde{m} = [u, v, 1]$: es el punto en coordenadas Homogeneas en 2D

s : es un escalar

$[R \ t]$: es la matriz de parametros extrínsecos

A : es la matriz de parámetros Intrínsecos

Que prácticamente expresa lo mismo que la ecuación (2). Sin perder generalidad se asume que el modelo plano esta sobre $Z=0$ en el sistema coordenado del mundo exterior a la cámara. Si denotamos la columna i -ésima de la matriz de rotación por r_i , obtenemos que:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A \begin{bmatrix} r_1 & r_2 & r_3 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = A \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = H \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (19)$$

Donde: $H = A \begin{bmatrix} r_1 & r_2 & t \end{bmatrix}$ es llamada homografía.

Entonces (19) se rescribe para $Z=0$ como:

$$s\tilde{m} = H\tilde{M} \quad (20)$$

$$\begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix} = A \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \quad (21)$$

Dado que R es una matriz de rotación, los vectores que la componen cumplen las restricciones de ortonormalidad, es decir, $r_1^T \cdot r_2 = 0$ y que $r_1^T \cdot r_1 = r_2^T \cdot r_2$, por lo tanto si se extraen los vectores de rotación a partir de la última expresión se tiene que:

$$h_1^T A^{-T} A^{-1} h_2 = 0 \quad (22)$$

$$h_1^T A^{-T} A^{-1} h_1 = h_2^T A^{-T} A^{-1} h_2 \quad (23)$$

Estas son las dos restricciones básicas de los parámetros intrínsecos dada una homografía. Para resolver este problema se propone una solución analítica seguida de una optimización no lineal. La matriz $A^T A^{-1}$ está compuesta por los parámetros intrínsecos de la cámara de la siguiente forma:

$$B = A^{-T} A^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} = \begin{bmatrix} \frac{1}{\alpha^2} & -\frac{c}{\alpha^2 \beta} & \frac{c\nu_0 - \nu_0 \beta}{\alpha^2 \beta} \\ -\frac{c}{\alpha^2 \beta} & \frac{c^2}{\alpha^2 \beta^2} + \frac{1}{\beta^2} & -\frac{c(c\nu_0 - \nu_0 \beta)}{\alpha^2 \beta^2} - \frac{\nu_0}{\beta^2} \\ \frac{c\nu_0 - \nu_0 \beta}{\alpha^2 \beta} & -\frac{c(c\nu_0 - \nu_0 \beta)}{\alpha^2 \beta^2} - \frac{\nu_0}{\beta^2} & \frac{(c\nu_0 - \nu_0 \beta)^2}{\alpha^2 \beta^2} + \frac{\nu_0}{\beta^2} + 1 \end{bmatrix} \quad (24)$$

Teniendo en cuenta que se trata de una matriz simétrica se puede definir con un vector de 6 elementos de la forma:

$$b = [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33}]^T \quad (25)$$

Por lo tanto si la columna i-esima de la matriz H es $h_i = [h_{i1} \ h_{i2} \ h_{i3}]^T$, se tiene que:

$$h_i^T B h_i = v_{ij}^T b \quad (26)$$

Donde:

$$v_{ij}^T = [h_{i1} h_{j1} \quad h_{i1} h_{j2} + h_{i2} h_{j1} \quad h_{i2} h_{j2} \quad h_{i3} h_{j1} + h_{i1} h_{j3} \quad h_{i3} h_{j2} + h_{i2} h_{j3} \quad h_{i3} h_{j3}] \quad (27)$$

Con estas expresiones se pueden escribir las restricciones de los parámetros intrínsecos en dos ecuaciones homogéneas en función de b de la siguiente forma:

$$\begin{bmatrix} v_{12}^T \\ (v_{11} - v_{22})^T \end{bmatrix} b = Vb = 0 \quad (28)$$

Donde, V es una matriz de dimensiones 2×6 . Si $n \geq 3$ se tiene una solución general con solución única de b definida con un factor de escala. Si $n=2$ se puede imponer la restricción $c = 0$ la cual se añade a las ecuaciones anteriores. La solución es el vector propio de $V^T V = 0$ asociado al valor propio más pequeño. Cuando se ha estimado b se pueden obtener los parámetros intrínsecos de la cámara que forman la matriz A según:

$$\nu_0 = (B_{12} B_{13} - B_{11} B_{23}) / (B_{11} B_{22} - B_{12}^2) \quad (29)$$

$$\lambda = B_{33} - [B_{13}^2 + v_0(B_{12}B_{13} - B_{11}B_{23})] / B_{11} \quad (30)$$

$$\alpha = \sqrt{\lambda / B_{11}} \quad (31)$$

$$\beta = \sqrt{\lambda B_{11} / (B_{11}B_{22} - B_{12}^2)} \quad (32)$$

$$c = -B_{12}\alpha^2\beta / \lambda \quad (33)$$

$$u_0 = cv_0 / \alpha - B_{13}\alpha^2 / \lambda \quad (34)$$

Ya que **A** es conocida los parámetros extrínsecos para cada imagen son fácilmente calculados como:

$$r_1 = \lambda A^{-1}h_1, \quad r_2 = \lambda A^{-1}h_2, \quad r_3 = r_1 \times r_2, \quad t = \lambda A^{-1}h_3, \quad \lambda = 1 / \|A^{-1}h_1\| = 1 / \|A^{-1}h_2\| \quad (35)$$

La distorsión radial es comúnmente la modelada por un polinomio, en este caso se utiliza un polinomio con dos términos, sean (u,v) las coordenadas ideales del píxel mientras (\tilde{u}, \tilde{v}) sus correspondientes coordenadas reales observadas en la imagen. De manera similar, (x,y) con (\tilde{x}, \tilde{y}) , son las coordenadas ideales y reales normalizadas en la imagen.

$$\tilde{x} = x + x[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2] \quad (36)$$

$$\tilde{y} = y + y[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2] \quad (37)$$

Donde k_1 y k_2 son los coeficientes de la distorsión radial, el centro de la distorsión radial se considera como el punto principal de nuestra cámara. $\tilde{u} = u_0 + \alpha\tilde{x} + c\tilde{y}$, $\tilde{v} = v_0 + \alpha\tilde{y}$ entonces tenemos:

$$\tilde{u} = u + (u - u_0) \left[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2 \right] \quad (38)$$

$$\tilde{v} = v + (v - v_0) \left[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2 \right] \quad (39)$$

2.10 Reconocimiento de patrones

El crear la implantación en una maquina de habilidades como la de detectar y determinar la identidad de los objetos que se encuentran a su alrededor continua siendo uno de los retos mas importantes en el mundo de la visión por computadora. Maquina como estas pudieran liberar al hombre de tareas peligrosas y en otros casos con poca necesidad de la presencia de una persona para realizarlas, así como realizar tareas imposibles de realizar por el ser humano, como tareas a altas temperaturas, con fuerza mayor a la del brazo humano. De forma general un sistema para el reconocimiento automatizado de objetos (SRAO) permite a una maquina encontrar (reconocer y posicionar en una imagen) objetos de su entorno a partir de una o más imágenes del mundo real, usando modelos de los objetos conocidos a priori. El problema del reconocimiento automático de objetos (SRAO) puede definirse como un problema de etiquetado que se basa en modelos de objetos conocidos. Formalmente el problema se puede describir de la siguiente forma:

Dada una imagen que contiene uno o más objetos de interés, además del fondo, y un conjunto de etiquetas; una para cada región de la imagen. El sistema debe asignar etiquetas correctas desde el punto de vista de correspondencias respecto a modelos conocidos a regiones o conjuntos de regiones en la imagen.

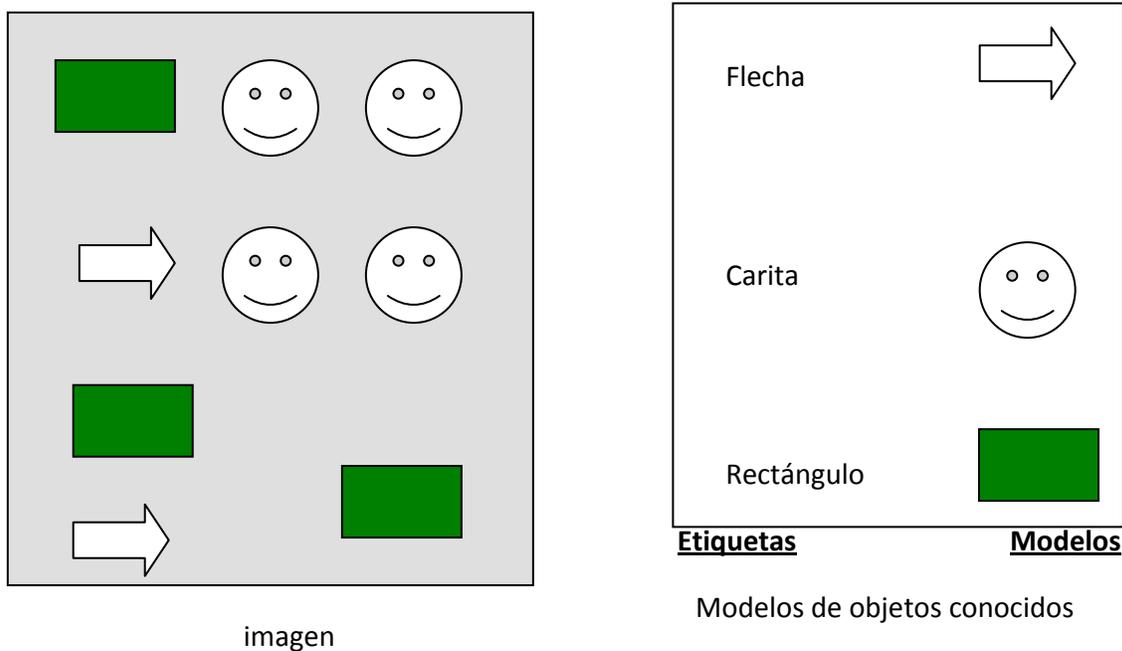


Figura 14. A la izquierda una imagen con algunos de los objetos conocidos que se enumeran a la derecha de la presente figura.

2.11 Componentes de un SRAO

En los últimos años se han propuesto muchas metodologías para el SRAO, pero generalmente se incluyen los siguientes componentes.

- ✦ Un banco de modelos
- ✦ Un módulo de preprocesamiento y acondicionado de la imagen
- ✦ Un módulo de segmentación o aislamiento
- ✦ Un módulo de extracción de rasgos
- ✦ Un módulo generador de hipótesis
- ✦ Un módulo verificador de hipótesis

En la Figura 14., se puede observar el diagrama de bloques donde se muestran las interacciones entre estos módulos típicos así como los flujos de información entre estos.

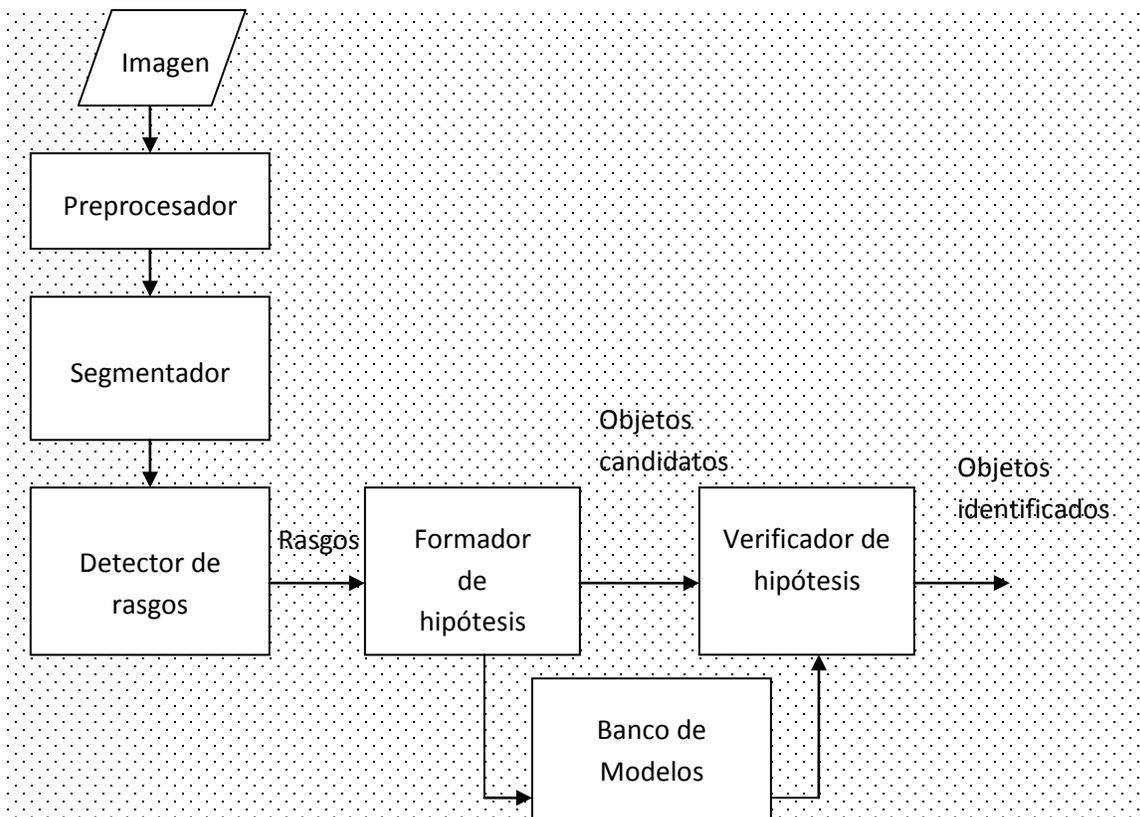


Figura 15. Diagrama a bloques de un SRAO típico con flujos de información.

Preprocesador. Este modulo tiene como objetivo mejorar la calidad de la imagen para tratamientos futuros, debe permitir que etapas posteriores del análisis tengan mejores posibilidades de éxito en su tarea. Usualmente se aplican filtros u otros métodos para reducir el ruido en las imágenes. Este modulo también es usado para reducir el efecto de las transformaciones geométricas sobre la imagen, como las producidas por los lentes.



Figura 16. Para obtener un mejor rango de trabajo se aumenta el contraste de una imagen, mejorándola para procesamiento posterior.

