



**INSTITUTO POTOSINO DE INVESTIGACIÓN
CIENTÍFICA Y TECNOLÓGICA, A.C.**

POSGRADO EN CIENCIAS APLICADAS

**Calibración de una cámara por seguimiento de
características en una secuencia de imágenes**

Tesis que presenta

Vicente Ferrer Gutiérrez

Para obtener el grado de

Maestro en Ciencias Aplicadas

En la opción de

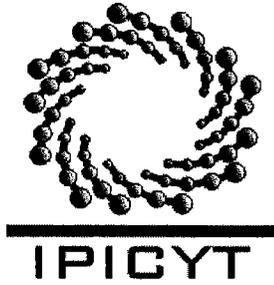
Control y Sistemas Dinámicos

Codirectores de la Tesis:

Dra. Rocío Díaz de León Torres

Dr. Jesús Leyva Ramos

San Luis Potosí, S.L.P., Septiembre de 2004.



Instituto Potosino de Investigación Científica y Tecnológica, A.C.

Acta de Examen de Grado

COPIA CERTIFICADA

El Secretario Académico del Instituto Potosino de Investigación Científica y Tecnológica, A.C., certifica que en el Acta 003 del Libro Primero de Actas de Exámenes de Grado del Programa de Maestría en Ciencias Aplicadas en la opción de Control y Sistemas Dinámicos está asentado lo siguiente:

En la ciudad de San Luis Potosí a los 14 días del mes de septiembre del año 2004, se reunió a las 16:00 horas en las instalaciones del Instituto Potosino de Investigación Científica y Tecnológica, A.C., el Jurado integrado por:

Dr. David Antonio Lizárraga Navarro	Presidente	IPICYT
Dra. Rocío Díaz de León Torres	Secretario	IPICYT
Dr. Emilio Jorge González Galván	Sinodal externo	UASLP
Dr. Jesús Leyva Ramos	Sinodal	IPICYT

a fin de efectuar el examen, que para obtener el Grado de:

**MAESTRO EN CIENCIAS APLICADAS
EN LA OPCIÓN DE CONTROL Y SISTEMAS DINÁMICOS**

sustentó el C.

Vicente Ferrer Gutiérrez

sobre la Tesis intitulada:

Calibración de una cámara por seguimiento de características en una secuencia de imágenes

que se desarrolló bajo la dirección de

Dra. Rocío Díaz de León Torres
Dr. Jesús Leyva Ramos

El Jurado, después de deliberar, determinó

APROBARLO

Dándose por terminado el acto a las 17:30 horas, procediendo a la firma del Acta los integrantes del Jurado. Dando fé el Secretario Académico del Instituto.

A petición del interesado y para los fines que al mismo convengan, se extiende el presente documento en la ciudad de San Luis Potosí, S.L.P., México, a los 14 días del mes septiembre de 2004.


Dr. Marcial Borja Marín
Secretario Académico


INSTITUTO POTOSINO
DE INVESTIGACIÓN
CIENTÍFICA Y TECNOLÓGICA, A.C.

IPICYT
SECRETARIA ACADEMICA


Mtra. Ma. Elisa Lucio Aguilar
Jefa del Departamento de Asuntos Escolares



**INSTITUTO POTOSINO DE
INVESTIGACIÓN CIENTÍFICA Y
TECNOLÓGICA, A.C.**

**DIVISIÓN DE MATEMÁTICAS APLICADAS
Y SISTEMAS COMPUTACIONALES**

**MAESTRÍA EN CIENCIAS APLICADAS CON OPCIÓN EN CONTROL Y
SISTEMAS DINÁMICOS**

**CALIBRACIÓN DE UNA CÁMARA POR SEGUIMIENTO DE
CARACTERÍSTICAS EN UNA SECUENCIA DE IMÁGENES**

Ing. Vicente Ferrer Gutiérrez

SINODALES

Dra. Rocío Díaz de León Torres
(Directora de Tesis)

Dr. Jesús Leyva Ramos
(Director de Tesis)

Dr. Emilio Jorge González Galván.

Dr. David Antonio Lizárraga Navarro.

SAN LUIS POTOSÍ, S.L.P., SEPTIEMBRE DE 2004

*A mis padres, hermanos y
amigos por su total apoyo,
comprensión
y muestras de cariño
incondicional.*

AGRADECIMIENTOS

A mis padres por su apoyo cuando más lo necesitaba y por confiar plenamente en mí.

A la Dra. Rocío Díaz de León y al Dr. Jesús Leyva por sus consejos y dirección brindada a lo largo de este proyecto.

A mis maestros Dr. Hugo Cabrera, Dr. Haret Rosu, Dr. David Lizárraga, Dr. Arturo Zavala y Dr. Ricardo Femat por sus enseñanzas, consejos y amistad.

A mis compañeros y amigos Michael, Misael y Raúl por su apoyo y amistad.

A mi novia Elvia por su amor, comprensión y amistad.

Al programa de becas crédito de CONACYT por el financiamiento que proporcionaron para poder concluir este trabajo.

RESUMEN.

En este trabajo se plantea que un sistema de visión realice la detección, ubicación y seguimiento de un objeto. Se considera como sistema de visión, una cámara de video con la suficiente capacidad técnica para realizar los estudios y lograr nuestros objetivos, así como el programa desarrollado para calibrar la cámara.

Muchas diferentes técnicas en la actualidad muestran como obtener información 3-D del mundo usando una imagen o secuencia de imágenes capturadas por una o más cámaras. Cada técnica incluye una serie de procesos que influyen directamente en lograr un rendimiento eficaz, dentro de estos el principal proceso es la calibración. La calibración es el problema básico en aplicaciones de sistemas de visión en las que se pretenda obtener información geométrica del espacio. Este problema consiste en encontrar los valores de la posición y orientación de una cámara, así como, sus propiedades ópticas, geométricas y digitales a partir de puntos conocidos en el espacio que son proyectados en una imagen.

En la mayoría de los medios actuales, las propiedades de una cámara pueden considerarse como estables y conocidas, por lo tanto, el problema se reduce a determinar la orientación y posición del sistema de referencia de una imagen. Un modelo matemático que describa y relacione correctamente la información del espacio 3-D y su correspondiente información 2-D de una o unas imágenes, es la base principal de realizar una buena calibración. De igual forma, la calibración depende de la precisión con que obtengamos la información del espacio y de la imagen.

El problema de correspondencia de información es la principal dificultad cuando se intenta tomar los datos en la imagen de un objeto en movimiento. Esto ocurre debido a que la ubicación de la cámara debe ser determinada por una secuencia de movimientos de la cámara y el objeto que está siendo observado. Tomando en consideración lo anterior,

planteamos nuestro trabajo desde dos situaciones diferentes: la imagen de una escena y la estructura real de la escena que toma el sistema de visión.

Aquí presentamos técnicas para atacar cada uno de los problemas. Primero hacer un análisis eficaz de la imagen por medio de aplicar un algoritmo de clasificación de color. La idea de llevar a cabo esta clasificación es detectar un objeto con características de color predefinidas. Además de ello utilizamos un espacio de color diferente al tradicional, pero, con propiedades más útiles para hacer la discriminación del color de interés en las imágenes.

El color de la piel es la característica principal en la que se baso para hacer la clasificación. Es claro que al decir color de la piel, se refiere a todo su posible conjunto de combinaciones de la piel humana. Se decidió usar este conjunto de color debido a que el interés por detectar o ubicar personas con algún sistema de visión sobresale por encima de otro u otros objetos. Además de que una amplia posibilidad de aplicaciones futuras a desarrollar con este proyecto involucran captar gente.

Posterior a la detección en base al color, establecemos la estructura de un objeto que sabemos que aparecerá en la escena tomada por la cámara. Este objeto cuenta con las características del color de la piel y el cual debemos detectar en todo momento que se encuentre frente a la cámara. Una vez encontrado el objeto en la imagen, obtenemos puntos de referencia del objeto en 2-D, los cuales, vienen definidos por un modelo del objeto antes definido. Definimos este modelo 3-D del objeto, con la idea de usar esta información en la aplicación de la calibración de la cámara. Relacionamos la información obtenida de la detección con la del modelo del objeto y podemos de esta forma ubicar nuestro objeto en la escena.

Es importante mencionar que la detección usando clasificación de color se puede probar con cualquier tipo de objeto o persona con las características de color antes descritas.

Una vez ubicado el objeto en la imagen y en la escena, lo restante es hacer su seguimiento en el caso de que se esté moviendo. Este seguimiento consiste en aplicar el proceso de detección a cada imagen captada, tomando en cuenta, la información inmediata de la imagen anterior procesada. Además de ello, tenemos que actualizar dinámicamente la ubicación y orientación de la cámara con respecto a las nuevas posiciones que guarde el objeto a medida que se vaya moviendo.

Al final de este documento se muestran resultados de pruebas realizadas para seguir un cubo en un ambiente real.

CONTENIDO.

ÍNDICE DE FIGURAS.	x
1. INTRODUCCIÓN.	1
1.1. Definición del problema.	1
1.2. Objetivo de la tesis.	4
1.3. Metodología.	4
1.4. Propuesta de tesis.	5
1.5. Organización de la tesis.	6
2. CALIBRACIÓN DE UNA CÁMARA.	8
2.1. Definición de calibración.	8
2.2. Métodos de calibración.	13
2.2.1. Calibración directa.	13
2.2.2. Calibración obtenida de la matriz de proyección.	17
2.3. Trabajo relacionado.	20
2.3.1. Calibración con objetos fijos.	21
2.3.2. Calibración con objetos en movimiento.	23
2.4. Conclusiones.	24
3. CLASIFICADORES DE COLOR.	26
3.1. Espacios de color.	26
3.1.1. Espacio de color RGB.	26
3.1.2. Espacio de color HSI.	28
3.1.3. Conversión de RGB a HSI.	30
3.2. Método de clasificación de datos.	38
3.2.1. Algoritmo de agrupamiento k-medias.	38

3.3. Trabajo relacionado.	45
3.4. Conclusiones.	47
4. CALIBRACIÓN DE UNA CÁMARA BASADA EN EL MOVIMIENTO DE UN OBJETO.	48
4.1. Detección y reconocimiento del objeto.	48
4.1.1. Modelo de color.	49
4.1.2. Clasificador de color.	51
4.1.3. Depuración de la imagen.	56
4.1.4. Algoritmo de detección.	57
4.2. Proceso de calibración de la cámara.	57
4.2.1. Algoritmo de calibración.	62
4.3. Conclusiones.	63
5. RESULTADOS EXPERIMENTALES DEL SEGUIMIENTO DE UN CUBO.	64
5.1. Detección del objeto.	65
5.2. Calibración de la cámara.	72
5.3. Seguimiento del objeto.	76
5.4. Algoritmo de seguimiento.	81
5.5. Conclusiones.	82
6. CONCLUSIONES.	83
6.1. Aportaciones.	84
6.2. Limitaciones.	85
6.3. Trabajo a futuro.	86
REFERENCIAS.	87
APÉNDICE A. Diagrama de flujo.	89
APÉNDICE B. Interfaz gráfica del programa de calibración.	

1. INTRODUCCIÓN.

El problema de la calibración de una cámara es importante en todas aquellas áreas de sistemas de visión en donde se pretenda adquirir información de una escena como aplicaciones fotogramétricas y de visión de máquina. De la misma forma, en sistemas de vigilancia se considera en gran medida que la cámara sea capaz de detectar, identificar y seguir a personas para determinar las actividades que realiza. En robótica, la información que pudiera adquirir una cámara o sensor como parte del sistema de visión de un robot sería muy útil para que al procesar las imágenes, el robot sea capaz de identificar personas, actividades u órdenes, y a partir de ello realizar posibles tareas. Además, la calibración es un paso importante en visión por computadora en tercera dimensión que extrae datos geométricos de imágenes en dos dimensiones.

La calibración se define como encontrar los valores de la posición y orientación de una cámara, así como, sus propiedades ópticas, geométricas y digitales a partir de puntos conocidos en el espacio que son proyectados en un plano (imagen).

Muchas diferentes técnicas en la actualidad muestran como obtener información de estructuras 3-D del mundo usando una imagen o secuencia de imágenes capturadas por una o más cámaras.

1.1. DEFINICIÓN DEL PROBLEMA.

En particular se explora la calibración de una cámara con puntos de referencia variables. Dado un objeto en movimiento se calibra la cámara para poderlo seguir, esto significa que primero se tendrá que detectar el objeto en movimiento y posteriormente calibrar la cámara. Se pretende adaptar un método de calibración que hace una estimación iterativa de los parámetros extrínsecos de la cámara con la mayor precisión posible. Se toma una estimación inicial de los parámetros, y en base a la información proporcionada de una imagen y un modelo predefinido de un objeto se puede realizar la calibración.

En fechas recientes se han implementado métodos para alcanzar la medición de los parámetros de una cámara a través de un modelo en proyección perspectiva por seguimiento de características en una imagen, y sin usar patrones especiales de calibración [Chiori y Parvin, 1996], [Dron, 1993]. Este tipo de técnicas tienen que ser capaces de seguir características entre imágenes a medida que el objeto se mueve y extraer la información de secuencias de movimiento para producir una estimación más precisa de los parámetros que son actualizados dinámicamente.

Para desarrollar una calibración eficiente se pretende asignar modelos de objetos con dimensiones conocidas al definir para todo objeto real detectado un modelo 3-D con la similitud precisa para poder establecer su ubicación con respecto a la cámara. Así, relacionando el modelo del objeto con la información del objeto en la imagen podemos encontrar los valores de los parámetros extrínsecos de la cámara, es decir, calibrarla.

Las suposiciones importantes a considerar dentro de esta tesis son: los parámetros intrínsecos de la cámara, es decir, la longitud focal, coordenadas del centro de imagen, tamaño efectivo del píxel y la distorsión radial son conocidos; lo cual es permisible considerando las características de los modelos de cámara actuales. En particular la distorsión radial se considerará como despreciable.

Dentro del algoritmo propuesto se pretende demostrar que los métodos de clasificación de color son una herramienta importante para identificar regiones de interés dentro de una imagen, donde la precisión de cualquier método de calibración depende de las correspondencias adecuadas entre puntos de la imagen y puntos en el espacio. Las metas inmediatas surgidas de los diferentes problemas a los que nos enfrentaremos son:

1. Realizar el método de calibración considerando que la cámara tiene una posición fija, con ello la traslación de la cámara no se puede llevar a cabo.

2. Realizar pruebas para llegar a una estimación inicial aceptable de los ángulos de rotación y el vector de traslación, de acuerdo al modelo del objeto y las posibles posiciones del objeto en las imágenes.
3. Lograr que el algoritmo del proceso de seguimiento sea lo suficientemente rápido para que su implementación en un sistema permita una eficiente detección de movimiento en una secuencia de imágenes.
4. Contemplar en las actualizaciones de rotación el movimiento simultáneo del objeto y la cámara.

Analizando en la literatura casos con o sin traslación del sistema de visión, nos da perspectiva de posibles aplicaciones a las que nos podemos enfrentar en un futuro. Un sistema de visión ubicado como los ojos de un robot que tuviera la capacidad de detectar, reconocer y seguir posibles objetos o personas sería entrenado con un algoritmo como el que proponemos para después llevar a cabo infinidad de tareas encomendadas. Una perspectiva a futuro indica que la relación persona-robot cada día va tomando más fuerza, y por tanto, manejar detección basado en métodos de detección como el propuesto en este trabajo nos daría la capacidad de reconocer personas u objetos utilizando clasificación de color en secuencias de imágenes captadas por el sistema de visión del robot. En este escenario es claro que las posibilidades de desplazamiento de un robot nos obligaría a considerar traslación del sistema de visión del robot.