Segmentador. Busca particionar o dividir la imagen en subgrupos de píxeles con la finalidad de que cada uno de estos se aproxime en forma y cantidad de píxeles en lo posible a la región de cada uno de los objetos en la imagen sin identificar si alguno pertenece a nuestros objetos conocidos. El caso mas simple existe cuando el fondo de la imagen y los objetos de interés tienen un contraste muy definido, también cuando los objetos no se traslapan. Debido a este tipo de problemáticas, a veces este modulo no se incluye en SRAO.

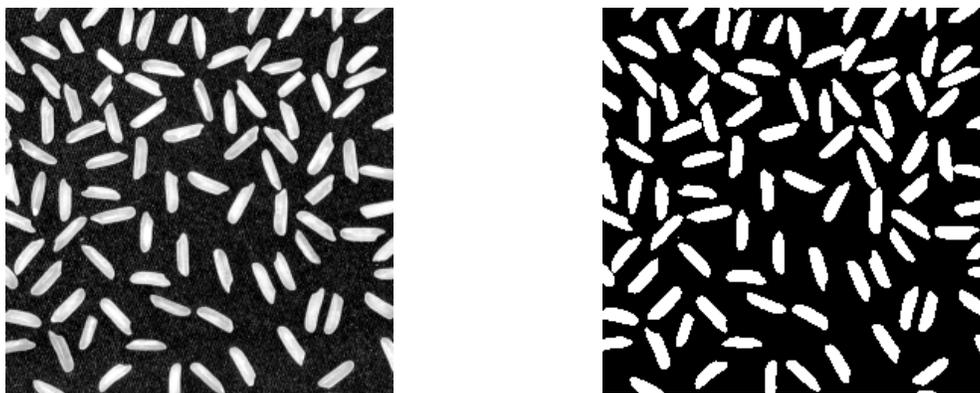


Figura 17. El segmentador se encarga de extraer sólo los objetos que interesan al análisis que se esta desarrollando en cada aplicación.

Extractor de rasgos. En este modulo se realizan operaciones sobre la imagen segmentada o no para detectar la presencia de objetos conocidos en la imagen, estas operaciones se aplican a los subconjuntos formados en la etapa de segmentación si existiera. Los extractores de rasgos u operadores dependen del tipo de objetos a reconocer y de los modelos almacenados para referencia.

Generador de hipótesis. Utiliza los rasgos extraídos de la imagen para asignar una probabilidad de presencia de un objeto de nuestra base de datos de modelos presente en la imagen. Esta etapa es para reducir el espacio de búsqueda del verificador de hipótesis.

Banco de modelos. Almacena todos los modelos de los objetos que el sistema reconocerá, dependiendo de la técnica utilizada, puede ser una descripción funcional o cualitativa general de cada objeto, ó una descripción geométrica precisa del mismo.

Módulo verificador. Compara los rasgos de los objetos extraídos por el generador de hipótesis y los compara con los modelos almacenados en el banco de modelos, de acuerdo a la identificación de las menores diferencias se marcan a los objetos que mayor certidumbre tienen de haber sido identificados, conforme a la regla definida por el diseñador del modulo.

Capítulo

3

“Quien detiene su formación se convierte en anciano, sea a los veinte o a los ochenta. Quien sigue aprendiendo se mantiene siempre joven”. **Henry Ford.**

Estado del Arte



3.1 Calibración estereoscópica

La calibración de una cámara nos permite conocer los parámetros internos de esta como: el skew, el punto principal y la geometría de los píxeles. Pero para manejar las coordenadas de un punto en el espacio, es necesario, utilizar por lo menos 2 cámaras que tomen fotografía del mismo punto en el mismo instante (si lo que se captará en las fotografías esta estático, se puede utilizar la misma cámara para tomar 2 fotografías de diferente punto de vista que incluya el punto de análisis). Esto se debe a que al utilizar una fotografía de la escena de interés, sólo se cuenta con la información proporcionada por el par coordenado (x,y) en dos dimensiones por lo tanto no se puede conocer la profundidad de algún punto en el espacio. Existen algunas técnicas probabilísticas que ayudan a estimar un punto 3D a partir de puntos 2D, pero es conveniente enfatizar que es una **estimación**. Para conocerla es necesario utilizar al menos 2 imágenes del mismo punto desde diferente punto de vista. En la Figura 9., se muestran las configuraciones de cámaras utilizadas en visión estereoscópica; donde se observa que lo importante es conocer la transformación rígida que relaciona la posición entre éstas. Esta relación se expresa simplemente conociendo el vector de posición \vec{r} que va del sistema de coordenadas de una cámara, hasta el sistema de coordenadas de la otra cámara, junto a los ángulos (α, β, θ) de orientación de una cámara respecto a la otra. Obviamente fijando primero a una de las dos como cámara de referencia y el sistema de referencia de la otra se expresará en términos de los parámetros antes mencionados.

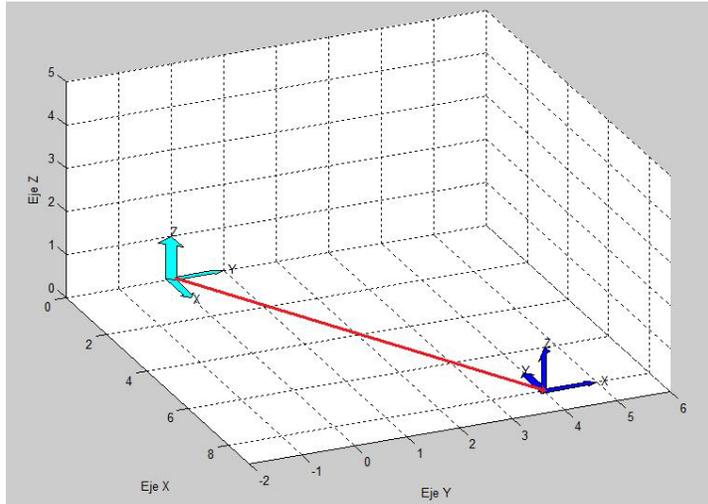


Figura 18. Se muestran dos sistemas de referencia en azul claro y azul oscuro, donde el primero localiza al segundo y los puntos expresados a partir del sistema de referencia azul oscuro pueden expresarse desde el sistema de referencia azul claro.

Las empresas que se dedican a vender equipos de visión estereoscópica especifican en sus equipos la distancia que existe entre los focos de las dos cámaras, la diferencia entre los ángulos de los sistemas de referencia de las cámaras; así como los parámetros intrínsecos de estas.

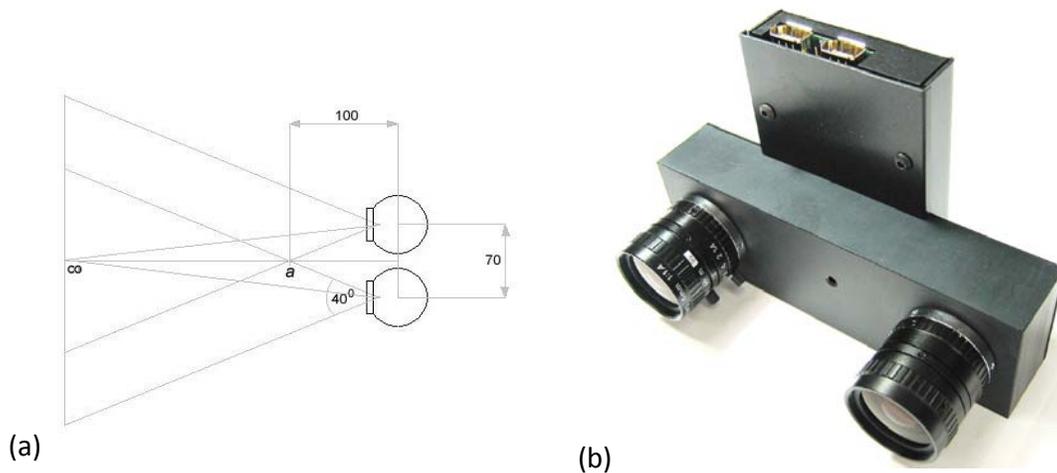


Figura 19. (a) Dos cámaras separadas 70 mm., a una distancia infinita la disparidad estéreo es cero. (b) Cámaras acopladas para uso en visión estereoscópica.

La calibración estereoscópica se conforma de tener 2 tipos de información: Calibración de cada una de las cámaras (particularmente los parámetros intrínsecos), y la transformación rígida (rotación y traslación) que relaciona la posición y orientación entre las 2 cámaras. En la figura 10., se muestran 2 cámaras calibradas para ser usadas de forma estereoscópica, es decir se conocen los parámetros intrínsecos de las cámaras y en la misma figura se muestran los parámetros extrínsecos, separación (vector de traslación) de las cámaras y orientación de estas. Una situación importante que debe existir que en algunos casos se considera obvia, es que el punto a analizar se debe encontrar en las 2 imágenes y saber que es el mismo punto en el espacio. Esto se puede lograr con los estudios realizados en otra de las áreas de la visión por computadora llamada *matching*, donde se resuelve el problema de correspondencias de puntos entre 2 o más imágenes.

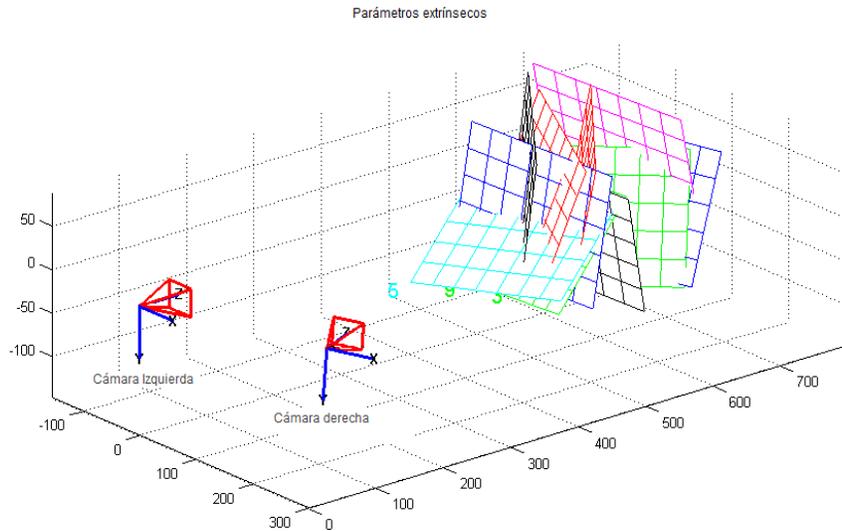


Figura 20. Cámaras calibradas, donde se conoce la transformación rígida del sistema de coordenadas de la cámara izquierda a la cámara derecha.

3.2 Aproximación de un Punto 3D en base a 2 fotografías.

A partir de que conocemos los parámetros intrínsecos y extrínsecos de las 2 cámaras \mathbf{M} (parámetros de la cámara izquierda) y \mathbf{M}' (parámetros de la cámara derecha). Podemos describir a un punto del espacio $P=(X,Y,Z,1)$, que en las 2 imágenes se identifica a este punto como $p'=(u,v)$ y $p''=(u',v')$.

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{12} & m_{12} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \quad M' = \begin{bmatrix} m'_{11} & m'_{12} & m'_{12} & m'_{12} \\ m'_{21} & m'_{22} & m'_{23} & m'_{24} \\ m'_{31} & m'_{32} & m'_{33} & m'_{34} \end{bmatrix} \quad (40)$$

Al multiplicar el punto 3D por la matriz \mathbf{M} y \mathbf{M}' (los parámetros de cada una de las cámaras) obtenemos el conjunto de ecuaciones (41) (42) (43) y (44):

$$u = \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \quad (41)$$

$$v = \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \quad (42)$$

$$u' = \frac{m'_{11}X + m'_{12}Y + m'_{13}Z + m'_{14}}{m'_{31}X + m'_{32}Y + m'_{33}Z + m'_{34}} \quad (43)$$

$$v' = \frac{m'_{21}X + m'_{22}Y + m'_{23}Z + m'_{24}}{m'_{31}X + m'_{32}Y + m'_{33}Z + m'_{34}} \quad (44)$$

Las ecuaciones (41) (42) (43) y (44) pueden ser agrupadas y arregladas para escribirse en forma matricial:

$$(um_{31} - m_{11})X + (um_{32} - m_{12})Y + (um_{33} - m_{13})Z = m_{14} - um_{34} \quad (45)$$

$$(vm_{31} - m_{21})X + (vm_{32} - m_{22})Y + (vm_{33} - m_{23})Z = m_{24} - vm_{34} \quad (46)$$

$$(u'm'_{31} - m'_{11})X + (u'm'_{32} - m'_{12})Y + (u'm'_{33} - m'_{13})Z = m'_{14} - u'm'_{34} \quad (47)$$

$$(v'm'_{31} - m'_{21})X + (v'm'_{32} - m'_{22})Y + (v'm'_{33} - m'_{23})Z = m'_{24} - v'm'_{34} \quad (48)$$

Entonces obtenemos un sistema de ecuaciones de la forma $Ax = b$, que puede ser resuelto si primero multiplicamos ambos lados de la ecuación por A^T :

$$A^T Ax = A^T b \quad (49)$$

Después despejamos el vector solución x :

$$x = (A^T A)^{-1} A^T b \quad (50)$$

Es aquí donde obtenemos las coordenadas de un punto del espacio (3D) a partir de 2 fotografías.

3.3 Reconocimiento de objetos a partir del reconocimiento de patrones.

Al Reconocimiento de patrones llamado también lectura de patrones, identificación de figuras y reconocimiento de formas[1] consiste en el reconocimiento de patrones de señales. Los patrones se obtienen a partir de los procesos de segmentación, extracción de características y descripción dónde cada objeto queda representado por una colección de descriptores. El sistema de reconocimiento debe asignar a cada objeto su categoría o clase (conjunto de entidades que comparten alguna característica que las diferencia del resto). Para poder reconocer los patrones se siguen los siguientes procesos como se muestra en la figura:

1. Adquisición de datos
2. Extracción de características
3. Toma de decisiones

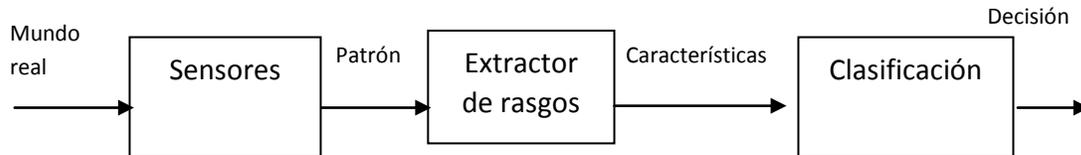


Figura 21. Diagrama a bloques de un sistema de reconocimiento de patrones en general donde el sensor puede ser de audio, imágenes, señales electromagnéticas y otros.