En otro ámbito mencionado anteriormente, la generalidad de sistemas de vigilancia con cámaras no permite movimiento de las mismas, por lo que en esta situación la única capacidad de rotación de la cámara a considerar es de interés.

El algoritmo de calibración presentado en este trabajo está vinculado a los métodos que requieren la capacidad de seguir características entre imágenes considerando movimiento de la cámara y la escena. Para la detección y seguimiento de características de las imágenes se usa un método de clasificación de color que describimos a fondo en el Capítulo 3, donde utilizamos el espacio de color *HSI* (*H*=Matiz, *S*=saturación e

I =intensidad) y sus diferentes canales como patrones para la clasificación. La detección y seguimiento de objetos o personas se realiza de acuerdo al color de la piel, para lo cual, como se demuestra más adelante, el espacio *HSI* es el más apropiado.

Una vez extraída la información de una secuencia de imágenes, se produce una estimación de los parámetros extrínsecos de la cámara, las cuales son actualizados dinámicamente por medio de métodos iterativos en la resolución de un sistema lineal que contiene la información tanto de la escena 3-D como de la secuencia de imágenes en 2-D.

1.2. OBJETIVO DE LA TESIS.

Realizar y definir un esquema de detección de un objeto en movimiento basado en el reconocimiento de color en una secuencia de imágenes, que proporcione información a un método de calibración con el propósito de seguir el objeto. Así como desarrollar un proceso que permita realizar el seguimiento del objeto. El sistema de visión debe ser capaz de apreciar el movimiento del objeto y seguirlo hasta donde su campo de visión le sea posible.

1.3. METODOLOGÍA.

La metodología desarrollada en este trabajo consiste en ubicar la posición y orientación de la cámara con respecto a un solo objeto en movimiento. Para detectar el objeto se exploró un algoritmo de clasificación de color basado en estimar probabilidades de pertenencia como es el algoritmo de agrupamiento *K-medias*, el cual fue alimentado con patrones de clasificación basados en el espacio *HSI*, con la misión de discriminar un determinado patrón de color. Para nuestro objetivo hemos determinado el color de piel con su gran variedad de tonalidades como conjunto de interés a detectar en la secuencia de imágenes y con ello ubicar las regiones de cada imagen con las características de color así definidas. Posteriormente, se llevó a cabo el reconocimiento de la o las regiones capturadas para identificar si es o no el objeto que buscamos. Esto último se realizó partiendo de un

modelo predefinido del objeto el cual nos permitió determinar la posición del objeto en el espacio. Lo siguiente fue relacionar los puntos 3-D del modelo con los puntos imagen y calcular la matriz de rotación R y el vector de traslación T por medio de un método iterativo de predicción. Esto se hizo por medio de actualizar estimaciones de los puntos 3-D en las coordenadas de la cámara y proyectándolos en un plano imagen que define la diferencia entre éste y la imagen original, hasta que la diferencia entre éstos sea cero o lo suficientemente pequeña. La calibración se realizó con la base de mantener el eje óptico de la cámara siempre en el centro del objeto y con ello en el centro de su campo de visión hasta donde la capacidad de rotación de la cámara le sea posible.

1.4. PROPUESTA DE TESIS.

De manera global, la tesis plantea realizar la calibración de una cámara en base a la información obtenida por la detección y seguimiento de un objeto, y el modelo predefinido del mismo objeto. Así, el método involucra dos procesos diferentes: (a) la detección del objeto, y (b) la calibración de la cámara.

La detección del objeto se realiza a través del algoritmo de clasificación de datos *K-medias*, el cual se aplica usando patrones de color en el espacio *HSI* para realizar la clasificación, que en nuestro caso es para sólo dos clases. Este proceso se llevó a cabo para cada una de las imágenes que componen una secuencia de video. Posteriormente, se realizó la calibración de la cámara, la cual proporciona la información requerida para ubicar el objeto y orientar la cámara a la posición en la que se encuentre. Ejecutando la detección en la imagen siguiente en base a la información de la imagen previa, además de actualizar la orientación de la cámara por medio de la calibración podemos realizar el seguimiento del objeto.

El algoritmo del método propuesto se resume en los siguientes pasos:

1. Ubicación fija de la cámara con la capacidad de rotar en forma horizontal y vertical.
2. Iniciar el proceso de detección del objeto que consiste en:
 - a) Análisis de cada imagen de una secuencia (video) captada por la cámara.
 - b) Se toma la primera imagen de la secuencia y se le aplica el método de clasificación *k-medias* en base al color de la imagen, la cual es analizada en el espacio de color *HSI*, de donde tomamos cada canal del mismo como patrones para la clasificación.
 - c) La clasificación se lleva a cabo con la premisa de ubicar objetos con un color determinado, que en nuestro caso hemos decidido como color de interés el color de la piel y sus diferentes tonalidades.
 - d) Si en la primera imagen no se detecta ningún objeto, se toma la siguiente imagen de la secuencia, y así sucesivamente hasta encontrar una imagen que contenga un objeto con el color que hemos asignado a detectar.
3. Una vez localizado el objeto, se procede a la calibración de la cámara, la que se restringe a la estimación de los parámetros extrínsecos de la cámara, ya que suponemos que los parámetros intrínsecos son conocidos. Lo anterior permite ubicar el objeto en el espacio y con ello se tiene la información completa tanto de la escena, la imagen captada y los parámetros buscados de la cámara.
4. El proceso incluye la actualización de los parámetros de la cámara a medida que el objeto detectado o la cámara se encuentren en movimiento, por lo que se tomará en cuenta el seguimiento del objeto y la historia del movimiento, es decir, la trayectoria que este tome.

1.5. ORGANIZACIÓN DE LA TESIS.

En este Capítulo se dio una introducción sobre el problema de la calibración de una cámara, se definió de una forma clara y concisa en que consiste este proceso, así también se

planteó el problema desde nuestro punto de vista. Se atacó el problema usando un método de clasificación de color basado en el modelo *HSI* aplicado a una secuencia de imágenes tomadas por la cámara. Por último se mostró un resumen del algoritmo completo con las suposiciones y restricciones principales, además de definir cual constituye nuestro objetivo al final del estudio.

En el siguiente Capítulo se analiza más a fondo el proceso de calibración, así como, las técnicas principales de este tipo de procesos. Se analiza algunos de los trabajos anteriores realizados sobre el campo de la calibración de sistemas de visión en los que se comenta puntos de vista sobre los mismos, así como sus ventajas y desventajas, ligando información de las técnicas y datos de estudio llevadas a cabo con el trabajo aquí desarrollado.

En el tercer Capítulo se describe el proceso aplicado para la detección del objeto basado en un método de clasificación de color. Se detalla como se trabajó en el espacio de color *HSI*, el cual es más tolerante a la variación de luz y con mejores propiedades para lograr los objetivos planteados. Aquí también se revisan trabajos anteriores respecto a estos tópicos y de igual forma se relacionan a este trabajo.

Posteriormente, el cuarto Capítulo se centra en desarrollar un algoritmo de reconocimiento de objetos que consiste en inspeccionar la o las regiones en una imagen detectadas por el proceso anterior de detección de color e identificar cual es el que reconocemos como el objeto buscado. Una vez realizado lo anterior detallamos la calibración de la cámara.

Al final se integra toda la información de cada uno de los procesos para realizar la implementación del seguimiento del objeto. Los resultados y las situaciones de prueba se presentan en el Capítulo 5. En las conclusiones se describen los alcances y las limitaciones del trabajo realizado, así como posibles desarrollos futuros.

2. CALIBRACIÓN DE UNA CÁMARA.

2.1. DEFINICIÓN DE CALIBRACIÓN.

La idea principal de calibrar una cámara es escribir las ecuaciones de proyección uniendo las coordenadas conocidas de un conjunto de puntos en 3-D con sus correspondientes proyecciones, y resolverlas para los parámetros de la cámara. En una o más imágenes se toma un patrón de calibración, que es un objeto en 3-D de geometría conocida y posiblemente ubicado en una posición también conocida en el espacio y generar características de imagen que se pueden ubicar con precisión [Trucco y Verri, 1998]. La calibración implica el diseño de un modelo matemático para estimar los parámetros extrínsecos e intrínsecos de la cámara dadas imágenes de un patrón de calibración, y su precisión depende de que tan eficazmente ubiquemos los puntos en el espacio y los puntos de referencia de la cámara.

Los parámetros de calibración de una cámara se dividen en 2 tipos:

a) Los Parámetros Extrínsecos. Estos identifican la orientación y posición de la cámara con respecto a las coordenadas mundiales son:

- i) Vector de Traslación T definido en el espacio describiendo las posiciones relativas de los orígenes de los dos marcos de referencia, el de la cámara y las mundiales.
- ii) Matriz de Rotación R trayendo los correspondientes ejes de los dos marcos uno sobre el otro, con la condición de ser una matriz ortogonal ($R^T R = R R^T = I$).

En la figura 2.1 se muestran los parámetros extrínsecos de una cámara en un modelo en proyección perspectiva en donde se identifican a (X^w, Y^w, Z^w) como los ejes en coordenadas mundiales, (X^c, Y^c, Z^c) como los ejes en coordenadas de la cámara y (x, y) las coordenadas en el plano de la imagen. Se aprecia que w_o denota la transformación de (R, T)

de los parámetros extrínsecos para alinear los ejes de coordenadas mundiales y de la cámara, es decir trasladar el origen de las coordenadas mundiales con el origen de las coordenadas de la cámara y la rotación para alinear los respectivos ejes.

b) Los Parámetros Intrínsecos. Estos caracterizan las propiedades ópticas, geométricas y digitales de la visión de la cámara que son necesarias para unir las coordenadas en píxeles de un punto imagen con las correspondientes coordenadas en el marco de la cámara. Estos parámetros se muestran en la figura 2.1 y son:

- i) *Longitud Focal f* que es la distancia del plano imagen al centro de la lente.
- ii) *Coordenadas del centro de imagen o punto principal (O_x, O_y)* que intersecta el eje óptico de la cámara y el plano imagen.
- iii) *Tamaño efectivo del píxel en dirección horizontal y vertical S_x, S_y* (en milímetros), que dan el aspecto de una escena tomada según el escalamiento y los elementos receptores de la cámara. También es conocida como *la razón de aspecto α* , que es igual al tamaño del píxel en dirección vertical entre el tamaño del píxel en dirección horizontal.
- iv) *Coefficiente de distorsión radial $K1$* que se presenta debido a que por la naturaleza de la lente o del proceso de adquisición de una imagen al captar objetos formados por líneas rectas, estas aparecen en la imagen como líneas curvas, es decir representa el desplazamiento radial dependiendo de la calidad del lente de la cámara usado y la distancia del punto en el espacio al centro de imagen. La distorsión es un fenómeno no deseable y más en modelos geométricos.

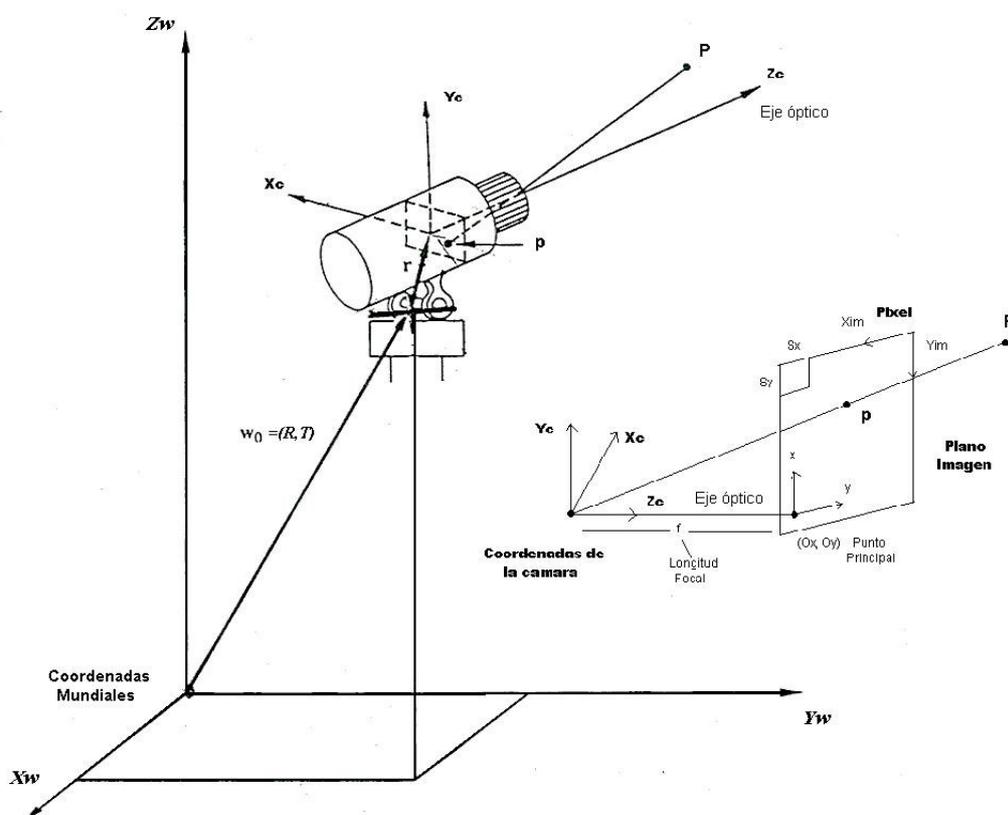


Fig. 2.1. Modelo de cámara en proyección perspectiva [Fu et al, 1987].

Ahora describiremos el aspecto geométrico de la formación de una imagen. La meta es unir la posición de puntos escena con sus correspondientes puntos imagen. Para hacer lo anterior es necesario modelar la proyección geométrica ejecutada por la cámara.

La cámara con perspectiva.

El modelo basado en la proyección perspectiva consiste de un plano imagen π y un punto O en 3-D, el foco o centro de proyección (ver figura 2.2). La distancia entre π y O es la longitud focal f . La línea a través de O y perpendicular a π es el eje óptico y o , la intersección entre π y el eje óptico es el llamado punto principal o centro de imagen.

Como se ve en la figura 2.2, la imagen de P es el punto al cual la línea recta a través de P y O intersecta al plano imagen π . Considerando el marco de referencia 3-D donde O es el origen y el plano π es ortogonal al eje Z y sea $P = [X, Y, Z]^T$ y $p = [x, y, z]^T$. El marco de referencia 3-D es el conocido como marco de la cámara.