El punto esencial del reconocimiento de patrones es la clasificación: se quiere clasificar una señal dependiendo de sus características. Señales, características y clases pueden ser de cualquiera forma, por ejemplo se puede clasificar imágenes digitales de letras en las clases «W» a «Y» dependiendo de sus píxeles o se puede clasificar ruidos de cantos de los niños en clases de niños ó niñas dependiendo de las frecuencias. Esencialmente existen 5 esquemas para el reconocimiento de patrones que después también se utilizan para el reconocimiento de imágenes en mayor o menor grado dependiendo de la aplicación:

Geométrico (Clustering): Los patrones deben ser graficables, en éste enfoque se emplea el cálculo de distancias, geometría de formas, vectores numéricos, puntos de atracción, etc.

Estadístico: Se basa en la teoría de la probabilidad y la estadística, utiliza análisis de varianzas, covarianzas, dispersión, distribución, etc.

Sintáctico-Estructural: estudia la estructura de los objetos, es decir, usa teoría de lenguajes formales, gramáticas, teoría de autómatas, etc.

Neuro-Reticular: Se utilizan redes neuronales que se ‘entrenan’ para dar una cierta respuesta ante determinados valores.

Lógico-Combinatorio: se basa en la idea de que la modelación del problema debe ser lo más cercana posible a la realidad del mismo, sin hacer suposiciones que no estén fundamentadas. Se utiliza para conjuntos difusos y utiliza lógica simbólica, circuitos combinatoriales y secuenciales, etc.

3.4 Reconocimiento de objetos en 2D por componentes principales.

Un problema central en el análisis de datos multivariantes es la reducción de la dimensionalidad: si es posible describir con precisión los valores de p variables por un pequeño subconjunto $r < p$ de ellas, se habrá reducido la dimensión del problema a costa de una pequeña pérdida de información. El análisis de componentes principales tiene este objetivo: dadas n observaciones de p variables, se analiza si es posible representar adecuadamente esta información con un número menor de variables construidas como combinaciones lineales de las originales. Por ejemplo, con variables con alta dependencia es frecuente que un pequeño número de nuevas variables (menos del 20 por 100 de las originales) expliquen la mayor parte (más del 80 por 100 de la variabilidad original).

La técnica de componentes principales es debida a [19], aunque sus orígenes se encuentran en los ajustes ortogonales por mínimos cuadrados introducidos por K. Pearson [20]. Su utilidad es doble:

1. Permite representar óptimamente en un espacio de dimensión pequeña, observaciones de un espacio general p -dimensional. En este sentido, componentes principales es el primer paso para identificar las posibles variables latentes, o no observadas que generan los datos.
2. Permite transformar las variables originales, en general correladas, en nuevas variables incorreladas, facilitando la interpretación de los datos.

Supongamos que se dispone de los valores de p -variables en n elementos de una población dispuestos en una matriz X de dimensiones $n \times p$, donde las columnas contienen las variables y las filas los elementos. Supondremos que previamente hemos restado a cada variable su media, de manera que las variables de la matriz X tienen media cero y su matriz de covarianzas vendrá dada por $\frac{1}{n} X^T X$.

El problema que se desea resolver es encontrar un espacio de dimensión más reducida que represente adecuadamente los datos. Puede abordarse desde tres perspectivas equivalentes.

a) Enfoque descriptivo

Se desea encontrar un subespacio de dimensión menor que p tal que al proyectar sobre él los puntos conserven su estructura con la menor distorsión posible. Veamos como convertir esta noción intuitiva en un criterio matemático operativo. Consideremos primero un subespacio de dimensión uno, una recta. Se desea que las proyecciones de los puntos sobre esta recta mantengan, lo más posible, sus posiciones relativas como se muestra en la figura 22.

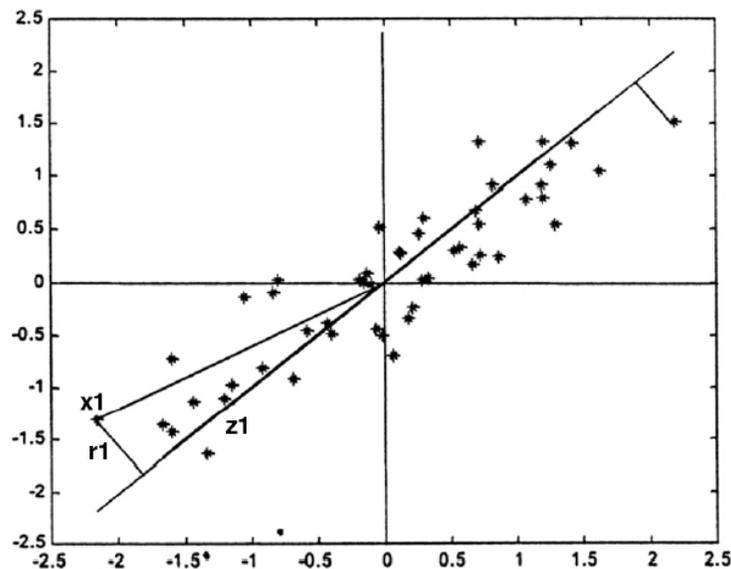


Figura 22. Ejemplo de la recta que minimiza las distancias ortogonales de los puntos a Ella.

Para concretar, consideremos el caso de dos dimensiones ($p = 2$). La Figura 1., indica el diagrama de dispersión y una recta que, intuitivamente, proporciona un buen resumen de los datos, ya que la recta pasa cerca de todos los puntos y las distancias entre ellos se mantienen aproximadamente en su proyección sobre la recta. La condición de que la recta pase cerca de la mayoría de los puntos puede concretarse exigiendo que las distancias entre los puntos originales y sus proyecciones sobre la recta sean lo más pequeñas posibles. En consecuencia, si consideramos un punto x_i y una dirección $\vec{a}_1 = (a_{11}, \dots, a_{1p})^T$, definida por un vector \vec{a}_1 de norma unidad, la proyección del punto x_i sobre esta dirección es el escalar:

$$z_i = a_{11}x_{i1} + \dots + a_{1p}x_{ip} = \vec{a}_1^T \vec{x}_i \quad (51)$$

y el vector que representa esta proyección será $z_i \vec{a}_1$. Llamando r_i a la distancia entre el punto x_i , y su proyección sobre la dirección \vec{a}_1 , este criterio implica:

$$\text{Minimizar} \quad \sum_{i=1}^n r_i^2 = \sum_{i=1}^n |\vec{x}_i - z_i \vec{a}_1|^2 \quad (52)$$

Donde $|\vec{u}|$ es la norma euclídea o módulo del vector \vec{u} .

La Figura 22 muestra que al proyectar cada punto sobre la recta se forma un triángulo rectángulo donde la hipotenusa es la distancia del punto al origen, $(\vec{x}_i^T \vec{x}_i)^{1/2}$, y los catetos la proyección del punto sobre la recta (z_i) y la distancia entre el punto y su proyección (r_i). Por el teorema de Pitágoras, podemos escribir:

$$\vec{x}_i^T \vec{x}_i = z_i^2 + r_i^2, \quad (53)$$

y sumando esta expresión para todos los puntos, se obtiene:

$$\sum_{i=1}^n \vec{x}_i^T \vec{x}_i = \sum_{i=1}^n z_i^2 + \sum_{i=1}^n r_i^2 \quad (54)$$

Como el primer miembro es constante, minimizar $\sum_{i=1}^n r_i^2$, la suma de las distancias a la recta

de todos los puntos, es equivalente a maximizar $\sum_{i=1}^n z_i^2$, la suma al cuadrado de los valores de

las proyecciones. Como las proyecciones z_i son variables de media cero, maximizar la suma de sus cuadrados equivale a maximizar su varianza, y obtenemos el criterio de encontrar la dirección de proyección que maximice la varianza de los datos proyectados. Este resultado es intuitivo: la recta de la Figura 22., parece adecuada porque conserva lo más posible la variabilidad original de los puntos. El lector puede convencerse considerando una dirección de

proyección perpendicular a la de la recta en esta figura: los puntos tendrían muy poca variabilidad y perderíamos la información sobre sus distancias en el espacio.

Si en lugar de buscar la dirección que pasa cerca de los puntos buscamos la dirección tal que los puntos proyectados sobre ella conserven lo más posible sus distancias relativas llegamos al mismo criterio. En efecto, si llamamos $d_{ij}^2 = \vec{x}_i^T \vec{x}_j$ a los cuadrados de las distancias originales entre los puntos y a las distancias entre los puntos $\hat{d}_{ij}^2 = (z_i - z_j)^2$ proyectados sobre una recta, deseamos que

$$D = \sum_i \sum_j (d_{ij}^2 - \hat{d}_{ij}^2) \quad (55)$$

sea mínima. Como la suma de las distancias originales es fija, minimizar D requiere maximizar $\sum_i \sum_j \hat{d}_{ij}^2$, las distancias entre los puntos proyectados.

Capítulo

4

“El aprendizaje es un simple apéndice de nosotros mismos; dondequiera que estemos, está también nuestro aprendizaje”. **William Shakespeare**

Reconocimiento de Objetos en 3D



El reconocimiento de objetos en tres dimensiones (3D) se lleva al cabo a partir de imágenes en dos dimensiones (2D), debido a que en las fotografías convencionales (analógicas) y en las cámaras digitales por naturaleza se obtienen de esta manera. El reconocimiento de objetos en fotografías y las imágenes se estima con algún conocimiento a priori de la aplicación en estudio. De este hecho parte la teoría y práctica del reconocimiento de objetos en imágenes 2D y 3D dentro de la visión por computadora. En este trabajo se propone un tipo de reconocimiento de objetos en 3D basados en la visión estereoscópica y en la estadística del objeto a reconocer.

4.1 ¿Por qué hacer reconocimiento de objetos en 3D?

La respuesta simplemente es:

El reconocimiento de objetos en 3D sirve para caracterizar completamente un cuerpo en sus 3 dimensiones de tipo volumétricas (no del tipo temporal o matemática) y diferenciar estas de distintas dimensiones de otros cuerpos.

4.2 ¿Cuál es la aportación de la presente tesis?

El reconocimiento de objetos se realiza a partir de la perspectiva de la visión por computadora estereoscópica, utilizando 2 imágenes tomadas por diferentes puntos de vista y diferentes cámaras digitales considerando como elemento importante: las 3 dimensiones de los objetos de interés. Este sistema se basa en la estadística de las formas, esto lo dota de un grado de generalidad muy alto, ya que los sistemas **Geométrico (Clustering)**, **Sintáctico - Estructural**, **Neuro - Reticular** y **Lógico - Combinatorio** necesitan cambiar su estructura cuando se requiere de cambio de aplicación. La estructura de este sistema puede utilizarse para muchas aplicaciones sin modificar la estructura del sistema y pueden ser perfeccionados para la aplicación que se quiera utilizar. Este los encuentra de manera automática los detalles para

encontrar matemáticamente los rasgos de cada objeto en la imagen, esta tarea no se deja en manos del talento de una persona, por lo tanto se comprueba la versatilidad del sistema.

Actualmente para realizar el reconocimiento de objetos, la mayoría de los métodos sólo información de fotografías desde el punto de vista de un plano, que como se explicó, es información únicamente del tipo de dos dimensiones (2D), la técnica que proponemos aquí determina la tercera dimensión y no se apoya de:

- ✦ Métodos probabilísticos como teoría de procesos bayesianos para estimar el reconocimiento de objetos de 3 dimensiones.
- ✦ La textura de algunos cuerpos, que marcan la tercera dimensión del cuerpo en estudio.
- ✦ Las esquinas de los paralelepípedos.
- ✦ Las sombras de los cuerpos.
- ✦ Las redes neuronales.
- ✦ El problema en estos sistemas es que siempre se debe tener una referencia dimensional dentro de la imagen ó alguna de las tres dimensiones fijada en el sistema.

Aunque para mejorar su desempeño en alguna aplicación podrían usarse. En cuanto a técnicas para determinar la tercera dimensión, también existen sistemas que utilizan la triangulación de rayos láser para determinar las tres dimensiones de los objetos. Que comúnmente llamamos reconstrucción tridimensional de los cuerpos.

Debido a que se visión por computadora ha solucionado muchos problemas en distintas áreas del conocimiento, se continua utilizando mucho en sistemas de reconocimiento de objetos pero se utilizan algunas restricciones para estimar las dimensiones de los objetos, por ejemplo; en un sistema de control de calidad automático por visión, en el que se deben medir las dimensiones de una pieza elaborada en el torno y la fresa, se coloca una cámara a una distancia fija de donde se colocará la pieza a medir para que se revisen posibles defectos del producto terminado.



Figura 23. Sistema de control de calidad automático, trabaja por visión, compara las dimensiones de la pieza a revisar con la pieza modelo.

El sistema antes mencionado tiene la desventaja de la flexibilidad ante cambios de distancia entre los objetos a reconocer y la cámara o cámaras que se encargaran de capturar las imágenes. El sistema (estructura) que proponemos no necesita de una distancia fija.

Se ha reportado el funcionamiento de la propuesta de reconocimiento de objetos 3D de esta tesis. En los cuales podemos observar que se logra el reconocimiento de objetos similares donde la diferencia entre los objetos radica en las dimensiones volumétricas de los cuerpos, concretamente en las imágenes se presentan esferas de diferentes tamaños, del mismo color (blancas) y el mismo material (unicel). Queremos identificar a la esfera que mide 6 cm. de diámetro.

La metodología propuesta se describe en el siguiente diagrama de flujo, considerando que antes de utilizarla se calibraron ambas cámaras de manera estereoscópica:

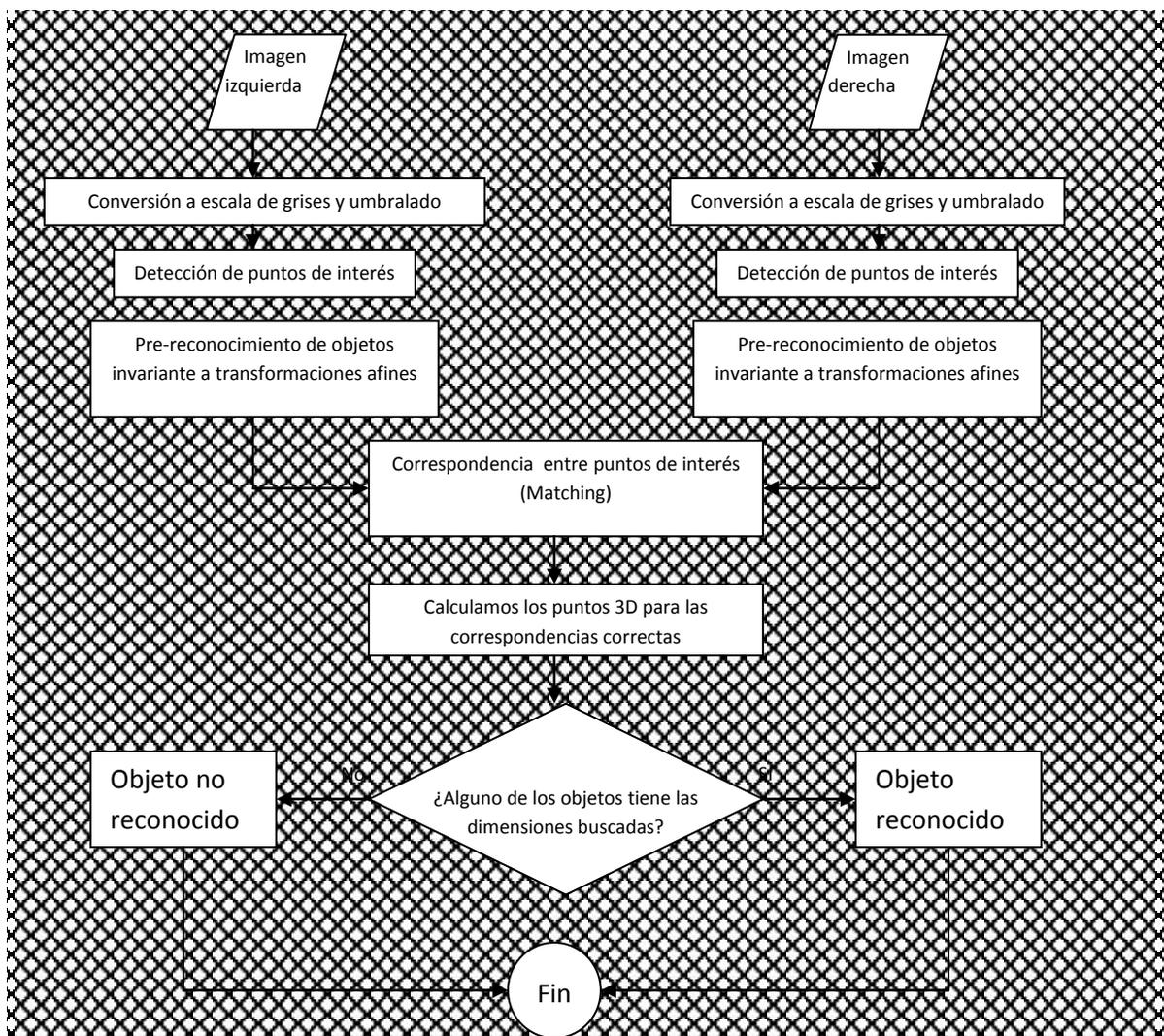


Figura 24. Diagrama de flujo de la metodología propuesta para el reconocimiento de objetos en tres dimensiones.

Como se puede observar la metodología propuesta tiene algunos supuestos, como que las dos cámaras deben observar en su mayoría las mismas cosas sólo que en distinto punto de vista, las cámaras deben manejar el mismo tipo de representación de imágenes.

4.3 Experimentación

4.3.1 Aproximación de un Punto 3D en base a 2 fotografías usando visión estereoscópica

Introducción.

Se tiene por objetivo mostrar la metodología propuesta paso a paso, así como demostrar su funcionamiento y compararlo con metodología que actualmente se utilizan para el reconocimiento de objetos.

Para comenzar con la experimentación es necesario, proponer una escena, calibrar las cámaras digitales e indicar a la computadora el modelo del objeto a reconocer en las escenas. En la Figura 20., se muestra una escena donde se pretende conocer el diámetro de cada esfera partir de las coordenadas conocidas de la imagen derecha (x_d, y_d) e izquierda (x_i, y_i) y después utilizarlas para reconocer a la esfera seleccionada en el modelo. Para lograrlo utilizamos lo expuesto en el apartado anterior. Calibrar de manera individual cada una de las cámaras, calibrar de forma estereoscópica y por ultimo encontrar la triangulación entre estas.

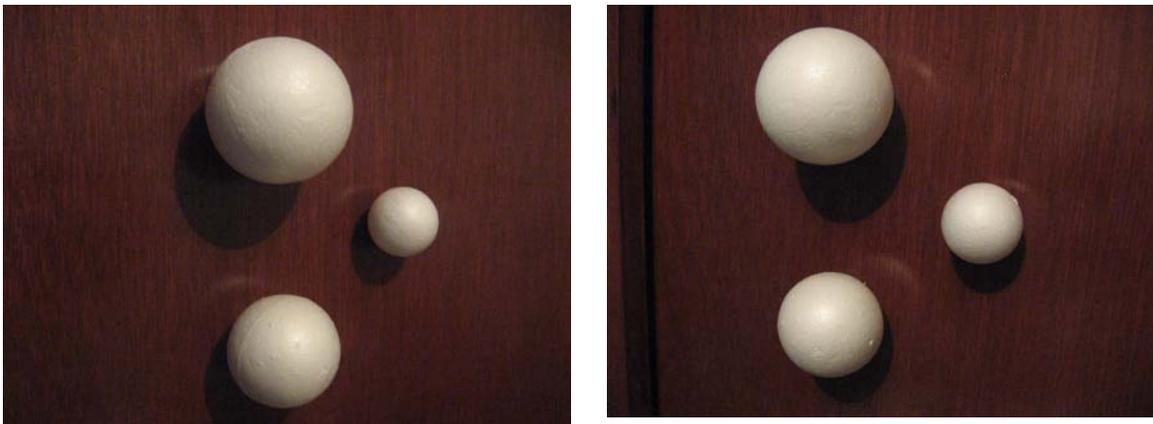


Figura 25. Se muestra una escena tomada de 2 diferentes puntos de vista. El punto de interés es la esquina superior derecha de la vista frontal del libro.

Como se puede observar en la figura 20, se tienen 3 esferas del mismo color y material, por lo tanto si utilizamos visión plana (2D) el reconocimiento de objetos debía de utilizar una distancia fija entre los objetos de estudio y la cámara. La notoria ventaja entre el sistema

propuesto no necesita de esta restricción, debido a que utilizamos visión estereoscópica y con ayuda de la triangulación podemos estimar las dimensiones de cada esfera y reconocer a la esfera seleccionada utilizando la información de sus dimensiones.

4.3.2 La calibración de las cámaras

Al utilizar la metodología de calibración individual de cámaras digitales expuesta en el capítulo 2, se obtuvieron los siguientes resultados, mostrados en las tablas 2 y 3.

Distancia Focal:	$fc = [529.38517 \ 532.00371] \pm [10.51844 \ 11.35446]$
Punto principal :	$cc = [205.75486 \ 97.02537] \pm [20.45086 \ 16.15632]$
Skew:	$\alpha_c = [0.00000] \pm [0.00000] \Rightarrow$ angle of pixel axes = 90.00000 ± 0.00000 degrees
Distorsión:	$kc = [-0.33528 \ 1.30429 \ -0.00769 \ 0.00282 \ 0.00000] \pm [0.13930 \ 1.32092 \ 0.00592 \ 0.00700 \ 0.00000]$
Error en pixel:	$err = [0.41502 \ 0.28730]$

Tabla 2. Los resultados (parámetros intrínsecos) de calibrar la cámara izquierda del sistema en formato Heikkilä.

Distancia Focal:	$fc = [569.86012 \ 570.82381] \pm [6.97541 \ 7.72294]$
Punto principal :	$cc = [182.92097 \ 126.68038] \pm [10.92345 \ 11.54650]$
Skew:	$\alpha_c = [0.00000] \pm [0.00000] \Rightarrow$ angle of pixel axes = 90.00000 ± 0.00000 degrees
Distorsión:	$kc = [-0.39631 \ 2.26238 \ -0.00844 \ -0.00010 \ 0.00000] \pm [0.11383 \ 1.47155 \ 0.00385 \ 0.00410 \ 0.00000]$
Error en pixel:	$err = [0.28760 \ 0.26163]$

Tabla 3. Los resultados (parámetros intrínsecos) de calibrar la cámara derecha del sistema en formato Heikkilä.

También se muestra el resultado de la calibración estereoscópica utilizando el método estudiado en el capítulo 3 en la Tabla 4., donde se expresan los valores de los ángulos de rotación del sistema de coordenadas de la cámara derecha respecto a los de la cámara izquierda en radianes, así como el valor del vector de traslación que describe la posición de la cámara derecha respecto de la izquierda.

Rotation vector:	$\omega = [0.01131 \ 0.55594 \ 0.03174]$
Translation vector:	$T = [-311.29408 \ 4.63580 \ 62.21443]$

Tabla 4. Resultados de calibración estéreo, donde se indican los 3 ángulos de rotación en el vector de traslación respecto de la cámara izquierda.

El formato de presentación de los datos se encuentra en el propuesto por Janne Heikkilä. Descrito en el apéndice A.

4.3.3 Obtención de imágenes y conversión a escala de grises

Las cámaras digitales actuales realizan la captura de fotografías a color (generalmente RGB). Pero para reducir el tiempo de procesamiento a razón de 3, gracias a la redundancia de información desde el punto de vista de bordes en las imágenes, se opta por convertir la imagen a color a escala de grises como se muestra en la figura 21., donde se tienen las imágenes de la cámara derecha e izquierda con sus respectivas conversiones en escala de grises.

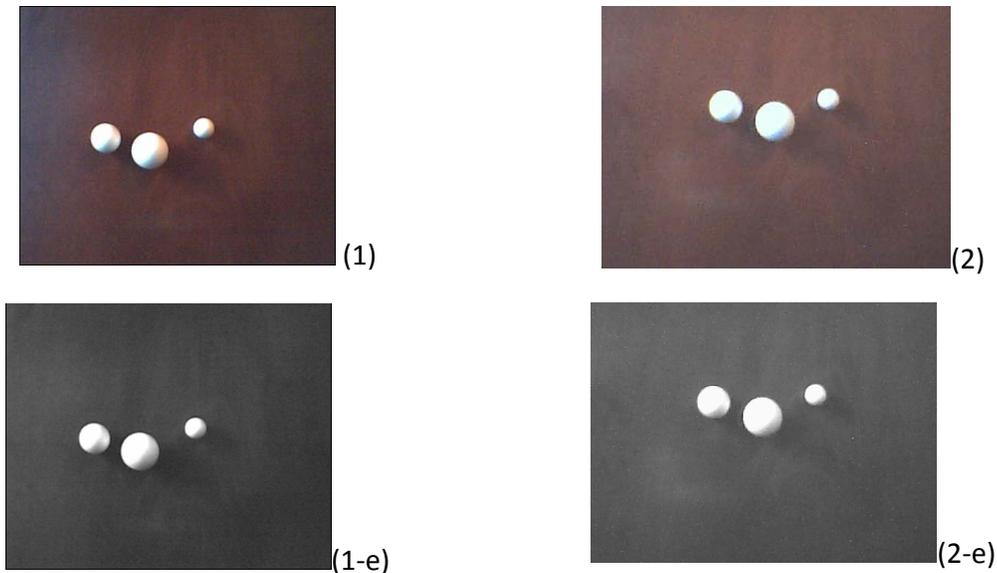


Figura 21. Las dos imágenes tomadas por el sistema estereoscópico: (1) es la imagen tomada por la cámara derecha y (2) la imagen tomada por la cámara izquierda. Debajo de cada imagen su respectiva versión en escala de grises.

4.3.4 Filtrado y Umbralado

Se ha optado por utilizar una etapa de filtrado pasabajas para reducir la probabilidad de procesar ruido en las siguientes etapas, como impurezas del lente, suciedad alrededor de los objetos u objetos muy lejanos que no pueden ser reconocidos. El filtro pasabajas utilizado es del tipo mediana con tamaño de ventana 10 x 10, los resultados del filtrado se muestran en la figura 22. El umbralado de la imagen filtrada también es un proceso importante, pues aquí es donde se definen las formas que se procesarán más adelante y las que no obteniendo una imagen binarizada (ceros y unos) de las formas de interés que posiblemente son los objetos de estudio, la metodología empleada es la propuesta por Otsu, el resultado del proceso de umbralado también se muestra en la figura 22.

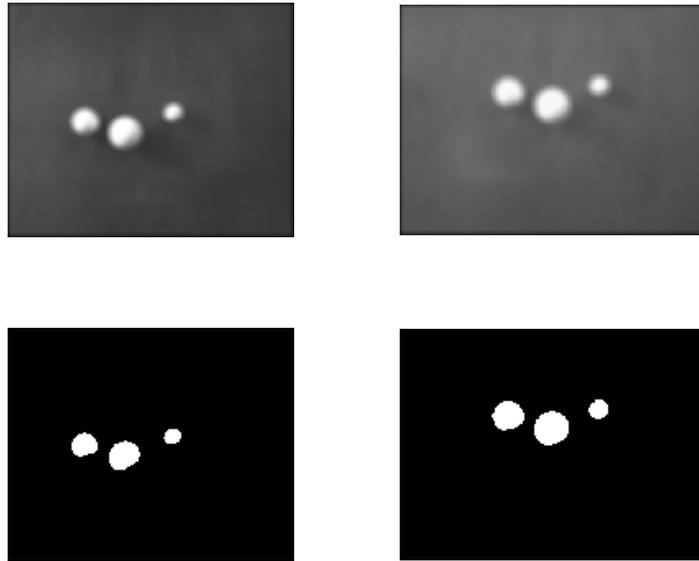


Figura 26. *En la parte superior se muestran las imágenes de las fotografías en escala de grises filtradas del tipo mediana de 10 x 10, en la parte inferior las imágenes binarizadas por el método de Otsu.*

4.3.5 Detección de bordes

Debido a que se desea que este sistema funcione para la detección de objetos de diferentes formas y este sea fácil de adaptar, se ha optado por utilizar un detector de rasgos del tipo bordes llamado operador de Sobel. Este operador aplicado sobre una imagen digital umbralada, obtiene el gradiente de la intensidad de brillo de cada punto de la imagen dando la dirección del mayor incremento posible (de zonas oscuras a blancas) además calcula la cantidad de cambio normalizado en esa dirección, es decir, devuelve un vector. El resultado dice qué tan abruptamente o suavemente cambia una imagen en cada punto analizado, y a su vez que tanto un punto determinado representa un borde en la imagen y también la orientación a la que tiende ese borde, aunque en la práctica solo se utilice la magnitud y no la dirección. Matemáticamente, el gradiente de una función de dos variables (para este caso la función de intensidad de la imagen) para cada punto es un vector bidimensional cuyos componentes están dados por las primeras derivadas de las direcciones verticales y horizontales. Para cada punto de la imagen, el gradiente del vector apunta en dirección del incremento máximo posible de intensidad, y la magnitud del gradiente del vector corresponde a la cantidad de cambio de intensidad en esa dirección. Lo anterior implica que el resultado de aplicar el operador sobel sobre la región de una imagen con intensidad de brillo constante es un vector cero, y el resultado de aplicarlo en un punto sobre un borde es un vector que

apunta cruzando el borde (perpendicular) en sentido de los puntos más oscuros hacia los más claros.