Las ecuaciones de proyección perspectiva en el marco de la cámara son:

$$x = f \frac{X}{Z} \quad (2.1)$$

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

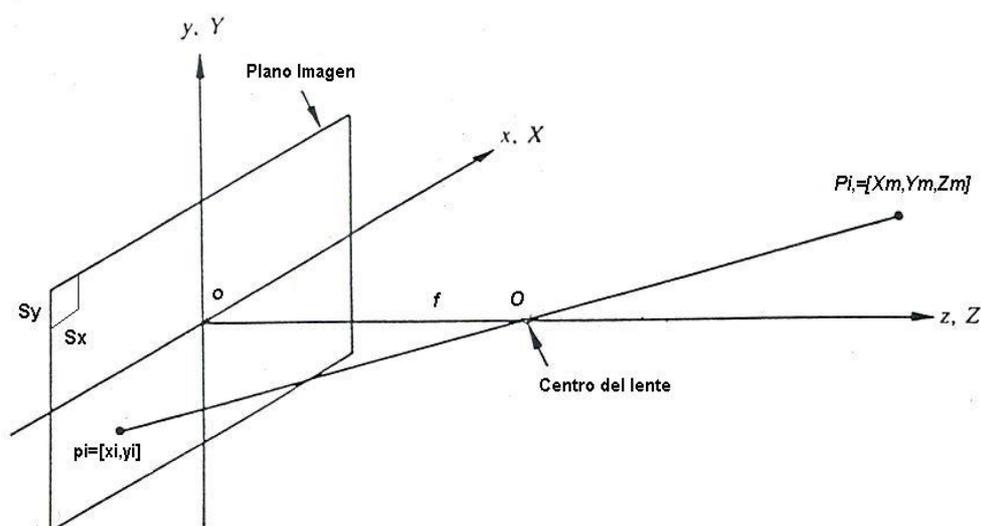


Fig. 2.2. Proyección perspectiva [González y Woods, 1987].

La cámara con perspectiva débil.

Este modelo requiere que la distancia relativa a lo largo del eje óptico de un par de puntos escena sea mucho menor que la distancia promedio Z' de los puntos de la vista de la cámara. Así, las ecuaciones básicas de proyección perspectiva son:

$$x = f \frac{X}{Z} = f \frac{X'}{Z'} \quad (2.2)$$

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

Estas ecuaciones describen una secuencia de dos transformaciones: una proyección ortográfica (figura 2.3) en la que los puntos del espacio son proyectados a lo largo de rayos paralelos al eje óptico y que es:

$$x = X \quad (2.3)$$

$$y = Y$$

seguida de un escalamiento isotrópico por el factor f/Z' . Nótese en la figura 2.3 que como resultado de las transformaciones, la figura de la izquierda es una versión escalada de la derecha.

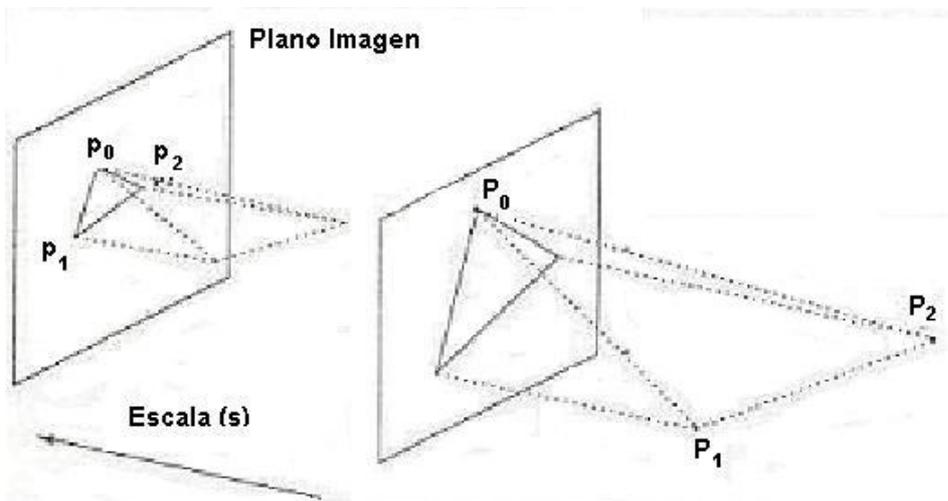


Fig. 2.3 Puntos objeto P_0 , P_1 y P_2 , y sus imágenes bajo perspectiva débil p_0 , p_1 , y p_2 , generada por una proyección ortogonal seguida por un factor de escalamiento [Trucco y Verri, 1998].

2.2. METODOS DE CALIBRACIÓN.

En esta sección se discuten dos métodos clásicos para calibración de una cámara. Ambos métodos que son los más populares dentro de la literatura y se encuentran dentro de aquellos que realizan el proceso basados en una o más imágenes de un patrón de calibración. Obviamente la precisión de ambos métodos va de la mano con la precisión de las mediciones realizadas del patrón de calibración, además de que la calibración de la cámara implica la estimación de sus parámetros intrínsecos y de su orientación y posición. *Calibración directa*, recupera directamente los parámetros extrínsecos e intrínsecos de la cámara mientras *calibración basada en la matriz de proyección* [Trucco E. y Verri A., 1998], calcula los parámetros de la cámara como funciones en lazo cerrado de las entradas de una matriz de proyección. Existen otros métodos que a diferencia de los dos anteriores suponen que los parámetros intrínsecos de la cámara son conocidos, por lo que se enfocan en hacer únicamente la estimación de su posición y orientación.

2.2.1. Calibración Directa.

Este método recupera directamente los parámetros extrínsecos e intrínsecos de la cámara y se define en principio en términos geométricos para posteriormente expresar nuestros resultados en los valores de los parámetros de la cámara.

Consideremos el punto $P^w = [X^w, Y^w, Z^w, 1]^T$ en 3-D en el marco de referencia mundial (ver figura 2.1). Sean $P^c = [X^c, Y^c, Z^c, 1]^T$ el punto P en el marco de referencia de la cámara. ($Z^c > 0$, si P es visible). Usualmente el origen del marco de la cámara está en el centro de proyección y el eje Z es el eje óptico. Se desconocen la matriz R de rotación y el vector T de traslación, tales que unen los marcos de referencia mundial conocido y el marco de la cámara desconocido, de la forma:

$$[X^c, Y^c, Z^c]^T = R * [X^w, Y^w, Z^w]^T + T^T \quad (2.4)$$

y escritas en componentes son:

$$\begin{aligned}
X^c &= r_{11}X^w + r_{12}Y^w + r_{13}Z^w + T_x \\
Y^c &= r_{21}X^w + r_{22}Y^w + r_{23}Z^w + T_y \\
Z^c &= r_{31}X^w + r_{32}Y^w + r_{33}Z^w + T_z
\end{aligned} \tag{2.5}$$

Escribimos la imagen de $[X^c, Y^c, Z^c]^T$ en 3-D en el sistema de coordenadas de la imagen como:

$$x_{im} = -f \frac{X^c}{S_x Z^c} + o_x \tag{2.6}$$

$$y_{im} = -f \frac{Y^c}{S_y Z^c} + o_y$$

Juntando (2.4), (2.5) y (2.6) obtenemos:

$$x_{im} - o_x = -f \frac{(r_{11}X^w + r_{12}Y^w + r_{13}Z^w + T_x)}{(r_{31}X^w + r_{32}Y^w + r_{33}Z^w + T_z)} \tag{2.7}$$

$$y_{im} - o_y = -f \frac{(r_{21}X^w + r_{22}Y^w + r_{23}Z^w + T_y)}{(r_{31}X^w + r_{32}Y^w + r_{33}Z^w + T_z)}$$

Así pasamos del inaccesible marco de la cámara y unimos directamente las coordenadas mundiales con sus coordenadas correspondientes en la imagen, y usando un conocido patrón de calibración ambos vectores son conocidos. Lo anterior sugiere que dado un número de puntos sobre el patrón de calibración se resuelve (2.7) para los parámetros de la cámara, lo que llevamos a cabo en dos partes:

- a) Asumir coordenadas del centro de imagen conocidas y estimar los parámetros restantes.
- b) Encontrar las coordenadas del centro de imagen.

Ahora de (2.7), como ambas ecuaciones tienen el mismo denominador podemos escribir una sola de la forma:

$$x_i f \left[r_{21} X^w + r_{22} Y^w + r_{23} Z^w + T_y \right] = -y_i f \left[r_{11} X^w + r_{12} Y^w + r_{13} Z^w + T_x \right] \quad (2.8)$$

Mientras $\alpha = 1$ se puede tomar como una ecuación lineal para las ocho incógnitas $v = (v_1, v_2, \dots, v_8)$:

$$x_i X_i^w v_1 + x_i Y_i^w v_2 + x_i Z_i^w v_3 + x_i v_4 - x_i X_i^w v_5 - y_i Y_i^w v_6 - y_i Z_i^w v_7 - y_i v_8 = 0 \quad (2.9)$$

donde

$$v_1 = r_{21} \quad v_2 = r_{22} \quad v_3 = r_{23} \quad v_4 = T_y \quad v_5 = \alpha r_{11} \quad v_6 = \alpha r_{12} \quad v_7 = \alpha r_{13} \quad v_8 = \alpha T_x$$

La última ecuación para los N correspondientes pares, lleva al sistema homogéneo de N ecuaciones lineales $Av=0$ con la matriz A dada por:

$$A = \begin{bmatrix} x_1 X_1^w & x_1 Y_1^w & x_1 Z_1^w & x_1 & -y_1 X_1^w & -y_1 Y_1^w & -y_1 Z_1^w & -y_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ x_8 X_8^w & x_8 Y_8^w & x_8 Z_8^w & x_8 & -y_8 X_8^w & -y_8 Y_8^w & -y_8 Z_8^w & -y_8 \end{bmatrix} \quad (2.10)$$

Si $N > 7$ y los N puntos son no coplanares (puntos que no se encuentran en el mismo plano), A tiene rango 7 y el sistema tiene solución no trivial la cual puede ser determinada de la descomposición en valores singulares (SVD) de A , $A=UDV^T$ como la columna de V correspondiente al único valor singular nulo a lo largo de la diagonal de D .

La meta es determinar el factor de escala γ desconocido del vector solución $v = \bar{v}$.

$$\bar{v} = \gamma (r_{21}, r_{22}, r_{23}, T_y, \alpha r_{11}, \alpha r_{12}, \alpha r_{13}, \alpha T_x) \quad (2.11)$$

como $r_{21}^2 + r_{22}^2 + r_{23}^2 = 1$, de las primeras tres componentes de \bar{v} se obtiene:

$$[\bar{v}_1^2 + \bar{v}_2^2 + \bar{v}_3^2] = [r_{21}^2 + r_{22}^2 + r_{23}^2] = |\gamma| \quad (2.12)$$

y similarmente $r_{11}^2 + r_{12}^2 + r_{13}^2 = 1$ y $\alpha > 0$ de la quinta, sexta y séptima componentes de v' se obtiene :

$$[\bar{v}_5^2 + \bar{v}_6^2 + \bar{v}_7^2] = [r_{11}^2 + r_{12}^2 + r_{13}^2] = \alpha |\gamma| \quad (2.13)$$

y se resuelve para $|\gamma|$ y para la razón de aspecto α .

El tercer renglón de la matriz R puede obtenerse del producto vectorial de los primeros dos renglones. Esto implica que el signo del tercer renglón ya está fijo como las entradas del tercer renglón permanecen sin cambio si los signos de todas las entradas de los primeros dos renglones son cambiados.

Finalmente se determina el signo del factor de escala γ para concluir la estimación de los parámetros. Para ello se regresa a la ecuación (2.7) y se renombra para todo punto $Z^c > 0$ y entonces x y $r_{11} * X^w + r_{12} * Y^w + r_{13} * Z^w + T_x$ deben tener signo opuesto. Consecuente con esto es checar el signo de $x (r_{11} * X^w + r_{12} * Y^w + r_{13} * Z^w + T_x)$ para uno de los puntos.

Si $x(r_{11} * X^w + r_{12} * Y^w + r_{13} * Z^w + T_x) > 0$ los signos de los dos primeros renglones de R , de las primeras dos componentes de T y de α deben ser cambiados. De otra forma no se hará ningún cambio. Se puede llevar un argumento similar para y y $x (r_{21} * X^w + r_{22} * Y^w + r_{23} * Z^w + T_x)$.

A este punto se encontraron la matriz R , las primeras dos componentes de T y α . Solo falta encontrar T_z y la longitud focal en unidades de píxel f_x que se pueden encontrar

por mínimos cuadrados de un sistema de ecuaciones como (2.7) escrita para N puntos. Para esto se escribe:

$$x_i [x(r_{31} * X^w + r_{32} * Y^w + r_{33} * Z^w + T_x)] = -fx [x (r_{11} * X^w + r_{12} * Y^w + r_{13} * Z^w + T_x)] \quad (2.14)$$

y se resuelve el sistema de N ecuaciones lineales $A(T_z, fx)^T = b$ para las 2 incógnitas T_z y fx donde:

$$A = \begin{bmatrix} x_1 & r_{11}X_1^w + r_{12}Y_1^w + r_{13}Z_1^w + T_x \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ x_8 & r_{11}X_8^w + r_{12}Y_8^w + r_{13}Z_8^w + T_x \end{bmatrix} \quad \text{y} \quad b = \begin{bmatrix} -x_1(r_{11}X_1^w + r_{12}Y_1^w + r_{13}Z_1^w + T_x) \\ \cdot \\ \cdot \\ \cdot \\ -x_8(r_{11}X_8^w + r_{12}Y_8^w + r_{13}Z_8^w + T_x) \end{bmatrix} \quad (2.15)$$

2.2.2. Calibración obtenida de la matriz de Proyección.

El método consiste en dos etapas secuenciales:

- Estimar la matriz de proyección uniendo coordenadas mundiales y coordenadas imagen.
- Calcular los parámetros de la cámara como funciones en lazo cerrado de las entradas de la matriz de proyección.

La relación entre las coordenadas 3-D $[X^w, Y^w, Z^w, 1]^T$ de un punto en el espacio y las coordenadas 2-D (x, y) de su proyección en el plano imagen puede escribirse por medio de la matriz de proyección M de acuerdo a la ecuación:

$$(u, v, w)^T = M * [X^w, Y^w, Z^w, 1]^T \quad (2.16)$$

con:

$$x = \frac{u}{w} = \frac{m_{11}X^w + m_{12}Y^w + m_{13}Z^w + m_{14}}{m_{31}X^w + m_{32}Y^w + m_{33}Z^w + m_{34}} \quad (2.17)$$

$$y = \frac{v}{w} = \frac{m_{21}X^w + m_{22}Y^w + m_{23}Z^w + m_{24}}{m_{31}X^w + m_{32}Y^w + m_{33}Z^w + m_{34}}$$

y M definida sobre un factor arbitrario de escala y tiene 11 entradas independientes las que pueden determinarse a través de un sistema lineal homogéneo por escribir la ecuación anterior para al menos 6 correspondencias de puntos mundiales-imagen. Asumiendo N correspondencias para el sistema se tiene $Am = 0$ con

$$A = \begin{bmatrix} X_1^w & Y_1^w & Z_1^w & 1 & 0 & 0 & 0 & 0 & -x_1 X_1^w & -x_1 Y_1^w & -x_1 Z_1^w & -x_1 \\ 0 & 0 & 0 & 0 & X_1^w & X_1^w & X_1^w & 1 & -y_1 X_1^w & -y_1 Y_1^w & -y_1 Z_1^w & -y_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ X_N^w & Y_N^w & Z_N^w & 1 & 0 & 0 & 0 & 0 & -x_N X_N^w & -x_N Y_N^w & -x_N Z_N^w & -x_N \\ 0 & 0 & 0 & 0 & X_N^w & Y_N^w & Z_N^w & 1 & -x_N X_N^w & -x_N Y_N^w & -x_N Z_N^w & -y_N \end{bmatrix} \quad (2.18)$$

y $m = [m_{11}, m_{12}, \dots, m_{33}, m_{34}]$

Mientras A tenga rango 11, el vector m puede obtenerse de SVD y siendo m , la columna de V correspondiente al cero o valor singular más pequeño de A .