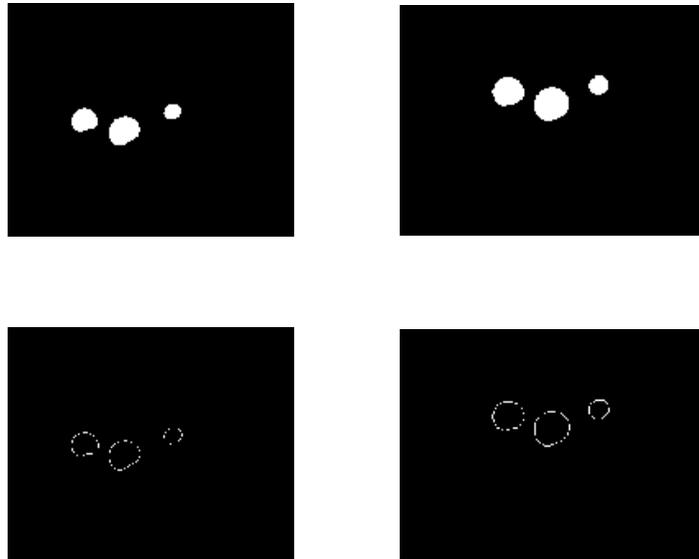


Figura 27. Las imágenes de la parte inferior muestran el resultado de la detección de bordes de las imágenes de la parte superior por el método de sobel.

4.3.6 Etiquetado y separación de objetos en la imagen

Este proceso sigue el algoritmo reportado en Harilack (1991), donde primero se realiza un barrido hasta encontrar un 1 y asignarle una etiqueta E, después asignar de forma recursiva la etiqueta E a todos sus vecinos, según la métrica elegida, detenerse hasta no encontrar más píxeles de ese objeto y regresar al paso 1 donde se hace el barrido de la imagen. Ya que se etiquetaron los objetos en la imagen binarizada se procede a separar a cada uno de ellos en diferentes imágenes como se muestra en la figura 24.

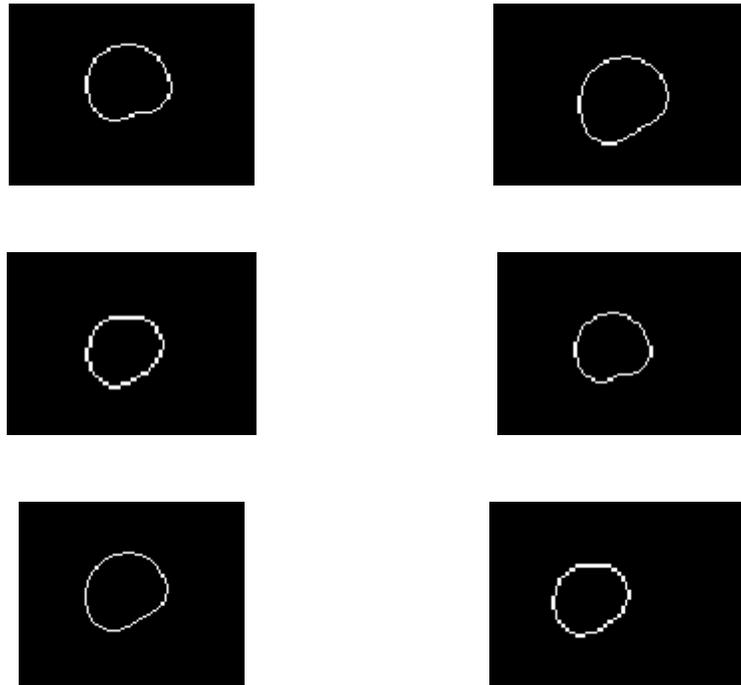


Figura 28. Elementos de las imágenes binarizadas presentados en imágenes independientes para ser analizados.

4.3.7 Prerreconocimiento de objetos sin dimensionalidad

En esta etapa se realiza el reconocimiento de objetos del tipo estadístico invariante a transformaciones afines, pues lo único que se necesita es analizar si la figura que buscamos reconocer corresponde a la que requerimos (en este caso debe corresponder a la silueta de una esfera). El proceso se realiza mediante el uso de los invariantes de hu, que se derivan de los momentos estadísticos centrados en la media y que básicamente mapean a cada silueta formada en una imagen de 2 dimensiones binaria a un espacio de 7 dimensiones (los 7 invariantes de Hu). Dichos coeficientes se calculan como sigue:

$$\phi_1 = \eta_{20} + \eta_{02} \tag{56}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{57}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + 3(\eta_{21} - \eta_{03})^2 \quad (58)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (59)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + 3(\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (60)$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (61)$$

$$\phi_7 = (3\eta_{21} - \eta_{30})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (62)$$

Donde:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad \gamma = \frac{p+q}{2} + 1, \quad p+q = 2,3,\dots \quad (63)$$

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (64)$$

$$\text{Con: } \bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}}, \quad m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (65)$$

Donde $f(x, y)$ es el valor del píxel en el punto (x, y) en la imagen binaria que contiene la forma que deseamos identificar. Debido a que p y q en (36) toman todos los valores enteros no negativos, se puede demostrar que dichos índices generan un conjunto infinito de momentos y que dicho conjunto determina en forma única cada función $f(x, y)$, e inversamente dichos valores quedan determinados unívocamente por $f(x, y)$, [papo1965]. Los invariantes de hu son invariantes a traslaciones, rotaciones y cambio de escala de una forma en la imagen, lo que permite identificar la forma que requerimos reconocer, si sabes que esta última está ligeramente modificada, podemos definir rangos en nuestras variables para identificar a la forma deseada. Así es como se realiza en el ejemplo de las esferas donde los coeficientes de Hu normalizados resultantes se muestran en la tabla 5. De ahí se pueden observar las magnitudes de los coeficientes de hu y valorarlos para establecer los límites que se tomarán en cuenta para clasificar a las posibles esferas contenidas en una imagen.

	F_{i_1}	F_{i_2}	F_{i_3}	F_{i_4}	F_{i_5}	F_{i_6}	F_{i_7}
Objeto1	2.6960	0.0413	0.0790	0.0002	-0.0011e-3	0.0018	0.0001e-4
Objeto2	3.1792	0.0520	0.1831	0.0088	0.1775e-3	0.0059	-0.7470e-4
Objeto3	1.7845	0.0323	0.0215	0.0048	-0.0453e-3	0.0049	0.1322e-4

Tabla 5. Valores de los 7 invariantes de Hu para las 3 esferas contenidas en la imagen de la cámara derecha.

El proceso desarrollado para establecer dichos límites en esta tesis se realizó observando los valores típicos de cada coeficiente para 30 esferas, después se realizó un redondeo de los valores. Los límites para este caso se presentan en la tabla 6. Cabe destacar que este proceso también se puede automatizar con alguna técnica de optimización, redes neuronales y/o inteligencia artificial, pero dicho procedimiento no fue propósito de este trabajo.

	F_{i_1}	F_{i_2}	F_{i_3}	F_{i_4}	F_{i_5}	F_{i_6}	F_{i_7}
Límite Superior	3.50	0.60	0.20	0.0060	0.3e-3	0.0100	0.3e-4
Límite Inferior	1.10	0.25	0.01	0.0001	0.06e-3	0.0005	-0.9e-4

Tabla 6. Los límites para la clasificación de esferas tomados de forma experimental solo seleccionando máximos y mínimos y redondeando.

4.3.8 Medición automática de dimensiones

Después de la etapa de prerreconocimiento, donde se verificó si alguna de las formas encontradas en la imagen correspondía con la forma buscada (en este caso la silueta de una esfera), se trabajará con la etapa de medición automática de dimensiones de las siluetas. Para esto se escoge una dimensión característica del objeto, que para efectos de esta tesis llamaremos rasgo dimensional. En el caso de estudio presentado en esta tesis medimos el radio de las esferas encontradas en la imagen, para esto contamos con un sistema de visión estereoscopio calibrado como se planteo en la sección de calibración de cámaras.

El rasgo dimensional seleccionado fue el diámetro de la esfera medido desde el extremo superior al inferior de esta, así como se muestra en la figura 24. Se escogen estos puntos

debido a que en las 2 imágenes del sistema estereoscópico, estos puntos son el mismo, ya que asumimos que el sistema esta calibrado y colocado sobre una línea horizontal.

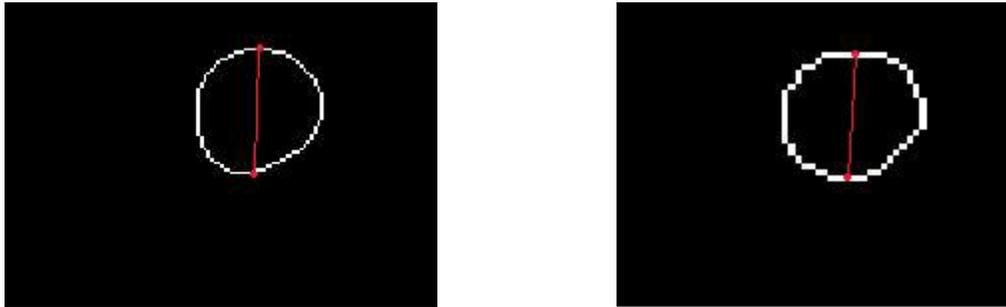


Figura 29. Se detecta el punto superior de la esfera y el punto inferior de la misma para medir su diámetro.

Al encontrar los puntos (x,y) de los extremos superior e inferior de cada una de las esferas en las imágenes derecha e izquierda, se procede a calcular el diámetro de dicha esfera correlacionando directamente el extremo superior de cada una de las imágenes (derecha e izquierda) de la esfera. Por lo tanto utilizamos nuestro modelo de cámaras estereoscópico calibrado, donde sus entradas son los puntos: (x_I, y_I) y (x_D, y_D) y la salida es un punto en el espacio (x, y, z) . Este proceso se ilustra en la Tablas 7., y Tabla 8., Donde se muestran los resultados para los puntos superiores e inferiores respectivamente.

Coordenada	x_left_1	x_right_1	Xc_1_left
X	93 píxeles	220 píxeles	-70.6037 mm.
Y	102 píxeles	71 píxeles	41.6543 mm.
Z			846.0288 mm.

Tabla 7. Se presentan las coordenadas (en píxeles) de los dos puntos superiores de la esfera en la cámara izquierda (x_{left_1}) y derecha (x_{right_1}), así como sus coordenadas en el espacio respecto a la cámara izquierda (Xc_{1_left}).

Coordenada	x_left_1	x_right_1	Xc_1_left
X	97	230	69.7538
Y	185	159	51.9970
Z			365.9866

Tabla 8. Se presentan las coordenadas (en píxeles) de los dos puntos inferiores de la esfera en la cámara izquierda (x_left_1) y derecha (x_right_1), así como sus coordenadas en el espacio respecto a la cámara derecha.

El último paso es verificar si las dimensiones de la esfera corresponden con el diámetro buscado. En este caso nosotros queremos reconocer a la esfera de diámetro 5 cm. Con ayuda de los dos puntos en el espacio (superior e inferior) de la esfera se calcula el diámetro de la misma por medio de la ecuación de la distancia entre 2 puntos.

$$d = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2} \quad (66)$$

El resultado en el ejemplo que hemos venido manejando el resultado fue:

500.2476

Donde sus unidades se encuentran en milímetros. Lo cuál verifica que la esfera mediana que se encuentra en la imagen de las tres esferas es la esfera que buscábamos, y esto se puede calcular de manera automática.

Capítulo

5

“Copia de un libro y lo llamarán plagio. Copia de varios, y lo llamarán investigación”. Anónimo.

Conclusiones



Una vez aplicada la metodología de investigación para de recolección de datos, procesados los mismos con las técnicas propuestas y obtenido la información que de ello se generó conjuntamente con los respectivos análisis, se obtuvieron resultados que le permite al investigador presentar el siguiente conjunto de conclusiones:

1. Se obtuvo una metodología novedosa y exitosa en el reconocimiento automático de objetos basado en las formas de los objetos y en las dimensiones de los mismos.
2. Se analizaron diferentes metodologías para realizar el reconocimiento de formas invariantes ante transformaciones afines como: Traslación, rotación y cambios de escala.
3. Se obtuvo un sistema de reconocimiento flexible respecto a diferentes formas de los objetos de interés.
4. Se analizaron diferentes maneras de calibrar una cámara digital y se decidió por utilizar la más conveniente a esta tesis.
5. La medición de operaciones de punto flotante necesarias para la realización del reconocimiento de formas y dimensiones de objetos.

Resumen de contribuciones:

- ♣ La metodología propuesta para el reconocimiento de objetos presentada en el capítulo 4, figura 22.
- ♣ Concepto de rasgo dimensional (RDRO) para el reconocimiento de objetos, para la discriminación de información innecesaria o redundante.

Investigación Futura o perspectivas:

- ♣ Mejorar la eficiencia de la metodología.
- ♣ Automatizar el proceso de medición de invariantes a transformaciones afines para los objetos de interés.
- ♣ Minimizar el uso de puntos de las formas a reconocer.

Referencias

- [1] Marchand, E., Chaumette, F. 1999. “**An autonomous active vision system for complete and accurate 3D scene reconstruction**”. Intern. Journal of C.V. 32(3): 171-194 p.
- [2] González R.C., Woods R.E. (1993). **Digital Image Processing**, Addison-Wesley, Reading, MA.
- [3] Rosenfeld A., SURVEY, **Image Analysis and Computer Vision: 1993**, Computer Vision Image Understanding, 62(1), 90-143.
- [4] Pajares G., Dela Cruz J., **Visión por computador**, 2002, Alfaomega, Ra-Ma, (1)-p1.
- [5] Barranco, G.A.I (2006): **Sistema Mecatrónico Controlado Telemáticamente**, México D.F.: Instituto Politécnico Nacional.
- [6] Barranco G.A.I., Medel J., Sánchez G. (2006): *Telecontrol sobre TCP/IP*, Aleph Zero, 45, Julio-Septiembre 2007.
- [7] Gonzáles P. De la Cruz J. M. (2004): **Visión por computador.**, Madrid España. AlfaOmega RaMa.
- [8] Voss K., Marroquin J.L., Gutiérrez S. J., Suesse H., (2006): **Análisis de Imágenes de Objetos Tridimensionales.**, México, Instituto Politécnico Nacional.
- [9] Hartley R., Zisserman A. (2003): **Multiple View Geometry in Computer Vision.**, Canberra, Australia. Cambridge University Press.
- [10] Lira J. (2002): **Introducción al tratamiento digital de imágenes.**, México. Instituto Politécnico Nacional.
- [11] Trucco E., Verri A., (1998): **Introductory Techniques For 3d Computer Vision.**, Genova Italia, Prentice Hall.
- [12] Image Processing Toolbox Team (2007): **User’s Guide**, USA, The MathWorks
- [13] Faugeras O. Lustman F. (1988): **Motion and Structure from Motion in a Piecewise Planar Environment**, France, INRIA.
- [14] Ritter G., Wilson J., (1996): **Handbook of Computer Vision Algorithms in Image Algebra**, Internet, ITKnowledge.
- [15] Espiau B. (1992), **A New Approach To Visual Servoing In Robotics**, IEEE Transaction in Robotics and Automatiion, Vol 18 N 3, Institute of Electrical and Electronics Engineers.
- [16] Sossa H., (2006), **Rasgos Descriptores para el reconocimiento de Objetos**, México D.F., Insituto Politécnico Nacional.
- [17] Mery D. (2004), **Visión por Computador**, Santiago de Chile, Universidad Católica de Chile.
- [18] Zhang Z. (1999), **Flexible Camera Calibration By Viewing a Plane From Unknown Orientations**, Redmon USA, IEEE.
- [19] Hotelling, H. (1933) **Analysis of a complex of statistical variables into principal components.** *Journal of Educational Psychology*, 24:417-441,498-520.
- [20] Pearson, K. 1901. **On lines and planes of closest fit to systems of points in space.** *Philosophical Magazine* 2:559-572

Apéndice A

Reconocimiento automático de objetos en imágenes estereoscópicas por medio de características dimensionales

A. I. Barranco Gutiérrez^a, J.J. Medel Juárez^{a,b}

^a*Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada,
Calzada Legarí 694 Col. Irrigación, 11500,
e-mail: barranco_alejandro@yahoo.com.mx*

^b*Centro de Investigación en Computación,
Calle Venus S/N, Col. Nueva Industrial Vallejo, 07738,
e-mail: jjmedelj@yahoo.com.mx*

El este trabajo se presenta una solución al problema del reconocimiento de objetos 3D en imágenes considerando la información de tres dimensiones (3D), proporcionada por un sistema de visión estereoscópico conformado por cámaras digitales convencionales. Esto resuelve el problema de la flexibilidad requerida por la dinámica de los sistemas en la vida actual, donde los objetos no pueden ser capturados por módulos de visión por computadora que requieren el objeto a reconocer a una distancia fija, respecto a las cámaras digitales que lo capturan, o donde las dimensiones de los objetos son factor importante para realizar el reconocimiento. En este trabajo se propone la solución con un tipo de reconocimiento de objetos en 3D basados en la visión estereoscópica y en la estadística del objeto a reconocer en la imagen. Se obtuvieron resultados que abordan este tema desde una perspectiva diferente a la que comúnmente se está haciendo alrededor del mundo. Ya que actualmente se encuentra una división entre los investigadores que perfeccionan las técnicas de reconocimientos de patrones y quienes reconstruyen objetos en 3 Dimensiones. La metodología puede utilizarse para muchas aplicaciones sin modificar su estructura y pueden modificarse algunos parámetros para la aplicación que se requiera utilizar de manera automática, esto comprueba la versatilidad de la técnica.

Descriptores: Reconocimiento de objetos en 3D, Visión estereoscópica, momentos estadísticos.

This paper presents the solution to the problem of 3D object recognition in images using the information from three dimensional (3D), provided by a stereo vision system consists of conventional digital cameras. This solves the problem of the flexibility required by the dynamics of systems in modern life, where objects can not be captured by computer vision modules that require the subject to recognize a fixed distance respect to the digital cameras, or where the dimensions of objects are important for recognition. This paper proposes a solution for the 3D object recognition based on stereo vision and the objects statistics in order to recognize them. Obtained results that address this issue from a different perspective than is commonly being done around the world. Because now exist a division between researchers who perfected the techniques of pattern recognition and who works on 3D object reconstruction. The methodology can be used for many applications without changing its structure and may change some parameters to the application that is required to use automatically, this shows the versatility of the technique.

Keywords: 3d Object recognition, stereo vision, statistical moments.

1.- INTRODUCCIÓN

De manera general, un sistema para el reconocimiento automatizado de objetos (SRAO) permite a una máquina encontrar (reconocer y posicionar en una imagen) objetos de su entorno a partir de una o más imágenes del mundo real, usando modelos de los objetos conocidos a priori [1]. El problema del reconocimiento automático de objetos (SRAO) puede definirse como un problema de etiquetado que se basa en modelos de objetos conocidos. Formalmente el problema se puede describir de la siguiente forma: dada una imagen que contiene uno o más objetos de interés, además del fondo y un conjunto de etiquetas; una para cada región de la imagen, el sistema debe asignar etiquetas correctas (desde el punto de vista de correspondencias respecto a modelos conocidos) a regiones o conjuntos de regiones en la imagen [2], así como se muestra en la Figura 1.

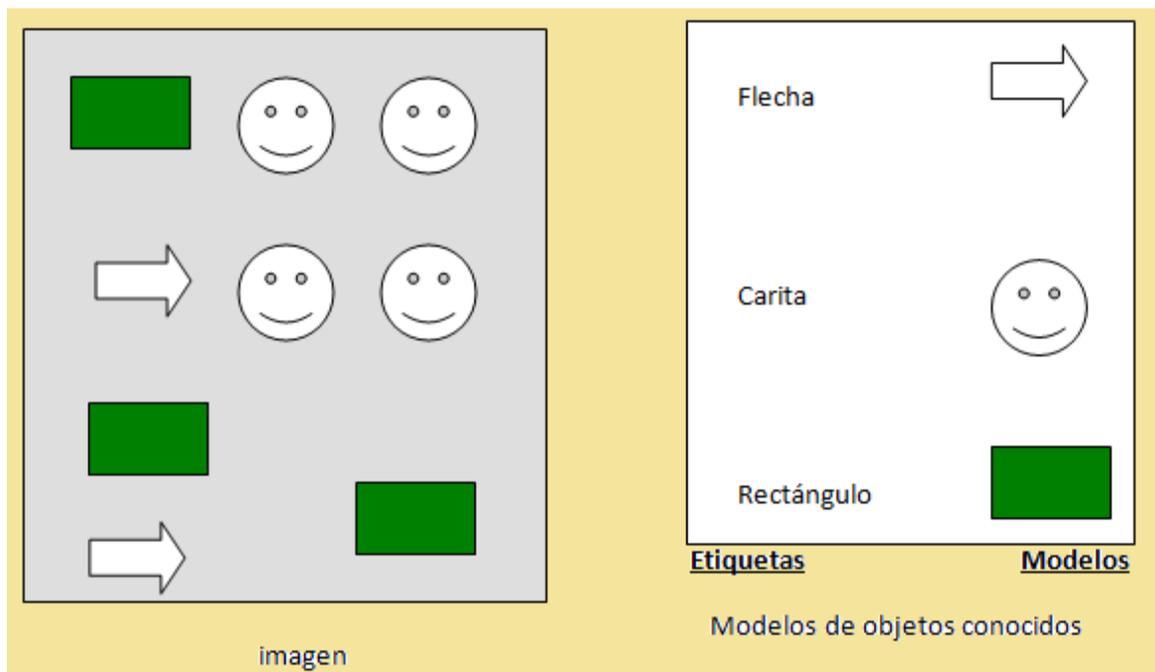


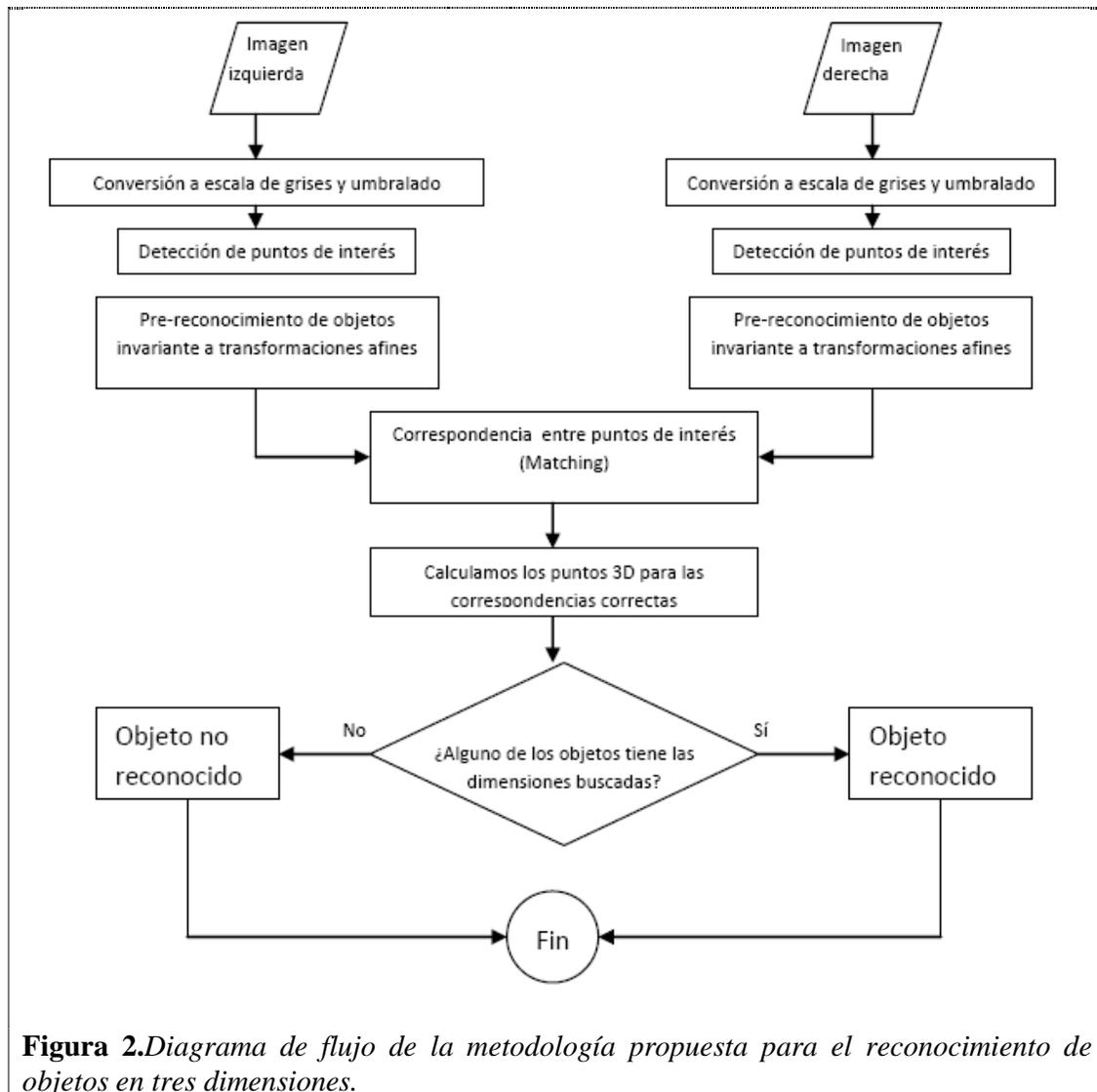
Figura 1. A la izquierda una imagen con algunos de los objetos conocidos que se enumeran a la derecha de la presente figura.

2.- COMPONENTES DE UN SRAO.

En los últimos años se han propuesto muchas metodologías para el SRAO, pero generalmente se incluyen los siguientes componentes: Un banco de modelos, Un módulo de preprocesamiento y acondicionado de la imagen, un módulo de segmentación o aislamiento, un módulo de extracción de rasgos, un módulo generador de hipótesis y un módulo verificador de hipótesis [3]. Estos componentes no consideran el análisis dimensional de los objetos, debido a que solo analizan las formas de los objetos, pero no pueden diferenciar entre dos objetos con la misma forma pero con diferentes dimensiones [1].

Se ha reportado el funcionamiento de un sistema de reconocimiento de objetos en 3D. En los cuales podemos observar que se logra el reconocimiento de objetos similares, donde la diferencia entre los objetos radica en las dimensiones volumétricas de los cuerpos, concretamente se presenta el ejemplo de esferas de diferentes tamaños, del mismo color (blancas) y el mismo material (poliestireno expandido). Donde el objeto a reconocer es la esfera que mide 6 cm., de diámetro.

La metodología propuesta se describe en el siguiente diagrama de flujo ilustrado en la Figura 2., considerando que antes de utilizarla se calibraron ambas cámaras de manera estereoscópica:



Como se puede observar la metodología propuesta tiene algunas restricciones por ejemplo, que las dos cámaras deben observar en su mayoría la escena desde distinto punto de vista, las cámaras deben manejar el mismo tipo de representación de imágenes: matricial, en escala de grises.

3.- DESCRIPCIÓN DEL RECONOCIMIENTO DE OBJETOS EN 3D.

El reconocimiento se basa en la representación matricial de imágenes en escala de grises, donde cada imagen se puede denotar por una matriz de $n \times m$, así entonces para la imagen derecha e izquierda respectivamente. $D \in R^{[m \times n]}$, $I \in R^{[m \times n]}$. Para eliminar algunas impurezas en las imágenes se utilizan filtros, el utilizado en este trabajo fue el del tipo mediana [4], para obtener sólo bajas frecuencias en las fotografías. Para el umbralado de las imágenes se considera un

sistema descrito por Otsu [5] en donde tenemos como resultado una representación de las imágenes del tipo binaria como se muestra en la Figura 3.

Definición 1. Una función de imagen o imagen es cualquier función real $i(x, y)$ con integral finita y soporte compacto S tal que, para todo punto $p \in S, x > 0, y > 0$.