Ahora lo que se hace es expresar los parámetros extrínsecos e intrínsecos como funciones de la matriz de proyección estimada. Para evitar confusión se llama \hat{M} a la matriz estimada y con ello \hat{m}_{ij} los elementos de \hat{M} . Rescribiendo la expresión completa para las entradas de M :

$$M = \begin{bmatrix} -f_x r_{11} + o_x r_{31} & -f_x r_{12} + o_x r_{32} & -f_x r_{13} + o_x r_{33} & -f_x T_x + o_x T_z \\ -f_y r_{21} + o_y r_{31} & -f_y r_{22} + o_y r_{32} & -f_y r_{23} + o_y r_{33} & -f_y T_y + o_y T_z \\ r_{31} & r_{32} & r_{33} & T_z \end{bmatrix} \quad (2.19)$$

Ahora se necesita definir vectores 3-D de la forma:

$$\begin{aligned}
 \mathbf{q}_1 &= [\hat{m}_{11}, \hat{m}_{12}, \hat{m}_{13}]^T \\
 \mathbf{q}_2 &= [\hat{m}_{21}, \hat{m}_{22}, \hat{m}_{23}]^T \\
 \mathbf{q}_3 &= [\hat{m}_{31}, \hat{m}_{32}, \hat{m}_{33}]^T \\
 \mathbf{q}_4 &= [\hat{m}_{41}, \hat{m}_{42}, \hat{m}_{43}]^T
 \end{aligned} \tag{2.20}$$

Mientras M es definida sobre un factor de escala escribimos $\hat{M} = \hat{M} \gamma$. El valor absoluto de $|\gamma|$ se obtiene notando que q_3 es el último renglón de la matriz de rotación. Entonces:

$$(\hat{m}_{31}^2 + \hat{m}_{32}^2 + \hat{m}_{33}^2)^{1/2} = |\gamma| (r_{31}^2 + r_{32}^2 + r_{33}^2)^{1/2} = |\gamma| \tag{2.21}$$

del último renglón de (2.19) se tiene

$$Tz = \sigma \hat{m}_{34} \quad y \quad r_{3i} = \sigma \hat{m}_{3i}, \quad i = 1, 2, 3 \quad \text{con } \sigma = \pm 1 \tag{2.22}$$

tomando los productos punto de q_3 con q_1 y q_2 se encuentra:

$$o_x = q_1^T q_3 \quad y \quad o_y = q_2^T q_3 \tag{2.23}$$

se determina f_x y f_y de la forma:

$$\begin{aligned}
 f_x &= (q_1^T q_1 - o_x)^{1/2} \\
 f_y &= (q_2^T q_2 - o_y)^{1/2}
 \end{aligned} \tag{2.24}$$

y se calcula los parámetros extrínsecos restantes como:

$$\begin{aligned}
 r_{1i} &= \sigma(o_x * \hat{m}_{3i} - \hat{m}_{1i})/f_x, \quad i = 1, 2, 3 \\
 r_{2i} &= \sigma(o_y * \hat{m}_{3i} - \hat{m}_{2i})/f_y, \quad i = 1, 2, 3 \\
 T_x &= \sigma(o_x * T_z - \hat{m}_{14})/f_x \\
 T_y &= \sigma(o_y * T_z - \hat{m}_{24})/f_y
 \end{aligned} \tag{2.25}$$

Lo restante es calcular la matriz ortogonal más cercana a la matriz de rotación \widehat{R} , y el signo de σ que se obtiene de $T_z = \sigma \widehat{m}_{34}$, debido a que se sabe que en todo caso, el origen del marco de referencia mundial se encontrará en frente ($T_z > 0$) o detrás ($T_z < 0$) de la cámara.

2.3. TRABAJO RELACIONADO.

Los métodos de calibración son divididos en dos categorías:

1. Calibración usando un patrón de calibración, que consiste de un objeto en 3-D de geometría conocida y su posición igualmente conocida en el espacio.
2. Métodos de auto-calibración, en los que se realiza un seguimiento de características en una imagen y posteriormente se obtienen sus coordenadas mundiales.

De ambos tipos de métodos, el primero de ellos requiere mediciones largas y tediosas, así como, configuraciones cuidadosamente controladas y el segundo necesita identificar información característica de una imagen para llevar a cabo la estimación de los parámetros del sistema de visión.

La calibración implica actualizar los parámetros de la cámara tomando algunos puntos de referencia, pero si se da el caso que tenemos puntos de referencia variables, es decir hay movimiento en mi escena o en la cámara, el problema de calibración se hará con base en la detección y seguimiento del movimiento. Tomando lo anterior se puede hacer una subdivisión de los métodos de calibración basada en la detección y seguimiento de objetos fijos y de aquellos que lo realizan para objetos en movimiento. La calibración orientada a objetos móviles implica mayor complejidad, ya que se tienen que considerar aspectos como variaciones de luz más considerables, cambios relativos entre cámara y escena y posibles deformaciones del objeto, tanto en posición como en forma. Aquí se comentan algunos trabajos anteriores en calibración, haciendo énfasis en la subdivisión de los métodos arriba mencionada.

2.3.1. Calibración con objetos fijos.

[Fremont y Chellali, 2002] describieron un método de calibración usando dos círculos concéntricos y basado en las propiedades de la elipse. Se estiman los centros de los círculos proyectados en perspectiva y de acuerdo a las propiedades elípticas de estos círculos se calculan la posición, orientación y foco de la cámara.

La detección de las elipses en una imagen se realiza por medio de la transformada de Hough y se determinan los coeficientes de la ecuación de la elipse. Estos coeficientes estarán en función de los parámetros de la cámara y el centro del círculo proyectado para formar un sistema de ecuaciones de los N puntos imagen de la elipse estimada. Con ello obtienen el foco, orientación y posición de la cámara.

El método también es útil para el problema de reconstrucción 3-D de una secuencia de imágenes de un objeto ubicado sobre una placa giratoria usada como patrón de calibración. Un complemento importante es que se incluyen posibles variaciones o errores en el proceso de adquisición de la imagen. El principal obstáculo expuesto es que sólo se toma una imagen, a diferencia de la mayoría de los trabajos actuales, que estudian secuencias de imágenes. Esto evita que no se consideren posibles variaciones al momento de la detección. Las perspectivas futuras de este método serán lograr la estimación del punto principal y considerar distorsiones radiales de la lente de la cámara.

[Zhang y Schenk, 1997] proponen auto calibración de sistemas de visión estéreo usando la información disponible en el espacio 3-D y de datos en el tiempo sin usar ningún aparato de calibración particular. Se contempla inicialmente un método de calibración clásico, a medida que se presentan cambios en el medio o en algunas de las cualidades ópticas de la cámara se realiza otro proceso independiente del usado inicialmente. En ese primer método, la idea es usar matrices de proyección de la cámara y correspondencia entre puntos imagen para alimentar la estructura euclidiana de la imagen previa, lo que permite realizar la recalibración del sistema de visión estéreo. Las incertidumbres son

sistemáticamente manipuladas y mantenidas para realizar cálculos más precisos de las matrices de proyección.

Es importante mencionar trabajos como este en el que se contempla el uso de más de una cámara para posibles extensiones de métodos de calibración de una sola cámara. En principio se muestra el proceso llevado a cabo para un sistema binocular (2 cámaras), pero fácilmente extensible para n cámaras. Una contribución importante en este trabajo es la caracterización de incertidumbres definidas dentro de las matrices de covarianza debidas a errores en la reconstrucción de puntos y posibles direcciones de los mismos. Las pruebas realizadas muestran un grado de error aceptable tanto en parámetros extrínsecos e intrínsecos.

[Eisert, 2002] propone una nueva técnica para la determinación de los parámetros extrínsecos e intrínsecos de la cámara basada en un modelo que consiste en absorber información característica de una imagen que vinculamos con uno o varios objetos de prueba. Después obtiene su representación 3-D, a la vez que compara en todo momento posibles proyecciones en imagen de los puntos del modelo con las imágenes originales capturados por la cámara. El método prevé un proceso iterativo para disminuir el error entre imágenes originales y de proyección del modelo. Se plantean dos posibles estrategias, la primera hacer sólo la estimación de los parámetros extrínsecos y la segunda de los dos tipos de parámetros. Así, el patrón de calibración utilizado no se restringe a mostrar simples características discretas detectables fácilmente, y sí pueden contemplarse objetos con un amplio rango de frecuencias espaciales.

Una búsqueda grande del error en este tipo de método para establecer correspondencias de los puntos característicos de una imagen se evita al considerar la posición anterior del objeto de prueba, con ello se simplifica en mucho los cálculos computacionales, que en la práctica de este proceso es vital para mostrar su eficiencia. Se muestra una buena precisión para imágenes reales o sintéticas, y mejora aún más al desarrollarse en múltiples imágenes.

2.3.2. Calibración con objetos en movimiento.

[Dron, 1993] particularizó el movimiento de una cámara y estructuras 3-D de secuencias de imágenes basado en seguimiento de características. Describe un método basado en un algoritmo de mínimos cuadrados para calcular la calibración de una cámara de una serie de secuencias de movimiento y cuya información produce estimación robusta de la matriz de calibración y se va actualizando dinámicamente de acuerdo a las nuevas mediciones adquiridas.

El considerar que la dirección de traslación de la cámara es conocida reduce en mucho la problemática de calibración y la precisión del proceso se basa en la precisión de la estimación de este parámetro. Los resultados experimentales mostraron buena precisión con respecto a la orientación de la cámara y las correspondencias de los puntos en la imagen con respecto a su localización al centro mostraron un grado de error estándar de alrededor de tres a cuatro píxeles.

[Tay y Parvin, 1998] desarrollaron un sistema simple de seguimiento de movimiento con calibración activa de una cámara. La detección de movimiento se basa en determinar el flujo óptico y agrupando los mismos vectores de movimiento lo que permite un seguimiento basado en correlación. A su vez el método describe como encontrar esquinas adecuadas de una imagen de un medio natural y con ello construyen las ecuaciones adecuadas para estimar los parámetros intrínsecos. Con lo anterior es claro que no se requiere un patrón de calibración, y el siguiente desarrollo consiste en un algoritmo perteneciente a un método de autocalibración. Así, los dos procesos mencionados corren simultáneamente para ir actualizando y corrigiendo el tamaño y ubicación de la región de movimiento detectada.

Este trabajo se distingue de otros aquí mencionados en que emplea una estimación del flujo óptico como una aproximación del campo de movimiento y que se calcula de

secuencias de imágenes variantes en el tiempo. Es importante hacer notar diferencias en las técnicas empleadas para hacer el seguimiento. En este trabajo, a pesar de las considerables restricciones que se hacen, los resultados mostrados son en un rango muy cercano a otros y el grado de error va muy relacionado en donde se ubiquen los puntos de referencia en la imagen.

[Tommaselli y Tozzi, 1999] abordaron el problema de calibración dinámica de una cámara considerando un cubo en movimiento constante sobre una banda transportadora usando líneas rectas como referencias con la cámara fija. El modelo se basa en la equivalencia del vector normal al plano de interpretación en el marco de la imagen y el vector normal al plano de interpretación rotado en el marco del objeto. Para resolver la calibración dinámica de una cámara se aplica un proceso iterativo basado en el filtro de *Kalman* usando los parámetros de orientación de la cámara secuencialmente estimados para retroalimentar el proceso de extracción de características de la imagen.

La solución propuesta se implementó usando datos reales del cubo moviéndose. Las restricciones que deben resaltarse son que únicamente se considera una única trayectoria y movimiento constante, por lo que estudiar variaciones en el movimiento y trayectoria sería una mejora sustancial al problema. Además se resalta que dentro de las pruebas, los parámetros de traslación no son muy confiables y los parámetros intrínsecos deben considerarse más precisos y con cámaras con mejor resolución. De igual forma no se aplicaron técnicas de procesamiento de imagen muy efectivas para poder alcanzar resultados más deseables.

2.4. CONCLUSIONES.

En este Capítulo se definió el término calibración, el cual, es esencial su entendimiento para los intereses de este trabajo; así también se mostró los parámetros extrínsecos e intrínsecos que constituyen la calibración de una cámara y como afectan cada uno de ellos un sistema de visión. Se explicaron a detalle dos métodos clásicos de

calibración y sus diferencias, que nos dan una idea de las herramientas matemáticas necesarias y condicionantes para alcanzar los resultados correctos de una calibración. Hay que remarcar que muchos de los métodos de calibración en la literatura se han desarrollado con la base de obtener los puntos 3-D en el espacio de patrones de calibración, los cuales se han explicado como funcionan en este ámbito y para con ello hacer notar diferencias con respecto a nuestro trabajo. De la misma forma se han revisado antecedentes de la calibración, desde aquellos trabajos que han propuesto el uso de los mencionados patrones de calibración a otros en esencia más complejos al estimar los parámetros de una cámara en base a objetos en movimiento, y que se pretendió hacer énfasis en ellos por encontrar ideas cercanas a la propuesta que se hace en este estudio.

3. CLASIFICADORES DE COLOR.

3.1. ESPACIOS DE COLOR.

El objetivo de un espacio o modelo de color es facilitar la especificación de colores en forma estándar. En esencia un modelo de color es una especificación de un sistema de coordenadas $3-D$ y un subespacio dentro de ese sistema donde cada color es representado por un punto.

La mayoría de los modelos de color son orientados hacia el equipo de un sistema computacional (monitores a color o impresoras) o hacia aplicaciones donde la tarea es la manipulación del color (creación de gráficos a color para animación). El modelo orientado al equipo de un sistema más comúnmente utilizado es el RGB , que es igualmente usado para monitores a color y como la clase de color usada en las cámaras de video. Existe también el modelo de color HSI , sobre el cual está enfocado este trabajo y que es utilizado frecuentemente para realzar diferentes características en imágenes, que no son posibles de apreciar en el espacio RGB .

Aunque existen otros espacios o modelos de color, se explicará únicamente a detalle estos dos, uno por ser el más común y el otro por ser el de interés y utilidad para este trabajo.

3.1.1 Espacio de color RGB.

El modelo de color RGB es el más conocido y frecuentemente utilizado para la especificación de colores. En él, cada color aparece en sus componentes espectrales primarias de rojo (R), verde (G) y azul (B) [González y Woods, 1992]. Este modelo está basado en un sistema de coordenadas cartesianas. El subespacio de color de interés es el cubo mostrado en la figura (3.1), en el cual los valores de RGB se encuentran en las tres

esquinas; cyan, magenta y amarillo están en las otras tres esquinas del cubo; el negro es el origen y el blanco es la esquina más lejana del origen. En este modelo, la escala de grises se extiende de negro al blanco a lo largo de la línea que une esos dos puntos, y los colores son puntos sobre o dentro de la superficie del cubo, definidos por vectores extendiéndose desde el origen.

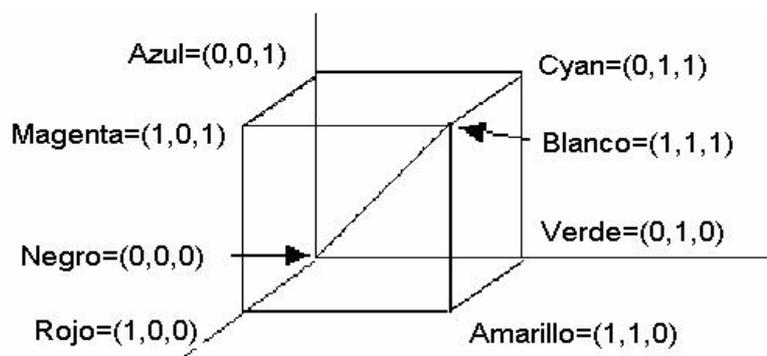


Figura 3.1. Cubo de color RGB.

Las imágenes en el modelo de color *RGB* consisten de tres planos-imagen independientes, uno para cada color primario. Cuando alimentamos en un monitor *RGB*, esas tres imágenes se combinan con la pantalla para producir una imagen compuesta del color (Figura 3.2). Entonces, el uso del modelo *RGB* para procesamiento de imágenes tiene sentido cuando las imágenes mismas son naturalmente expresadas en términos de los tres planos de color. Alternativamente, la mayoría de las cámaras de color usadas para adquirir imágenes digitales utilizan el formato *RGB*, lo cual lo hace un importante modelo en procesamiento de imágenes.



Figura 3.2. Imagen de color RGB y cada uno de sus planos de color.

3.1.2. Espacio de color HSI.

Existe dentro de la variedad de espacios de color, uno utilizado frecuentemente para la manipulación de color en una imagen. Es el conocido como *HSI* (matiz, saturación e intensidad) [González y Woods, 1992]. Este espacio de color surge a partir de que las características generalmente usadas para distinguir un color de otro son brillantez, matiz y saturación, donde brillantez consiste en la noción cromática de intensidad.

El matiz (H) define la frecuencia de la onda de luz. Es un atributo asociado con la longitud de onda dominante en una mezcla de ondas de luz. Entonces el matiz representa como es percibido el color dominante; cuando se llama a un objeto rojo, naranja o amarillo, se especifica su matiz. La asignación general de frecuencias de ondas de luz resulta de la división del ángulo de 360 grados en 3 partes de 120 grados cada uno. El ángulo en 0 grados o equivalentemente el ángulo en 360 grados corresponden al centro del color rojo. El ángulo de 120 corresponde al centro del color verde y el de 240 al centro del color azul. Los otros colores intermedios para su componente (H) son propiamente distribuidos

linealmente. El valor de matiz puede modificarse independientemente de otros elementos que compongan el espacio de color. El cambio de matiz cambia el componente cromático, pero de manera que todos los otros parámetros como saturación e intensidad no dependen de este cambio.

La saturación (S) se refiere a la pureza relativa o cantidad de luz blanca mezclada con el matiz del color. Los espectros de colores puros son completamente saturados. Así pues, la saturación de color como se conoce es la proporción del elemento cromático en el color entero del cual es parte. El valor de saturación puede variar de 0 a 1 y los colores como el rosa (rojo y blanco) y lavanda (violeta y blanco) son menos saturados, y así el grado de saturación es inversamente proporcional a la cantidad de luz blanca añadida.

La intensidad (I) define la intensidad del color y su valor igualmente varia de 0 a 1. Matiz y saturación son conocidos en forma conjunta como la cromaticidad de un color y entonces un color se caracteriza por su brillantez y cromaticidad.

La utilidad que da el uso de este modelo de color *HSI* se debe principalmente a dos factores:

- a) La intensidad I es separada de la información del color de una imagen.
- b) El matiz y la saturación están íntimamente relacionados a la forma en la cual, los seres humanos percibimos el color.

Dentro de las aplicaciones del espacio de color *HSI* pueden mencionar la inspección de la calidad de color de productos terminados, manipulación de muestras de colores en sistemas, diseño de programas de procesamientos de imágenes y la determinación de madurez de frutas y verduras. En las aplicaciones mencionadas la idea de usar el *HSI* es basar el sistema de operación sobre propiedades del color para alcanzar una meta en cuestión.

3.1.3. Conversión de RGB a HSI.

El espacio de color *RGB* se mide con respecto a un cubo unitario, mientras que *HSI* se define con respecto a un triángulo de color [González R.C. y Woods R.E, 1992]. Hay que hacer mención que todos los colores son obtenidos por la combinación de tres colores que están dentro de un triángulo cuyos vértices son definidos por los tres colores iniciales (Figura 3.3 (a)). Entonces el matiz de un punto de color *P* dentro del triángulo es el ángulo del vector que va del centro del triángulo a *P* con respecto al eje rojo. Cuando $H=0^\circ$ el color es rojo, si $H=60^\circ$ el color es amarillo y así sucesivamente. La saturación del punto de color *P*, que es el grado en que el color es diluido por el blanco es inversamente proporcional a la distancia de *P* al centro del triángulo (Figura 3.3 (a)).

La intensidad en el modelo *HSI* es medido, con respecto a una línea perpendicular al triángulo que pasa por su centro (Figura 3.3 (b)). Las intensidades a lo largo de esta línea que se encuentran abajo del triángulo tienden desde oscuro bajo a negro. Por el contrario intensidades arriba del triángulo tienden de luz alta a blanco. *H*, *S* e *I* combinados en un espacio de color 3-*D*, cumplen con la estructura de una pirámide de tres lados.

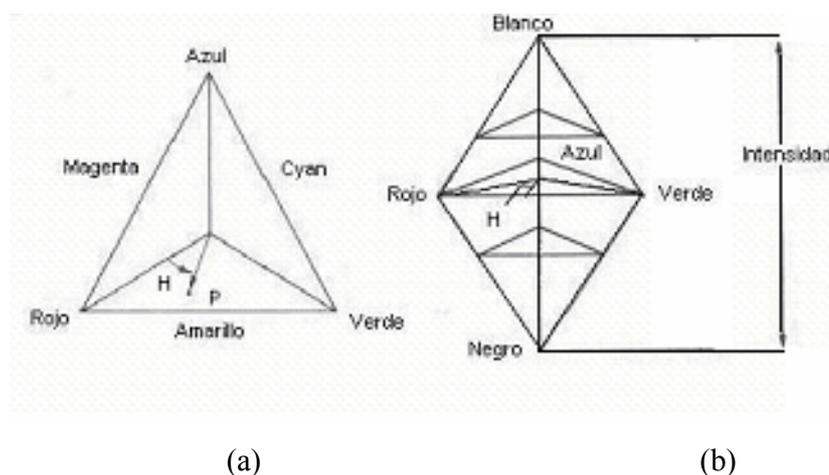


Figura 3.3. (a) Triángulo de color HSI y (b) Estructura de color HSI. [González y Woods, 1992]

Algún punto sobre la superficie de esta estructura representa un color puramente saturado. El matiz (H) de ese color es determinado por el ángulo con respecto al eje rojo y su intensidad por su distancia perpendicular desde el punto negro. De forma similar, aplica a puntos dentro de la estructura, la única diferencia es que los colores llegan a ser menos saturados conforme se vayan aproximando al eje vertical. Los colores en el espacio HSI son definidos con respecto a los valores de rojo, verde y azul normalizados dados en términos de los RGB primarios por:

$$r = \frac{R}{R+G+B} \quad g = \frac{G}{R+G+B} \quad b = \frac{B}{R+G+B} \quad (3.1)$$

donde R , G y B , como han sido normalizados están en el rango de $[0,1]$, por lo tanto, se tiene

$$r+g+b = 1. \quad (3.2)$$

Entonces para algunas componentes R , G y B , cada una en el rango de $[0,1]$, el componente de intensidad en el espacio HSI es definida como:

$$I = \frac{R+G+B}{3} \quad (3.3)$$

El siguiente paso es obtener H y S . Para obtener H requerimos la construcción geométrica del triángulo HSI mostrado en las figura 3.4, del cual se notan las siguientes condiciones:

- El punto W tiene coordenadas $(1/3, 1/3, 1/3)$.
- Un punto-color P arbitrario tiene coordenadas (r,g,b) .
- El vector extendiéndose desde el origen a W es denotado por w . Similarmente los vectores extendiéndose desde el origen a Pr y a P son denotados pr y p respectivamente.

- d) Las líneas P_iQ_i , $i = R, G, B$, intersectan a W por construcción.
- e) Siendo $ro = R/I$, $go = G/I$ y $bo = B/I$, donde I es dado en la ecuación (3.3), vemos de la figura 3.4 (a) que P_RQ_R es el foco de puntos (ro, go, bo) para la cual $go = bo$. Similarmente $ro = go$ a lo largo de P_BQ_B , y $ro = bo$ a lo largo de P_GQ_G .
- f) Algún punto en la región planar acotada por el triángulo $P_RQ_RP_G$ tiene $go \geq bo$. Algún punto en la región acotada por el triángulo $P_RQ_RP_B$ tiene $bo \geq go$. Entonces la línea P_RQ_R separa la región $go > bo$ de la región $go < bo$. Similarmente P_GQ_G separa la región $bo > ro$ de la región $bo < ro$, y P_BQ_B separa la región $go > ro$ de la región $go < ro$.
- g) Para $i = R, G$, o B , $|WQ_i|/|PiQ_i| = 1/3$ y $|WP_i|/|PiQ_i| = 2/3$, donde $|arg|$ denota la longitud del argumento.
- h) Por definición el sector RG es la región acotada por WP_RP_G , el sector GB es la región acotada por WP_GP_B , y el sector BR es la región acotada por WP_BP_R .

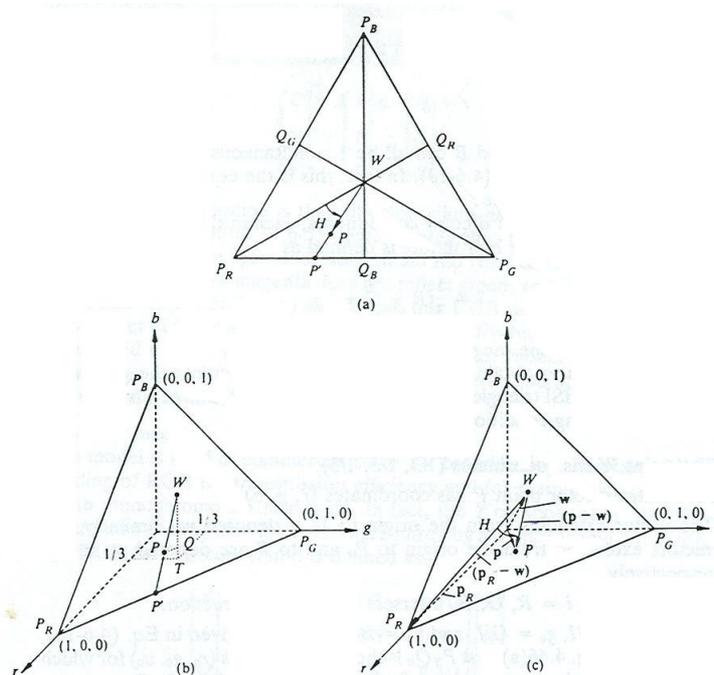


Figura 3.4. Detalles del triángulo de color HSI necesario para la obtención de expresiones para matiz H y saturación S [González R.C. y Woods R.E, 1992].

Con referencia en la figura 3.4 (b) el matiz H de un color arbitrario es definido por el ángulo entre los segmentos de línea WP_R y WP , o en forma vectorial (figura 3.4. (c)), por el ángulo entre los vectores $(P_R - W)$ y $(P - W)$.

En general la siguiente ecuación se cumple para valores entre 0 y 180 grados.

$$(P - W)(P_R - W) = \|P - W\| \|P_R - W\| \cos H \quad (3.4)$$

donde $(x)(y) = x^T y = \|x\| \|y\| \cos H$ denota el producto punto de dos vectores, y la doble barra denota la norma del argumento del vector. El problema ahora es expresar este resultado en términos de un conjunto de RGB primarios. De las condiciones a) y b) se tiene que:

$$\|P - W\| = \left[\left(r - \frac{1}{3}\right)^2 + \left(g - \frac{1}{3}\right)^2 + \left(b - \frac{1}{3}\right)^2 \right]^{1/2} \quad (3.5)$$

ya que la longitud de un vector A con componentes A_1, A_2 y A_3 es $\|A\| = [A_1^2 + A_2^2 + A_3^2]^{1/2}$.

Ahora sustituyendo ecuaciones (3.1) para r, g y b en (3.5) y simplificando nos lleva a:

$$\|P_R - W\| = \left\{ \frac{2}{3} \right\}^{1/2} \quad (3.6)$$

Tomando en cuenta que para dos vectores A y B , $A \bullet B = A^T B = A_1 B_1 + A_2 B_2 + A_3 B_3$, entonces

$$\begin{aligned} (P - W)(P_R - W) &= \frac{2}{3} \left(r - \frac{1}{3}\right) - \frac{1}{3} \left(g - \frac{1}{3}\right) - \frac{1}{3} \left(b - \frac{1}{3}\right) \\ &= \frac{2R - G - B}{3(R + G + B)} \end{aligned} \quad (3.7)$$

De la ecuación (3.4)

$$H = \cos^{-1} \left[\frac{(P-W)(P_R-W)}{(\|P-W\| \|P_R-W\|)} \right] \quad (3.8)$$

Sustituyendo ecuaciones (3.5) y (3.7) en la ecuación (3.8) y simplificando nos lleva a la siguiente expresión para H en términos de R , G y B :

$$H = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}} \right\} \quad (3.9)$$

Esta se cumple para valores de H entre 0° y 180° . Si $bo > go$, entonces H tiene que ser más grande que 180° . Así, donde $bo > go$, simplemente hay que definir $H = 360^\circ - H$.

El siguiente paso es encontrar una expresión para S en términos de un conjunto de valores RGB primarios. Para hacerlo se requiere nuevamente de las figuras 3.4 (a) y (b). Debido a que la saturación de un color es el grado de dilución de un color por el blanco, de la figura 3.4 (a) la saturación S de un punto color es dada por el rango $|WP|/|WP'|$, donde P' es obtenida extendiendo la línea WP hasta que intersecta al lado más cercano del triángulo.

Con referencia a la figura 3.4 (b), sea T la proyección de W en el plano rg paralelo al eje b y sea Q la proyección de P en WT , paralelo al plano rg , entonces

$$S = |WP|/|WP'| = |WQ|/|WT| = |WT| - |QT| / |WT| \quad (3.10)$$

donde el segundo paso se sigue de triángulos similares. Mientras $|WT| = 1/3$ y $|QT| = b$ en el sector mostrado,

$$S = 3\left(\frac{1}{3} - b\right) = 1 - 3b = 1 - bo \quad (3.11)$$

y donde el último paso sigue de la ecuación (3.2) y condición e). También se nota que $bo = \min(ro, go, bo)$ en el sector RG . Por supuesto, un argumento similar al expuesto muestra que la relación

$$S = 1 - \min(ro, go, bo) = 1 - \left(\frac{3}{R+G+B}\right)[\min(R, G, B)] \quad (3.12)$$

cumple en general para cualquier punto que está en el triángulo HSI .