Definición 2. Sean Z el conjunto de enteros. Una imagen digital denotada como $f(x, y)$, es un arreglo bidimensional $\in Z \times Z$ si (x, y) son enteros de $Z \times Z$ y es una función que asigna a cada par (x, y) un número de Z .

Definición 3. Una imagen binaria denotada por $b(x, y)$ es una imagen digital denotada como $f(x, y)$ que ha sido cuantizada a dos niveles de intensidad, 0 y 1.

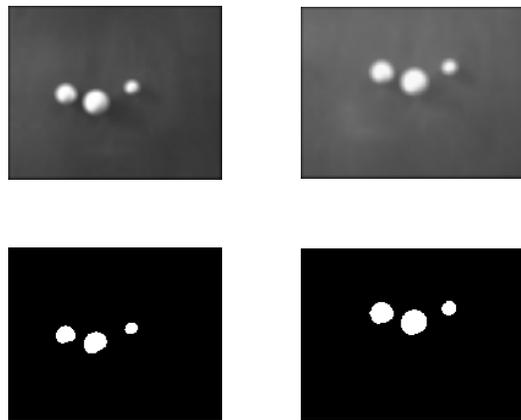


Figura 3. En la parte superior se muestran las imágenes de las fotografías en escala de grises filtradas del tipo mediana de 10×10 , en la parte inferior las imágenes binarizadas por el método de Otsu.

Al umbralar la imagen se obtiene como resultado, regiones que representan objetos de interés, además de una disminución de nuestro universo de búsqueda [2]. Esto nos proporciona objetos candidatos en la imagen para ser analizados y determinar si alguno corresponde con el objeto buscado. Antes de buscar esta correspondencia, en la metodología propuesta se propone la reducción de operaciones computacionales encontrando los contornos de las regiones en la imagen y así una menor cantidad de píxeles a procesar [6] como se muestra en la figura 4.

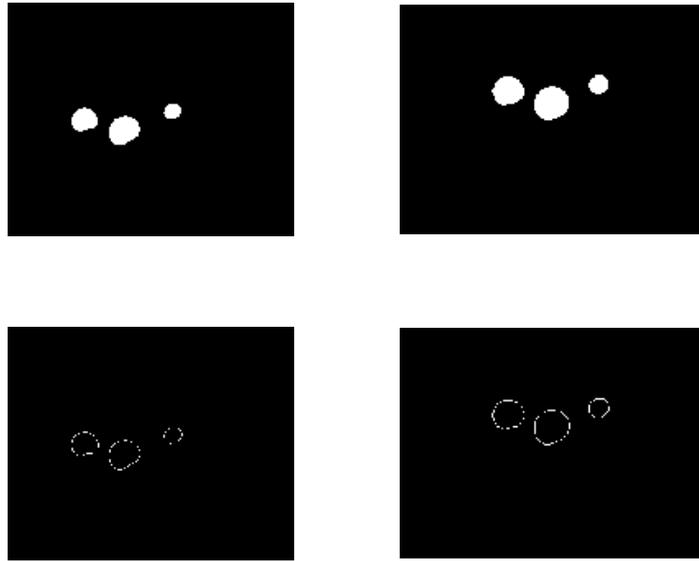


Figura 4. Las imágenes de la parte inferior muestran el resultado de la detección de bordes de las imágenes de la parte superior por el método de sobel[1].

Para hacer un análisis detallado de cada una de las formas obtenidas en la detección de bordes, se procede a separar los objetos aislándolos en una imagen independiente que conserve su posición dentro de la imagen original, para después analizar los rasgos estadísticos de la forma independiente y comparar esos rasgos con los rasgos del banco de modelos almacenados en nuestro SRAO y determinar conforme a una función de pertenencia [7] si dichos parámetros son similares o no, a alguno (s) de los almacenados en el banco de modelos, como se muestra en la figura 5.

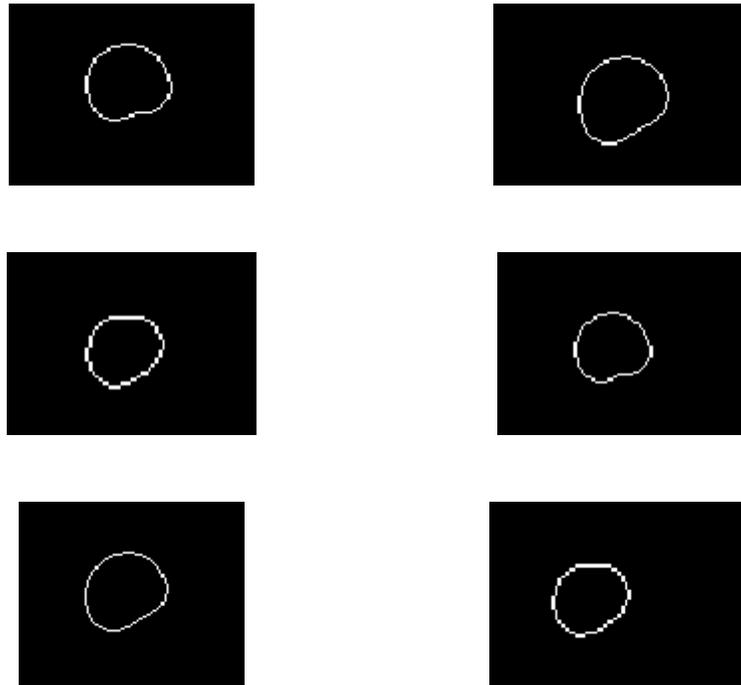


Figura 5. Elementos de las imágenes binarizadas presentados en imágenes independientes para ser analizados.

4.- RECONOCIMIENTO DE OBJETOS SIN DIMENSIONALIDAD

En esta etapa se realiza el reconocimiento de objetos del tipo estadístico invariante a transformaciones afines, pues lo único que se necesita es analizar si la figura que buscamos reconocer corresponde a la que requerimos (en este caso debe corresponder a la silueta de una esfera). El proceso se realiza mediante el uso de los invariantes de Hu [8], que se derivan de los momentos estadísticos centrados en la media y que básicamente mapean a cada silueta formada en una imagen de 2 dimensiones binaria a un espacio de 7 dimensiones (los 7 invariantes de Hu). Dichos coeficientes se calculan como sigue:

$$\phi_1 = \eta_{20} + \eta_{02} \quad (1)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (2)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + 3(\eta_{21} - \eta_{03})^2 \quad (3)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (4)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + 3(\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (5)$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (6)$$

$$\phi_7 = (3\eta_{21} - \eta_{30})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (7)$$

Donde:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad \gamma = \frac{p+q}{2} + 1, \quad p+q = 2,3,\dots \quad (8)$$

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (9)$$

$$\text{Con: } \bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}}, \quad m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (10)$$

Donde $f(x, y)$ es el valor del píxel en el punto (x, y) en la imagen binaria que contiene la forma que deseamos identificar. Debido a que p y q en (9) toman todos los valores enteros no negativos, se puede demostrar que dichos índices generan un conjunto infinito de momentos y que dicho conjunto determina en forma única cada función $f(x, y)$, e inversamente dichos valores quedan determinados unívocamente por $f(x, y)$ [9][10][11]. Los invariantes de Hu son invariantes a traslaciones, rotaciones y cambio de escala de una forma en la imagen, lo que permite identificar la forma que requerimos reconocer, si sabes que esta última está ligeramente modificada, podemos definir rangos en nuestras variables para identificar a la forma deseada. Así es como se realiza en el ejemplo de las esferas (mostrado en este trabajo) donde los coeficientes de Hu normalizados resultantes se muestran en la tabla 5. De ahí se pueden observar las magnitudes de los coeficientes de Hu y valorarlos para establecer los límites que se tomarán para clasificar a las posibles esferas contenidas en una imagen.

5.- RASGOS DIMENSIONALES

Los rasgos dimensionales proporcionan información del objeto a identificar un objeto por medio de alguna o algunas de sus dimensiones características, por ejemplo en este trabajo se muestra un conjunto de esferas del mismo color (blancas), que por tener la misma forma su análisis estadístico de Hu, muestra valores muy similares; pues de esa manera describe las esferas, pero si se requiere ubicar en la imagen a la esfera de diámetro 5 cm.; será necesario comprobar la dimensión del diámetro de la esfera.

Para realizar la medición automática de alguna dimensión a partir de imágenes, es necesario definir las dimensiones a medir y un sistema estereoscópico de cámaras calibradas, que en conjunto nos permitirán reconocer al objeto buscado por medio de su forma y algunos rasgos dimensionales, estos últimos son muy importantes para el reconocimiento en tres dimensiones, ya que este rasgo es la característica que nos ayudará a comprobar la identidad del objeto buscado.

El tipo de cámaras que se utiliza actualmente es del tipo digital, debido a las ventajas que ofrecen estos equipos [1]. La calibración de cámaras que se encuentra detalladamente documentada en [12] [13] [14], donde uno de los métodos más fáciles de implementar es el de Zhang [15], que entrega excelentes resultados que dependen directamente de los datos de entrada como la resolución y enfoque manejados en las cámaras digitales.

Básicamente la calibración de cámaras digitales que se utiliza en este trabajo consiste en conocer los parámetros un modelo matemático que nos relacione los píxeles

	F_{i_1}	F_{i_2}	F_{i_3}	F_{i_4}	F_{i_5}	F_{i_6}	F_{i_7}
Objeto1	2.6960	0.0413	0.0790	0.0002	-0.0011e-3	0.0018	0.0001e-4
Objeto2	3.1792	0.0520	0.1831	0.0088	0.1775e-3	0.0059	-0.7470e-4
Objeto3	1.7845	0.0323	0.0215	0.0048	-0.0453e-3	0.0049	0.1322e-4

Tabla 1. Valores de los 7 invariantes de Hu para las 3 esferas contenidas en la imagen de la cámara derecha.

El proceso desarrollado para establecer dichos límites en esta tesis se realizó observando los valores típicos de cada coeficiente para 30 esferas, después se realizó un redondeo de los valores. Los límites para este caso se presentan en la tabla 6. Cabe destacar que este proceso también se puede automatizar con alguna técnica de optimización, redes neuronales y/o inteligencia artificial, pero dicho procedimiento no fue propósito de este trabajo.

	F_{i_1}	F_{i_2}	F_{i_3}	F_{i_4}	F_{i_5}	F_{i_6}	F_{i_7}
Límite Superior	3.50	0.60	0.20	0.0060	0.3e-3	0.0100	0.3e-4
Límite Inferior	1.10	0.25	0.01	0.0001	0.06e-3	0.0005	-0.9e-4

Tabla 2. Los límites para la clasificación de esferas tomados de forma experimental solo seleccionando máximos y mínimos y redondeando.

6.- MEDICIÓN AUTOMÁTICA DE DIMENSIONES

Después de la etapa de prerreconocimiento, donde se verificó si alguna de las formas encontradas en la imagen correspondía con la forma buscada (en este caso la silueta de una esfera), se trabajará con la etapa de medición automática de dimensiones de las siluetas. Para esto se escoge una dimensión característica del objeto, que para efectos de esta tesis llamaremos rasgo dimensional. En el caso de estudio presentado en esta tesis medimos el radio de las esferas encontradas en la imagen, para esto contamos con un sistema de visión estereoscopio calibrado como se planteo en la sección de calibración de cámaras.

El rasgo dimensional seleccionado fue el diámetro de la esfera medido desde el extremo superior al inferior de esta, así como se muestra en la figura 6. Se escogen estos puntos debido a que en las 2 imágenes del sistema estereoscópico, estos puntos son el mismo, ya que asumimos que el sistema esta calibrado y colocado sobre una línea horizontal.



Figura 6. Se detecta el punto superior de la esfera y el punto inferior de la misma para medir su diámetro.

Al encontrar los puntos (x,y) de los extremos superior e inferior de cada una de las esferas en las imágenes derecha e izquierda, se procede a calcular el diámetro de dicha esfera tomando obteniendo directamente el extremo superior e inferior de cada una de las imágenes (derecha e

izquierda) de la esfera. Por lo tanto utilizamos nuestro modelo estereoscópico de cámaras calibrado, donde sus entradas son los puntos: (x_I, y_I) y (x_D, y_D) y la salida es un punto en el espacio (x, y, z) . Este proceso se ilustra en la Tablas 7., y Tabla 8., Donde se muestran los resultados para los puntos superiores e inferiores respectivamente.

Coordenada	x_left_1	x_right_1	Xc_1_left
X	93 píxeles	220 píxeles	-70.6037 mm.
Y	102 píxeles	71 píxeles	41.6543 mm.
Z			846.0288 mm.

Tabla 3. Se presentan las coordenadas (en píxeles) de los dos puntos superiores de la esfera en la cámara izquierda (x_{left_1}) y derecha (x_{right_1}), así como sus coordenadas en el espacio respecto a la cámara izquierda (Xc_{1_left}).

Coordenada	x_left_1	x_right_1	Xc_1_left
X	97	230	69.7538
Y	185	159	51.9970
Z			365.9866

Tabla 4. Se presentan las coordenadas (en píxeles) de los dos puntos inferiores de la esfera en la cámara izquierda (x_{left_1}) y derecha (x_{right_1}), así como sus coordenadas en el espacio respecto a la cámara derecha.

El último paso es verificar si las dimensiones de la esfera corresponden con el diámetro buscado. En este caso nosotros queremos reconocer a la esfera de diámetro 5 cm. Con ayuda de los dos puntos en el espacio (superior e inferior) de la esfera se calcula el diámetro de la misma por medio de la ecuación de la distancia entre 2 puntos.

$$d = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2} \quad (11)$$

El resultado en el ejemplo que hemos venido manejando el resultado fue: **500.2476**

Donde sus unidades se encuentran en milímetros. Lo cuál verifica que la esfera mediana que se encuentra en la imagen de las tres esferas es la esfera que buscábamos, y esto se puede calcular de manera automática.

7.- CONCLUSIONES

Una vez aplicada la metodología de investigación para de recolección de datos, procesados los mismos con las técnicas propuestas y obtenido la información que de ello se generó conjuntamente con los respectivos análisis, se obtuvieron resultados que le permite al investigador presentar el siguiente conjunto de conclusiones:

6. Se obtuvo una metodología novedosa y exitosa en el reconocimiento automático de objetos basado en las formas de los objetos y en las dimensiones de los mismos.
7. Se analizaron diferentes metodologías para realizar el reconocimiento de formas invariantes ante transformaciones afines como: Traslación, rotación y cambios de escala.
8. Se obtuvo un sistema de reconocimiento flexible respecto a diferentes formas de los objetos de interés.
9. Se analizaron diferentes maneras de calibrar una cámara digital y se decidió por utilizar la más conveniente.