Así, los resultados obtenidos dan las siguientes expresiones para obtener valores HSI en el rango $[0, 1]$ de un conjunto de valores RGB en el mismo rango:

$$I = \frac{1}{3}(R+G+B)$$

$$S = 1 - \left(\frac{3}{R+G+B}\right)[\min(R, G, B)] \quad (3.13)$$

$$H = \cos^{-1} \left\{ \frac{1}{2} \frac{[(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^2} \right\}$$

Como ya se indicó, definimos $H = H - 360^\circ$, si $(B/I) > (G/I)$. Además para normalizar al rango de $[0,1]$, definimos $H = H/360^\circ$. Finalmente, si $S = 0$, de la ecuación (3.10) vemos que $|WP|$ debe ser cero, lo cual indica que W y P son el mismo punto, haciendo esto menos significativo para definir el ángulo H . Por lo tanto se concluye que el matiz H no está definido cuando la saturación es cero. Similarmente de las ecuaciones (3.13), la saturación es indefinida para $I = 0$.

Las componentes *HSI* de una imagen *RGB* se pueden describir así:

La intensidad *I* de una imagen *RGB* se mide como una sola imagen de los tres canales *R*, *G* o *B* por separado. Cada píxel en la imagen de matiz *H* tiene valores expresados en grados. Con 8 bits, los 256 valores son representados en el rango de $[0^\circ, 360^\circ]$ en incrementos de $n(360/255)$, para $n = 0, 1, \dots, 255$. Similarmente píxeles en la imagen de la saturación pueden verse dentro de valores de 0 (no saturación) a 1 (completa saturación) en incrementos de $(1/255)$.

Con matiz *H* interpretado en grados, los grises deben ser tomados como ángulos. Entonces grises más claros corresponden a valores de ángulos más grandes. Debido a que el rojo tiene los valores de ángulos más pequeños, los rojos de una imagen original *RGB* en la imagen de *H* aparecerán como los grises más oscuros, el amarillo aparecerá como la siguiente sombra de gris más clara y así sucesivamente para verde, cyan, azul y magenta, en ese orden como se puede apreciar en la figura 3.5.

La imagen de la saturación *S* de una imagen *RGB* se muestra en base a la ecuación (3.13). Los máximos colores primarios y secundarios puramente saturados aparecen como blanco (máximo) en la imagen de saturación (ver figura 3.5). Como la saturación no está definida cuando la intensidad es cero todos los componentes negros de una imagen original *RGB* se muestran en blanco. Se debe mencionar que los blancos en *RGB* aparecen como negros en *S*, debido a que el blanco corresponde a saturación cero (+Saturado \rightarrow +Blanco) (-Saturado \rightarrow +Oscuro).

Los grises correspondientes a la banda de color variante en el espacio *RGB* son más claros hacia los finales de la banda, indicando la más grande saturación de color ahí. Los grises correspondientes a la porción central (verde) de la banda son más oscuros que grises en las porciones extremas de la banda, indicando menos saturación en la región verde (Figura 3.5). Esta condición es consistente con la forma en la que la parte verde dominante del patrón de prueba fue generado.

Se nota que el componente de intensidad mostrado en la imagen banda *I* de la figura 3.5 muestra que el negro, blanco y gris aparecen como en la imagen original *RGB*. Aquí también se ve que los componentes correspondientes a todos los colores primarios puros tienen el mismo valor y aparecen en un gris constante. Los colores puros secundarios también tienen un valor constante de intensidad, pero ellos son el doble del valor de la intensidad correspondiente a los primarios y entonces aparecen como grises más claros. Este resultado también es el esperado, debido a la forma secundaria de la luz que se obtiene por sumar dos primarios. Las intensidades correspondientes a la banda de color variante, también se esperan y varían de oscuro a brillante y regresa a oscuro.

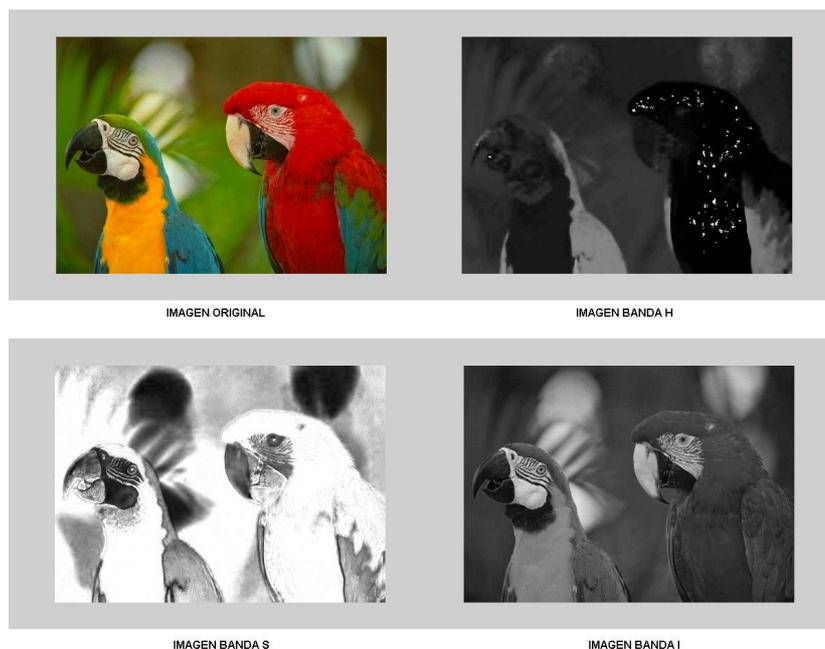


Figura 3.5. Descomposición de una imagen en canales *H*, *S* e *I*.

Como se ha mencionado, el modelo *HSI* es ideal para realizar mejoras de una imagen, esto debido a que la componente de intensidad es separada de la información de color en una imagen. Lo que se ha mostrado en esta sección es representativo de los modelos de color *RGB* y *HSI*, y la relación de uno con otro.

3.2. METODOS DE CLASIFICACIÓN DE DATOS.

3.2.1. Algoritmo de agrupamiento K-medias.

Este es uno de los algoritmos de agrupamiento más conocidos, también es llamado el *algoritmo de las medias móviles* debido a que en cada iteración se recalculan los centros de los agrupamientos [Duda, R. O., Hart P. E. y Stork, 2001]. Es un método no jerárquico que toma inicialmente el número de componentes de la población igual al número final requerido de clases o agrupaciones. En este paso el número final requerido de clases o agrupaciones es elegido de forma que los puntos son mutuamente apartados. Examina después, cada componente en la población y lo asigna a uno de las agrupaciones dependiendo de la distancia mínima. La posición del centroide es recalculada cada vez que un componente es sumado a la clase o agrupación y esto continúa hasta que todos los componentes son agrupados en el número final requerido de clases o agrupaciones.

El algoritmo requiere un único parámetro, K , el número de agrupamientos que debe encontrar. El objetivo es particionar o agrupar n patrones en subconjuntos disjuntos S_K conteniendo los N_j patrones así como minimizar el criterio de suma de cuadrados

$$J = \sum_{j=1}^K \sum_{n \in S_j} |x_n - \mu_j|^2 \quad (3.14)$$

donde x_n es un vector representando el n -ésimo patrón o punto dato y μ_j es el centroide geométrico de los puntos dato en S_K . En general, el algoritmo no aproxima un mínimo global de J sobre las asignaciones. Por supuesto, mientras el algoritmo usa asignaciones discretas mayores que un conjunto de parámetros continuos, el mínimo alcanzado no puede ser propiamente llamado un mínimo local. Sin tomar estas limitaciones, el algoritmo es usado frecuentemente como un resultado de su facilidad de implementación.

El algoritmo consiste de un simple proceso de reestimación como sigue:

1. Inicio. Consiste en inicializar arbitrariamente los centros de los S grupos, es decir, los patrones o puntos dato son asignados aleatoriamente a los S conjuntos. Entonces, el centroide es calculado para cada conjunto.
2. Asignación y actualización de los centros. En este paso se asigna cada patrón al grupo más cercano y se recalculan los centros en base a esta asignación.
3. Convergencia. En el paso anterior algunos patrones pueden cambiar de agrupamiento y en consecuencia, los centros de éstos. Si así ocurre, se repite el segundo paso hasta que no se cambien los centros. Si no hay modificaciones se considera que se ha encontrado una buena partición y se termina el agrupamiento.

Como se puede apreciar, el algoritmo es sencillo y eficiente debido a la simplicidad de los cálculos. Como inconvenientes, hay que destacar que el resultado final depende del valor de K , aunque el mayor problema está relacionado con la inicialización de los centros. Es fácil comprobar que el resultado está condicionado por éstos. Aunque los centros iniciales pueden seleccionarse aleatoriamente, si éstos se seleccionan de manera que se dispersen uniformemente sobre el espacio de representación, es más fácil que el resultado final sea mejor, además de acelerar la convergencia.

La idea entonces es que los pasos sean alternados hasta que el criterio de detección es alcanzado, es decir, cuando no haya cambios adicionales en la asignación de los puntos dato.

Ejemplo [Cortijo, 2001]:

De la figura 3.6(a) se ve el conjunto de patrones sobre el que se va a aplicar el algoritmo de la *K-medias* con $K=2$.

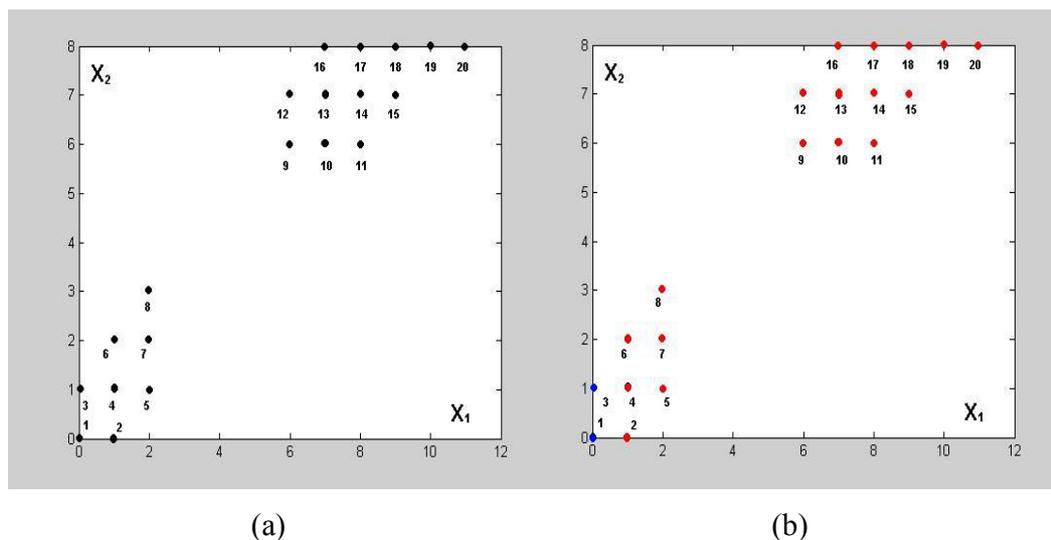


Figura 3.6. (a) Situación Inicial, y (b) Resultado de la primera asignación.

Paso 1. Inicio.

Se supone que los centros se toman en el orden en que están los patrones en X :

$S_1(0) = \{X_1\}$	$Z_1(0) = (0,0)$
$S_2(0) = \{X_2\}$	$Z_2(0) = (0,0)$

donde $S_i(j)$ es el conjunto de n elementos pertenecientes a la i -ésima clase a la j -ésima iteración y $Z_i(j)$ es el centroide de la i -ésima clase a la j -ésima iteración.

En los pasos siguientes se pondrán incisos a los pasos a realizar para detallar el número de ocasiones que se llevan a cabo dichos pasos.

Paso 2(a). Asignación y actualización de los centros

Si se asigna en base a la menor distancia entre $\delta(X_1, Z_1(0))$ y $\delta(X, Z_2(0))$ y finalmente se calculan $Z_1(1)$ y $Z_2(1)$ (figura 3.6 (b)) se tiene:

$S_1(1) = \{X_1, X_3\}$	$Z_1(1) = (0, 0.5)$
$S_2(1) = \{X_2, X_4, X_5, \dots, X_{20}\}$	$Z_2(1) = (5.8, 5.3)$

Los centroides $Z_i(j)$ a la j -ésima iteración se obtienen por medio de la ecuación:

$$Z_i(j) = [\sum_{n=a} (x_n)]/c \quad (3.15)$$

a = elementos pertenecientes a la i -ésima clase

c = número de elementos pertenecientes a la i -ésima clase

Paso 3(a). ¿Convergencia?

Como $Z_1(1) \neq Z_1(0)$ y $Z_2(1) \neq Z_2(0)$ se vuelve al paso 2.

Paso 2(b). Asignación y actualización de centros:

Si se asigna en base a la menor distancia entre $\delta(X_1, Z_1(1))$ y $\delta(X, Z_2(1))$ y finalmente se calculan $Z_1(2)$ y $Z_2(2)$ (figura 3.7 (a)) se tiene:

$S_1(2) = \{X_1, X_2, \dots, X_8\}$	$Z_1(2) = (1.1, 1.3)$
$S_2(2) = \{X_9, X_{10}, \dots, X_{20}\}$	$Z_2(2) = (8.0, 7.2)$

Paso 3(b). ¿Convergencia?

Como $Z_1(2) \neq Z_1(1)$ y $Z_2(2) \neq Z_2(1)$ se vuelve al paso 2.

Paso 2(c). Asignación y actualización de centros:

Si se asigna en base a la menor distancia entre $\delta(X_1, Z_1(1))$ y $\delta(X, Z_2(1))$ y finalmente se calculan $Z_1(2)$ y $Z_2(2)$ (figura 3.7 (a)) se tiene que:

$S_1(3)=\{X_1, X_2, \dots, X_8\}$	$Z_1(3)=(1.1, 1.3)$
$S_2(3)=\{X_9, X_{10}, \dots, X_{20}\}$	$Z_2(3)=(8.0, 7.2)$

Paso 3(c). ¿Convergencia?

Finalmente, como $Z_1(3) = Z_1(2)$ y $Z_2(3) = Z_2(2)$ se considera que la partición obtenida es estable y se termina. La partición final es la mostrada en la figura 3.7(b).

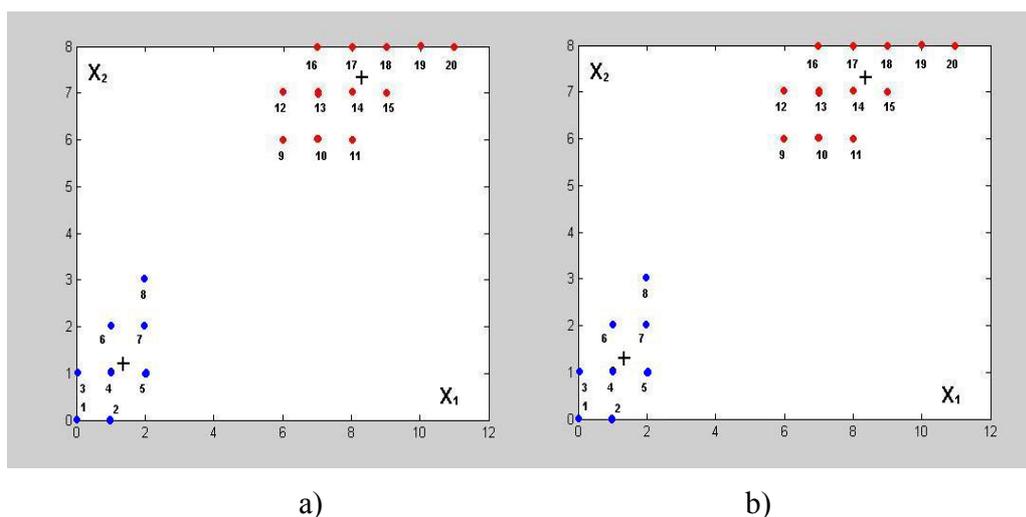


Figura 3.7. a) Segunda asignación, y b) Tercera y última asignación.

Métodos de agrupamiento KMD y KMC.

Dado un conjunto de clases C_1, C_2, \dots, C_c y un conjunto $X=(x_1, x_2, \dots, x_N)$ de N muestras, bajo la perspectiva difusa [Pajares y De la Cruz, 2002], se considera que dada una muestra x_j esta puede pertenecer a más de una clase. Esta pertenencia se mide por lo que se conoce como grado de pertenencia, que designaremos como $P(C_i/x_j)$ y que indica en que medida la muestra x_j pertenece a la clase C_i .

Las clases se caracterizan por tener un vector denominado centro de clase, en el caso que nos ocupa para las C clases serian los siguientes : m_1, m_2, \dots, m_c . De acuerdo a la forma en que se calculan estos grados de pertenencia se pueden definir dos tipos de algoritmo *k-medias* [Duda, et al, 2001]:

-*k-medias difuso KMD*

-*k-medias clásico KMC.*

El algoritmo *KMD* busca un mínimo de la función de costo global heurística siguiente:

$$J_D = \sum_{i=1}^c \sum_{j=1}^n [P(c_i/x_j)]^b \|x_j - m_i\|^2 \quad (3.16)$$

donde b es un parámetro libre, elegido para ajustar el solapamiento o mezcla de las diferentes clases, si b es 0, J_d es un mero criterio de suma de errores al cuadrado con cada patrón asignado a una clase única. Para $b > 1$, el criterio permite que cada patrón pertenezca a más de una clase. Los grados de pertenencia de cada patrón a las diferentes clases se normalizan como sigue:

$$\sum_{i=1}^c P(c_i/x_j) = 1 \quad (3.17)$$

La solución se obtiene encontrando los mínimos de J_d siguientes:

$$\partial J_D / \partial m_i = 0 \quad y \quad \partial J_D / \partial P(c_j) = 0 \quad (3.18)$$

donde $P(C_j)$ es la probabilidad a priori de que un patrón pertenezca a la clase C_j . Lo anterior conduce a las siguientes soluciones,

$$m_i = \frac{\sum_{j=1}^n [P(c_i / x_j)]^b x_j}{\sum_{j=1}^n [P(c_i / x_j)]^b} \quad (3.19)$$

con

$$P(c_i / x_j) = \frac{\left[\frac{1}{d_{ij}} \right]^{\frac{1}{b-1}}}{\sum_{r=1}^c \left[\frac{1}{d_{rj}} \right]^{\frac{1}{b-1}}} \quad y \quad d_{ij} = \|x_j - m_i\|^2 \quad (3.20)$$

donde $b > 1$ es un escalar denominado peso exponencial, de manera que cuanto mayor sea su valor, menor es la contribución de las muestras a la función objetivo. Función utilizada para obtener las ecuaciones (3.18) y (3.19). El número de clases C debe ser inferior al número de muestras y superior a 1, es decir, $1 < C < N$; d_{ij} es la distancia Euclídeana al cuadrado, si bien se podrían utilizar otras distancias.

Debido a que las ecuaciones (3.18) y (3.19) rara vez tienen soluciones analíticas, los centros de las clases y las probabilidades se estiman iterativamente de acuerdo con el siguiente algoritmo que se describe con todo detalle como sigue:

1. Extraer el número de muestras N a utilizar y el número de clases C .
2. Inicializar los centros de clases m_i y las probabilidades $P(c_i/x_j)$, $i = 1 \dots c$; $j = 1 \dots n$.
3. Normalizar las probabilidades por medio de la ecuación $\sum_{i=1}^c P(c_i/x_j)$, $i = 1 \dots c$; $j = 1 \dots n$.
4. Obtener m_i por medio de la ecuación (3.18).
5. Recalcular $P(c_i/x_j)$ por medio de la ecuación (3.19).
6. Repetir los pasos 3 a 5 hasta que m_i y $P(c_i/x_j)$ no cambien o el cambio sea pequeño.

El algoritmo de agrupamiento clásico *KMC* es un caso especial del anterior, en el que los grados de pertenencia a las clases se calculan como sigue:

$$P(c_i/x_j) = \begin{cases} 1 & \text{si } \|x_j - m_i\| < \|x_j - m_k\| \text{ para todo } k \neq i. \\ 0 & \text{de otro modo.} \end{cases} \quad (3.21)$$

Con el fin de acelerar la convergencia del método se puede añadir una condición adicional para el fin del algoritmo en el paso 6. Por supuesto, el proceso iterativo también puede detenerse si se alcanza un predeterminado número de iteraciones, aun cuando no se hayan cumplido ninguna de las dos condiciones de cambio dadas en dicho paso. Finalmente para determinar si un cambio es pequeño o no, se introduce un umbral E en base al cual se determina si los centros o los grados de pertenencia han cambiado o no respecto de la iteración anterior como lo exige el algoritmo. Es decir, si los valores de los cambios están por debajo de dicho umbral se consideran que no cambian.

3.3. TRABAJO RELACIONADO.

[Jones y Rehg, 1999] propusieron métodos estadísticos basados en modelos de color para realizar la detección de piel, estos consisten en calcular el histograma de datos de miles de imágenes o fotos de la World Wide Web para construir modelos genéricos de color. A partir de aquí se describe como construir modelos de color para clases de piel y no-piel. El uso del histograma beneficia según los autores tanto para la precisión del método de detección como para un bajo costo computacional.

Se han mostrado buenos resultados en la implementación de este método para detectar piel en imágenes aisladas. Requiere de hacer una gran recopilación de datos para tratar de abarcar la mayor cantidad de tonalidades de la piel, esto por supuesto, es un trabajo de mucho tiempo. En este trabajo se hace uso del *clasificador Bayesiano* que utiliza

la gran cantidad de información almacenada para realizar el proceso de captar piel en las imágenes que tenga que analizar.

[Zhang y Wang, 2002] proponen un nuevo método de segmentación de imágenes a color basado en el agrupamiento de matiz e intensidad, el método de agrupamiento utilizado es el algoritmo *k-medias* en el espacio *HSI*. Se describe la particularidad del agrupamiento realizado para la componente *H*, debido a su propiedad cíclica especial, utilizando definiciones de distancias y centros en el espacio de *H* sobre el que se implementan los análisis recopilados. Las pruebas de esta investigación se realizaron en el procesamiento de imágenes médicas en el que se muestra su buen comportamiento.

A diferencia de trabajos anteriores, en este se plantea el algoritmo *k-medias* pero usándolo para segmentar objetos de acuerdo a su textura y color basado en el espacio *HSI*, ya que era comúnmente utilizado en espacio *RGB*. Además presenta la importancia de tomar el canal de matiz *H*, de acuerdo a su propiedad cíclica, que como se demuestra influye directamente en el proceso de segmentación de la imagen para discriminar una textura de otras.

[Pajares y De la Cruz, 2002] aplicaron ambas versiones del algoritmo *k-medias*, *difuso* y *clásico*, para realizar una clasificación de texturas naturales en imágenes con base en los histogramas de color en el espacio *RGB*. Utilizan el *k-medias difuso* para realizar el agrupamiento dada la naturaleza aleatoria e irregular del tipo de texturas naturales, donde por sus propiedades daría el caso de que cierta textura pudiera pertenecer a más de una clase y utiliza de igual forma el *k-medias clásico* en el caso más simple de que la pertenencia de una textura sea a una única clase en forma indiscutible.

Se realiza un análisis muy detallado de todo el proceso de clasificación de texturas en este caso para nueve clases posibles y obteniendo avances en resultados en comparación con otros métodos de agrupamiento.

Aquí se realiza un trabajo muy completo en la detección y clasificación de texturas naturales usando como otros, histogramas en espacio *RGB* de una cantidad considerable de imágenes de entrenamiento que sirven para alimentar el algoritmo de clasificación *k-medias*, de igual forma compara el funcionamiento del *k-medias* con otros métodos de clasificación, mostrando las ventajas de rapidez y efectividad de este algoritmo en la obtención de buenos resultados para clasificar color.

3.4. CONCLUSIONES.

En esta parte del trabajo se ha analizado dos de las herramientas que se utilizarán en adelante, el espacio de color *HSI* y el algoritmo de clasificación de información. Se ha visto a detalle como afecta cada componente del modelo *HSI* en el análisis de imágenes, así como su relación con el espacio de color más común, el modelo *RGB*. Se repasó también el proceso de conversión del espacio *RGB* al *HSI* para hacer más clara la relación entre uno y otro.

En el ramo de la clasificación de datos se revisó paso a paso en que consiste el algoritmo *k-medias* que se probó para realizar la clasificación de color, para posteriormente, de acuerdo a resultados, mostrar su eficiencia para alcanzar los propósitos de detección de objetos. Es importante remarcar que para el buen funcionamiento de un clasificador de color en el procesamiento de imágenes, no sólo es importante la segmentación de las mismas, sino el comportamiento de los mismos en tiempo real.

4. CALIBRACIÓN DE UNA CÁMARA BASADA EN EL MOVIMIENTO DE UN OBJETO.

Este estudio plantea la calibración de una cámara asentada en una base con la capacidad de rotar por el movimiento de un motor ubicado en la base. Este le permite un movimiento horizontal-vertical. Se ubica la cámara en cierta posición en una escena con la premisa de tener la mayor perspectiva posible del lugar. Posteriormente, el interés se centra en lograr que al haber un objeto en movimiento en la escena, la cámara sea capaz de detectar primeramente al objeto y con ello seguir cualquier posible dirección que el objeto tome al moverse dentro de la escena. Esto último, involucra el manejo adecuado que la cámara deba realizar en la captación de imágenes y el eficiente trabajo del motor para dirigir la cámara hacia donde su campo de visión le permita seguir al objeto en movimiento.

Se define primeramente un proceso que consiste en realizar la detección del objeto con base en un algoritmo de clasificación de píxeles de cada una de las imágenes que componen una secuencia de video captada por una cámara. La clasificación de píxeles se realizará tomando como patrón de clasificación el modelo de color *HSI*, y utilizando el algoritmo *k-medias* se marca como objeto detectado aquel con las características del color de la piel. El segundo de ellos corresponde a realizar una calibración activa de la cámara para actualizar los parámetros de rotación de la misma y llevar a cabo el seguimiento del objeto detectado hasta donde al campo de visión de la cámara le sea posible.

4.1. DETECCIÓN Y RECONOCIMIENTO DEL OBJETO

La detección del objeto consiste de un proceso clasificatorio de píxeles, imagen por imagen de una secuencia de video. Para realizar ello se utiliza el método de clasificación *k-medias difuso* explicado con detalle en el Capítulo 3 para clasificar píxeles tomando como patrones de clasificación los canales de color *H*, *S* e *I* del espacio de color *HSI*. Aplicando

este método de clasificación se es capaz de hacer la separación de los píxeles, en píxeles de color piel y píxeles sin color de piel.

Se describe primeramente porque se eligió el espacio de color *HSI* para este trabajo y la forma en que se revisan los histogramas de los canales *H*, *S* e *I* para obtener los datos estadísticos necesarios para alimentar el algoritmo de clasificación. Después se explica el algoritmo de clasificación *k-medias* basado en datos de color tomados con anterioridad. Se verá a detalle como se tomaron los patrones de color para llevar a cabo la clasificación de píxeles en dos clases: Clase de piel y Clase de no-piel.

4.1.1. Modelo de Color.

El espacio de color *HSI* descrito en el Capítulo 3 debido a su estructura permite tomar cualquier color y obtener información de él en base a su cromaticidad y a su luminosidad. Es primordialmente útil para los propósitos de la detección tomar el rango de tonalidades de piel dentro de las miles de posibilidades que pudieran existir y separarlas de la cantidad de luz que pudiera incidir sobre el mismo matiz del color.

Trabajos anteriores han demostrado que la distribución de color de piel humana tiene la propiedad de formar grupos compactos y reducidos en la mayoría de los espacios cromáticos. Las diferencias de color entre las personas son debidas a cambios en la intensidad más que a variaciones del matiz del color. Se requiere de un espacio de representación de colores que elimine el efecto de la intensidad del color para la detección de piel.

En el modelo *HSI*, el canal *H* resulta más discriminante para resolver el problema de clasificación de piel o no-piel. Además, la componente *H* es la más adecuada para

describir el contenido de color de una imagen digital, y por ende su pérdida de información de color es menos considerable al momento de cambios en la iluminación.

Haciendo análisis de imágenes de piel, se ha comprobado que el color de la piel se encuentra siempre en valores bajos de H , lo que en el espacio RGB significarían los valores próximos al rojo. El canal de saturación aporta menos información, aunque es útil para distinguir el rojo del rosa-piel. Sin embargo, ninguno de los canales anteriores por sí mismos son suficientes para la detección de piel. De acuerdo a la información anterior, se decidió utilizar la representación HSI . Mientras los canales H y S son útiles para la segmentación de píxeles de color de piel, el canal I resulta interesante para segmentar los elementos faciales que pudieran ser más oscuros por variaciones de luz.

El modelo de color HSI , a pesar de ser menos conocido, ha sido muy utilizado en diferentes proyectos [García Mateos, et al, 2002], principalmente aquellos que de alguna u otra forma involucran la detección de personas debido a que se ha probado que el HSI por sus condiciones, permite ser más discriminante en cuanto a los tonos del color de piel, no tanto para la clasificación de otro tipo de texturas naturales [Pajares G. y De la Cruz J. M., 2002].

El objetivo de este trabajo se centra en la detección, ubicación y seguimiento de un objeto. Se contempla un conjunto de color de piel, basado en histogramas de una cantidad considerable de imágenes de piel y no piel, que son almacenados como datos estadísticos que nos permiten establecer los umbrales HSI necesarios para llevar a cabo la segmentación inicial. Con esta información el sistema es capaz de detectar las regiones correspondientes a piel en cualquier imagen mediante el proceso de clasificación de píxeles.

4.1.2. Clasificador de Color.

Se ha definido primeramente el tipo de datos que vamos a utilizar (*HSI*) para realizar nuestra clasificación de regiones en una imagen. Ahora se lleva a cabo una recopilación de datos que consisten de cientos de imágenes obtenidas de la red y otras obtenidas con una cámara digital con la idea de juntar la información necesaria de imágenes de piel y no piel.

Posteriormente hizo un estudio de los datos para definir el comportamiento de los mismos a través del método del histograma, el cual consiste en una gráfica que define la frecuencia medida en el número de píxeles, en la que determinado dato (color) aparece en una(s) imagen(es) y que además se ejecuta para cada canal *H*, *S* e *I*. En los histogramas, se analizan las propiedades o características del color de la piel con el fin de que el sistema pueda llevar a cabo el entrenamiento y la clasificación de los patrones. En la figura 4.1 se muestra una imagen con múltiples fragmentos de una gran variedad de tonos de piel de las imágenes recopiladas.