Investigación Futura o perspectivas:

1. Mejorar la eficiencia de la metodología.
2. Automatizar el proceso de medición de invariantes a transformaciones afines para los objetos de interés.
3. Minimizar el uso de puntos de las formas a reconocer

REFERENCIAS

- [1]. Barranco G. A. I., Medel J. J., “Visión estereoscópica por computadora”, Instituto Politécnico Nacional, Cáp. 1., (2008), pág. 3., Primera Edición, México D.F.
- [2]. Sossa H., “Rasgos descriptores para el reconocimiento de Objetos”, Instituto Politécnico Nacional, Cáp. 1., (2006), pág. 3., Primera Edición, México D.F.
- [3]. Y. Amit, “2D Object Detection and Recognition: Models Algorithms, and Networks”, MIT Press, Cáp. 1, (2002), pág. 10., Primer edición, Estados Unidos de América.
- [4]. R. C. Gonzalez y R. E. Woods.”Digital Image Processing”, Prentice Hall, Cáp. 3, (2006), Pág. 152-156, Tercera edición, Estados Unidos de América.
- [5]. N. Otsu, A threshold selection method from gray level histograms, IEEE transactions on systems, man and cybernetics, 9(1),, (2006), Pág. 62-66, 1979.
- [6]. L. Lam, S. W. Lee, Ch. Y. Suen, “Thinning Methodologies – A comprehensive Survey”, IEEE Transactions on pattern analysis and machine intelligence, 14(9): 869-865, 1992.
- [7]. J. J. J. Medel, J. C. I. Garcia, P. L. Guevara, Real-time neuro-fuzzy digital filtering: a technical scheme, Automatic Control and Computer Sciences, Volume 43, Number 1 / febrero de 2009, pag. 22-30.
- [8]. Pajares G. y De la Cruz J. “Visión por computador, imágenes digitales y aplicaciones”, Alfa Omega, Cáp. 9. (2002), Pág. 261-264.
- [9]. Peyton Z., Peebles Jr., “Principios de probabilidad, variables aleatorias y señales aleatorias”, McGrawHill, Cáp. 3 (2006), Pág. 81-84, Cuarta edición traducida al español, Estados Unidos de América.
- [10]. Klaus Voss, José Luis Marroquin, Salvador Jesús Gutiérrez, Hebert Suesse, “Análisis de imágenes de Objetos Tridimensionales”, Instituto Politécnico Nacional, Cáp. 6 (2006), Pág. 123-126, Primera Edición, México.
- [11]. A. I. Barranco, J. J. Medel, “Identificación de formas en imágenes a partir de la media y la varianza de los niveles de color RGB”, Primer Simposio de Tecnología avanzada, Instituto Politécnico Nacional, Pág. 79.
- [12]. Barranco G. A. I., Medel J. J. J., “Digital Camera Calibration Analysis Using Perspective Projection Matrix”, Proceedings of the 8th WSEAS International Conference on Signal Processing, Robotics and Automation, Pág. 321-325 (2009), Cambridge U. K.
- [13]. Barranco G. A. I., Medel J. J. J., “Proceso de calibración de cámaras digitales basado en el modelo pin-hole”, Primer Simposio de Tecnología avanzada, Instituto Politécnico Nacional, Pág. 66.

- [14]. Richard Hartley, Andrew Zisserman, "Multiple view geometry in computer vision", CAMBRIDGE, Cáp. 4 (2003), Segunda edición, Reino Unido
- [15]. Z. Zhang. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330-1334, 2000.

Apéndice B

Digital Camera Calibration Analysis Using Perspective Projection Matrix

BARRANCO GUTIÉRREZ ALEJANDRO ISRAEL¹, JOSÉ DE JESÚS MEDEL JUÁREZ^{1,2}

¹Applied Science and Advanced Technologies Research Center (CICATA)

²Computer Research Center (CIC)

National Polytechnic Institute

Legaria #694 Col. Irrigación México D.F. C.P. 11500

MEXICO

barranco_alejandro@yahoo.com.mx, jjmedelj@yahoo.com.mx

Abstract: This work presents a novel and simplified technique to estimate the perspective projective matrix elements of calibration matrix for pinhole model in Digital Camera Calibration. The "perspective projective matrix parameters" are variables depending of environmental changes and position and/or orientation camera so we propose a dynamical and stochastic method to model the uncertain in the parameters estimation. It is well suited for use without specialized knowledge of 3D geometry or computer vision. Two morphological matrix operations are introduced: *Central*($X_k Y_k$) and *Column*(M_k) to generalize the process which obtain the estimated parameters based on pseudo-inverse calculus without consider measures errors, this calibration procedure focus only on perspective projective matrix elements. The mean value respect to 3D points and 2D points used as input information. The theoretical results give a good enough approximation result considering the pseudo-inverse matrix calculus method. In the same sense, the experimental results showed a satisfactory advance in the parameters stochastic estimation theory, respect to velocity change gains.

Keywords: Camera Calibration, Pseudo-inverse, Stochastic process, filtering theory, estimation, Pin-Hole Model.

1 Introduction

Actually, camera calibration on computer vision becomes to be an important problem to Scientific Community because a machine can analyze the environment images automatically and we could use this theory to solve problems in different knowledge areas as medicine, robotics, physics, security and astronomy [1] [2]. We receive useful quantitative information from the scene in the image for example: real object dimensions, distance between objects and textures [3]. Moreover, this information used to control robots, which also take decisions, by itself for controlled motion in planed trajectories. The "Pin Hole" is very successful model used in the camera calibration field also this work based on this. The Fig. 1., shows the geometric scheme to describe the ideal "Pin Hole" model, a closed box with only one small hole and no lenses are used to focus light. The model does not describe geometric distortions or blurring of unfocused objects caused by lenses and finite sized apertures. Any external object emits light rays and those passes through the hole and an inverted object image projected on the image plane. This model is frequently used in robotics and computer sciences even does not consider all the

physic variables in the mathematic model [4] [5][6] and[7].

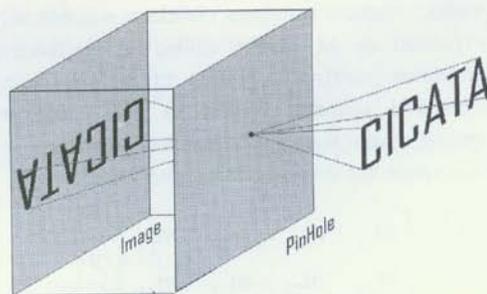


Fig. 1. The "PinHole" model principle consider the external objects image inversion.

2. Basic Equations

The "PinHole" model concept considering the correspondence between the fixed points modeled into finite computer image space, mapping from 3D real image to 2D compute pixel points. Phenomenon

studied by Zhang in [13] and [14]; with normalized transformation (1).

$$Y_m = A[R \ t]X_n \tag{1}$$

Where:

$X_n = [x, y, z, 1]$: Correspond to homogeneous 3D coordinates point.

$Y_m = [u, v, 1]$: Homogeneous 2D compute pixel points as coordinates point

$[R \ t]$: Extrinsic parameters matrix

A : Intrinsic known parameters matrix

$$A = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

The vector $\tilde{u} \in R^3$ allows identifying the object coordinates associated to the vector $\tilde{y} \in R^4$ in the image showed in Fig. 1. The normalized transformation (1) describes this correspondence and includes two important matrixes: Extrinsic parameters matrix $[R|t]$ models the rotation and translation of the camera respect to external reference and intrinsic parameters matrix models A as the pixel shape and the focal length elements [7] [9] [10] [11] [12] and [13]. Conforming mixed matrix respect to inner product symbolically expressed as M matrix calling as perspective projection matrix (2). Inside of it, considering dynamical elements respect to real coordinate system. Such that, the simplified transformation description has the linear description form:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{2}$$

The realistic problem consists in estimate the parameters transformation. Commonly could be use the traditional Wiener filter description by this kind of systems; but input and output vectors have different lengths. In imaginary form, respect to the Homogeneous 2D compute pixel points as

coordinates point:

$$E\{Y_k Y_k^T\} = M_k E\{X_k Y_k^T\} \tag{3}$$

But $X_k Y_k^T$ did not a conformal structure and did not express the common solution as:

$$M_k = E\{Y_k Y_k^T\} (E\{u_k y_k^T\})^{-1} \tag{4}$$

In the "Pin Hole" model, the second probability moment depicted as $E\{X_k Y_k^T\}$ did not mathematical conform structure and the matrix inner product $X_k Y_k^T$ did not realizable.

3. Dynamical model Solution

The "perspective projection matrix parameters" expressed as variables depending of environmental changes and position and/or orientation camera, represented in first instance by basic equation in order to calibrate a digital camera using a reference comparing measurements results and introducing some uncertainty, either using the expectation respect to each of the 3D coordinates points and their projections photography. Such that could be solve the camera calibration problem considering the Pseudo-inverse matrix resolution considering dynamic time-varying internal parameters, as shown in following theorem.

Theorem: Considering the MIMO stochastic system $Y_{k,m \times 1} = M_{m \times n} X_{k,n \times 1} + W_{X_{k,m \times 1}}$ (3); such that its matrix parameter estimator, respect to fix evolution:

$$\text{Column}(M)_{(m \times n) \times 1} = \text{Central}(X_k, Y_k)_{(m \times n) \times m}^+ (Y_{k,m \times 1} + W_{X_{k,m \times 1}}) \tag{4}$$

Proof: The stochastic system model conformed symbolically has the expanded form:

$$Y_{k,m \times 1} = \text{Central}(X_k, Y_k)_{m \times (m \times n)} \text{Column}(M)_{(m \times n) \times 1} + W_{X_{k,m \times 1}}$$

Where $\text{Central}(X_k, Y_k)$ has the form:

$$\text{Central}(X_k, Y_k)_{m \times (m \times n)} := \begin{bmatrix} [x_i, x_{i+1}, \dots, x_m] & \dots & 0 & y_i [x_i, x_{i+1}, \dots, x_m] \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & [x_i, x_{i+1}, \dots, x_m] & y_{i-1} [x_i, x_{i+1}, \dots, x_m] \end{bmatrix}$$

$\text{Column}(M)$ Expanded as vector:

$$Column(M)_{(m \times n) \times 1} := \begin{bmatrix} m_{11} \\ \vdots \\ m_{mn} \end{bmatrix}$$

With W_x as the uncertain added by the environment.

The Central pseudo inverse:

$$Central(X_k)_{(m \times n) \times m}^+ := \left(\left(Central(X_k)_{(m \times n) \times m} \right)^T Central(X_k)_{(m \times n) \times m} \right)^{-1} \left(Central(X_k)_{(m \times n) \times m} \right)^T$$

Such that the parameter set for a fixed evolution is described by (4). ■

4. Optimal Solution

The method described in the previous theorem has an optimal solution considering:

$$Y_k = Central(X_k) Column(M_k) + W_x$$

Where the gradient expressed as:

$$\frac{\partial}{\partial Column(M_k)} Y_k Y_k^T = 0$$

Allow to obtain:

$$Column(M_k) = -Central(X_k)^+ W_x$$

5. Experimental Results

5.1. Data input

The proposed algorithm tested into Matlab. The closed form solution involves finding a group of points of the same image. Inverse methodology allows to compare the certain using an perspective projective matrix given in Zhang[8][13] and [14]:

$$M = 1.0e+003 * \begin{bmatrix} 0.6573 & 0.0003 & 0.3027 & 2.5257 \\ 0 & 0.6577 & 0.2423 & 1.6271 \\ 0 & 0 & 0.0010 & 0.0040 \end{bmatrix}$$

where R :

$$R = R_x R_y R_z = \begin{bmatrix} \cos\beta \cos\theta & \sin\alpha \sin\beta \cos\theta + \cos\alpha \sin\theta & -\cos\alpha \sin\beta \cos\theta + \sin\alpha \sin\theta \\ -\cos\beta \sin\theta & \sin\alpha \sin\beta \sin\theta + \cos\alpha \cos\theta & \cos\alpha \sin\beta \sin\theta + \sin\alpha \cos\theta \\ \sin\beta & -\sin\alpha \cos\beta & \cos\alpha \cos\beta \end{bmatrix}$$

with $\alpha = 0$, $\beta = 0$ and $\theta = 0$

and intrinsic matrix A:

$$A = \begin{bmatrix} 657.3025 & 0.2761 & 302.7166 \\ 0 & 657.7439 & 242.3339 \\ 0 & 0 & 1.0000 \end{bmatrix}$$

The next input homogeneous vectors in fixed evolution used to test the method error.

$$X = \begin{bmatrix} 2 & 2 & 50 & 23 & 6 & 2 & 5 \\ 3 & 34 & 56 & 12 & 23 & 42 & 2 \\ 4 & 2 & 5 & 7 & 8 & 4 & 4 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Using the vector X as 3D point input and applied on the equation (1), it obtained the next Y vector output, which represents the correspondences of X points in the images.

$$Y = 1.0e+004 * \begin{bmatrix} 0.5052 & 0.4455 & 3.6920 & 1.9766 & 0.8898 & 0.5063 & 0.7024 \\ 0.4570 & 2.4475 & 3.9672 & 1.1216 & 1.8694 & 3.0222 & 0.3912 \\ 0.0008 & 0.0006 & 0.0009 & 0.0011 & 0.0012 & 0.0008 & 0.0008 \end{bmatrix}$$

5.2. Central () and Column () operations implementation.

Column() function description using the Matlab function representation code where the % character enclose code comments:

```
% The M matrix is the input data%
% CM is the matrix output%

function [CM] = column(M)

[m,n]=size(M); %To obtain the M dimensions%
k=1;
for j=1:n
    for i=1:m
        CM(k)=M(i,j)
        k=k+1
    end
end
CM=CM' %To obtain the CM transpose%
```

Central() function using the Matlab representing code description:

- [11] Sossa H., Rasgos Descriptores para el reconocimiento de Objetos, *Instituto Politécnico Nacional*. 2006.
- [12] Mery D., Visión por Computador, *Universidad Católica de Chile*, 2004.
- [13] Zhang Z., Flexible Camera Calibration By Viewing a Plane From Unknown Orientations, *IEEE*. 1999.
- [14] Espiau B., A New Approach To Visual Servoing In Robotics, *IEEE Transaction in Robotics and Automation*, Vol.18, No.3, 1992.