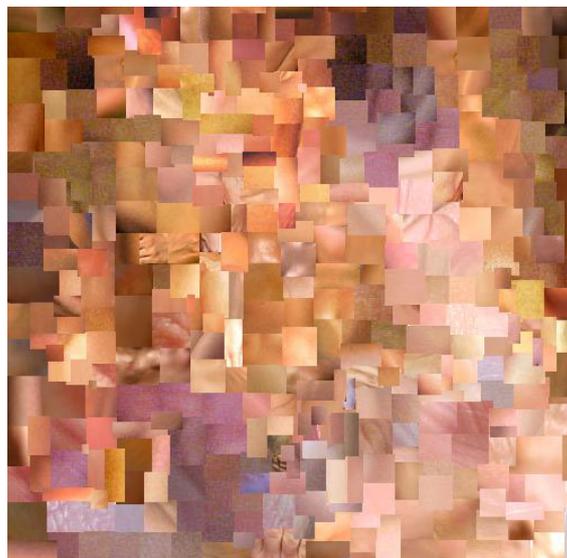


Figura 4.1. Parte de recopilación de muestras de piel.

A partir de las imágenes de muestras, se obtiene la secuencia de histogramas correspondientes para cada una de las tres bandas espectrales H , S , e I , donde los valores se han normalizado y se obtiene el histograma para los tres canales en el rango de $[0,1]$ como se muestran en las figuras 4.2 (a), (b) y (c).

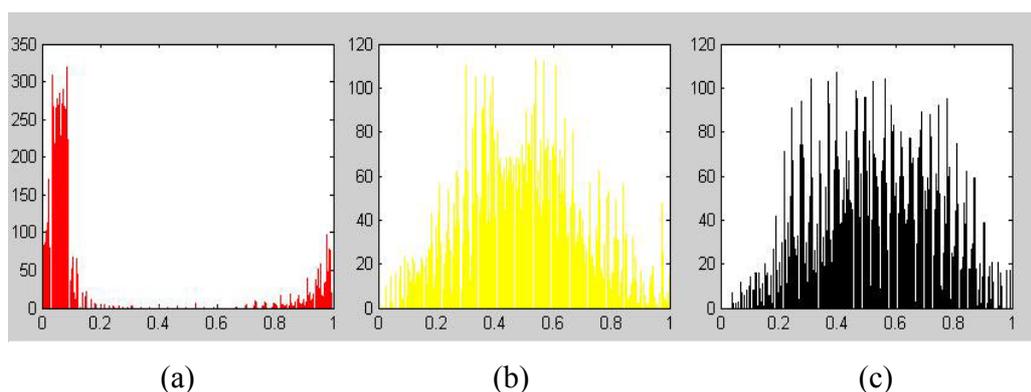


Figura 4.2. (a) Histograma canal H , (b) Histograma canal S y (c) Histograma canal I .

En primer plano se observa la inclinación del canal H a ubicarse en los valores bajos principalmente $[0,0.2]$, y de igual forma una ligera acumulación en los valores altos $[0.8, 1]$, esto muestra como en estudios anteriores que el tono de piel se ubica en tonos próximos al rojo en escala RGB . El canal S aunque menos útil para nuestros propósitos, sí nos puede ayudar a discriminar el tono de piel del tono rosa que podría en cierto momento dar pie a confusión en la clasificación. Se aprecia que toma valores intermedios con picos en los rangos $[0.3, 0.4]$ y $[0.5, 0.6]$. Por último el canal I , que no es más que la imagen original pero en escala de grises igualmente se concentra en valores intermedios, pero los picos se distribuyen más uniformemente en un rango de $[0.25, 0.78]$. Tomando ahora el histograma conjunto de los tres canales obtenemos la siguiente figura 4.3:

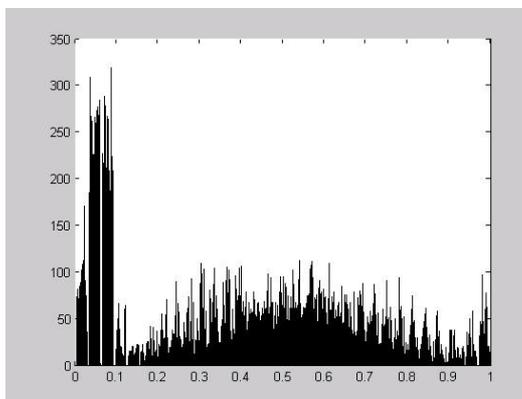


Figura 4.3. Histograma conjunto de los tres canales H , S e I .

Observando las Figuras 4.2(a), (b) y (c), y 4.3 se deduce lo siguiente:

1. El grado de solapamiento en la figura 4.3 es considerable, principalmente en valores intermedios y entre los canales S e I .
2. La variación de intensidad se debe en cierto grado a las condiciones de captura de las imágenes (iluminación, captura de la cámara digital, distancias, etc.) y por otro lado al tipo de textura de las diferentes variedades de piel humana.

De la primera consideración se deduce que el método KMD puede ser el adecuado para texturas con alto grado de solapamiento en los histogramas de las bandas espectrales, ya que el problema es visto desde la perspectiva de lógica difusa y trata el problema de los valores de píxeles solapados como pertenecientes a más de una clase según su grado de pertenencia.

La segunda consideración por su parte es bien tratada por el algoritmo KMD ya que permite un aprendizaje de un representante para cada clase, dos en nuestro caso (piel y no piel), que estiman mediante un proceso de aprendizaje, el centro de cada clase.

Basados en el método *KMD* explicado en el Capítulo 3 aplicado a la información anterior, se detalla el algoritmo de clasificación de píxeles de una imagen en piel y no-piel para detectar un objeto de la siguiente forma:

X = imagen

$\{x_1, x_2, x_3, \dots, x_n\} = n$ píxeles de la imagen.

C_1 = clase piel. C_2 = clase no-piel

Bajo la perspectiva difusa se considera que dado un píxel x_j este puede pertenecer a una de las dos clases.

$P(C_1/x_j)$ = Grado de pertenencia del píxel x_j a la clase de piel C_1 .

$P(C_2/x_j)$ = Grado de pertenencia del píxel x_j a la clase de no-piel C_2 .

Al definir el color de la imagen en el espacio *HSI*, cada píxel de la imagen está asociado a un vector $x_j = [H_j, S_j, I_j]$, donde cada H_j , S_j , y I_j son los valores de cada canal en el píxel.

Se caracteriza cada clase asociando el vector centro de las dos clases que se definen como:

$m_1 = [H_1, S_1, I_1]$ y $m_2 = [H_2, S_2, I_2]$ para las clases C_1 y C_2 respectivamente.

Del algoritmo definido en el Capítulo 3 se llega a las siguientes soluciones:

$$m_1 = \frac{\sum_{j=1}^n [P(c_1 / x_j)]^b x_j}{\sum_{j=1}^n [P(c_1 / x_j)]^b} \quad \text{y} \quad m_2 = \frac{\sum_{j=1}^n [P(c_2 / x_j)]^b x_j}{\sum_{j=1}^n [P(c_2 / x_j)]^b} \quad (4.1)$$

con

$$P(c_1/x_j) = \frac{\left[\frac{1}{d_{1j}} \right]^{\frac{1}{b-1}}}{\left[\left(\frac{1}{d_{1j}} \right)^{\frac{1}{b-1}} + \left(\frac{1}{d_{2j}} \right)^{\frac{1}{b-1}} \right]} \quad y \quad P(c_2/x_j) = \frac{\left[\frac{1}{d_{2j}} \right]^{\frac{1}{b-1}}}{\left[\left(\frac{1}{d_{1j}} \right)^{\frac{1}{b-1}} + \left(\frac{1}{d_{2j}} \right)^{\frac{1}{b-1}} \right]} \quad (4.2)$$

y

$$d_{1j} = \|x_j - m_1\|^2 \quad d_{2j} = \|x_j - m_2\|^2 \quad (4.3)$$

En principio se fija $b = 2$, pero en la practica se probarán otros valores para ver cual da mejores resultados, obviamente el número de clases debe ser menor que el número de píxeles, es decir, $1 < c < n$; y d_{1j} y d_{2j} son las distancias Euclidianas al cuadrado.

Calculamos iterativamente las ecuaciones de arriba ejecutando los siguientes pasos:

1. El número de muestras será de 5,578,979 píxeles para las dos clases C_1 y C_2 .
2. Inicializar los centros de las clases m_1 y m_2 de acuerdo al cálculo de los histogramas y los fijamos en: $m_1 = [H_1, S_1, I_1] = [11.21, 122, 146]$ y $m_2 = [H_2, S_2, I_2] = [180, 155, 150]$
3. Calcular las probabilidades $P(c_1/x_j)$ y $P(c_2/x_j)$ de acuerdo a (4.2)
4. Obtener m_1 y m_2 por medio de la ecuación (4.1)
5. Recalcular $P(c_1/x_j)$ y $P(c_2/x_j)$ por medio de (4.2)
6. Repetir los pasos 4 y 5 hasta notar que el cambio de m_1 y m_2 es despreciable.

4.1.3. Depuración de la imagen.

Siempre que se realiza un procesamiento de una imagen puede haber lugar a errores. En este sentido, la clasificación de píxeles en dos clases en base al color da como resultado dos conjuntos de píxeles, cada uno correspondiente a cada clase. Sin embargo la posibilidad de agrupar píxeles en un conjunto que pudieran pertenecer al otro siempre esta presente, es decir la aparición de píxeles erróneos. Por ello, un proceso adicional de depuración de errores con las condiciones necesarias es de gran utilidad.

Este proceso consiste en analizar los píxeles tomados como color piel después de la clasificación realizada por el *k-medias*, tomando el valor de estos y ver cuales de ellos están en los rangos de color de los canales *H*, *S* e *I*, para el color de la piel, (Información obtenida de los histogramas). *Hpiel(295-35)*, *Spiel(52-125)* e *Ipiel(60-125)*.

Si el valor de píxeles se encuentra en estos rangos de los tres canales, se mantendrán en la clase piel, si algún o algunos valores de los píxeles están fuera de rango se toman como píxeles erróneos y se toman como píxeles pertenecientes a la clase de no-piel durante la detección del objeto.

De la misma clasificación de píxeles, es posible que se ubiquen dentro de la imagen más de una región de piel. Esto para los propósitos de la detección no es lo ideal, en tal situación es conveniente elegir una para seguir adelante. Por lo tanto se realiza un método con la idea de encontrar la región de piel más grande en la imagen y tomarla como el objeto de interés, desechando de esa forma las otras regiones posibles de piel. Este paso se lleva de la forma simple de llevar un conteo del número de píxeles de la o las regiones detectadas de piel y elegir la de mayor cantidad.

La salida de este paso nos da una imagen con una sola región de interés, que se muestra experimentalmente como una imagen segmentada con la región de interés (piel) en blanco y el fondo en negro.

En este punto termina el proceso completo de detección y es importante señalar que del objeto ubicado dentro de la imagen se obtienen las coordenadas de los puntos de referencia del objeto. Estos puntos sirven para alimentar el algoritmo de calibración de la cámara explicado más adelante.

4.1.4. Algoritmo de detección.

1. La entrada es una imagen a color en formato *RGB*.
2. Obtener los valores de la imagen en el modelo *HSI* usando las ecuaciones de conversión 3.13.
3. Aplicar el algoritmo de clasificación *k-medias* usando los valores de los píxeles en cada uno de los canales *HSI*.
4. Si se encuentra región de piel en la imagen seguir adelante, si no terminar el proceso.
5. Hacer la depuración de la imagen, que incluye eliminar errores y obtener la región de piel más grande en caso de haber más de una.
6. Mostrar en la imagen el objeto encontrado al ubicarlo dentro de un rectángulo rojo.

4.2. PROCESO DE CALIBRACIÓN DE LA CÁMARA.

Ahora toca describir el segundo proceso fundamental que comprende nuestro trabajo, el proceso de calibración de la cámara. Esto se realiza con la idea fundamental de que al conocer la existencia de un objeto de interés particular en el campo de visión de la cámara, ésta sea capaz de realizar el seguimiento del objeto dentro de alguna escena e ir actualizando constantemente los parámetros de la cámara. Hecho esto se puede seguir eficazmente al objeto por medio de la rotación horizontal y/o vertical de la cámara.

El proceso en general se describe a detalle en los siguientes pasos:

1. La cámara se encuentra ubicada en cierta posición fija en un lugar preferentemente cerrado (figura 4.4). Se supone que la cámara puede moverse en el sitio donde se encuentra, es decir rotar a determinados ángulos sobre la horizontal y la vertical (pan-tilt). Además se asume que los parámetros intrínsecos de la cámara anteriormente mencionados longitud focal, distorsión radial, centro del plano de imagen y razón de aspecto son conocidos.
2. La cámara tendrá un patrón inicial de movimiento similar al del ojo humano empezando a inspeccionar de izquierda a derecha, y de arriba hacia abajo. Este patrón de movimiento se mantiene hasta que dentro del campo de visión de la cámara se detecta el objeto de color de piel de mayores dimensiones según el proceso de detección descrito.
3. La calibración consistirá en la actualización de los ángulos de rotación de la cámara a partir de las imágenes que está procesando. La imagen que contiene al objeto (región de piel) viene dada en coordenadas de imagen (píxeles). Mientras que, se asigna al objeto puntos en 3-D a los que llamamos coordenadas del modelo. El modelo va a ir de acuerdo al objeto que pretendemos detectar y seguir, por tanto, es completamente necesario tener un conocimiento aproximado de las dimensiones del objeto. Por último, se toman las coordenadas de la cámara, las cuales, son la base de la calibración y sobre las que se toman los ángulos de rotación.

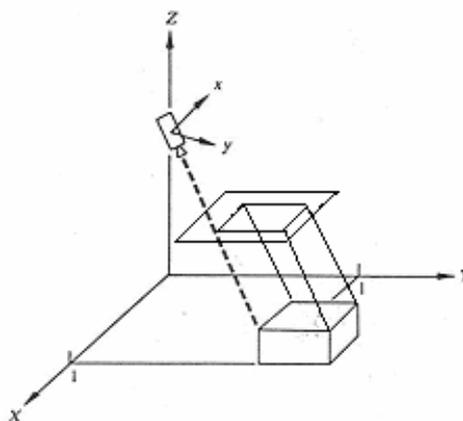


Figura 4.4. Posición y orientación inicial de la cámara

Los ángulos de rotación se toman sobre los ejes X (ϕ_1), Y (ϕ_2) y Z (ϕ_3) del marco de la cámara. La rotación de la cámara se hace únicamente en forma horizontal y vertical. La idea de los ángulos de rotación y la traslación es poder alinear los ejes de la cámara con los del modelo, lo que se logra a través de la siguiente transformación:

$$[X_c, Y_c, Z_c]^T = R * [X_i, Y_i, Z_i]^T + T \quad (4.4)$$

donde R es la matriz de rotación dada por y T es el vector de traslación dados por:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad \text{y} \quad T = [T_1, T_2, T_3] \quad (4.5)$$

En particular para este proceso se pretende que el eje óptico o eje Z de la cámara, posterior a la calibración, se ubique en posición perpendicular al centro del objeto y también se ubica el mismo objeto en el centro del plano imagen por medio de configurar las coordenadas del punto principal al centro del plano.

Como la cámara se encuentra en posición fija no habrá traslación para las posibilidades de movimiento de la cámara; sin embargo, obtenemos el valor de T como

parte de la estructura 3-D de la escena. Para otra aplicación en que la cámara pueda desplazarse, ya sea situada como parte del sistema de visión de un robot o montada sobre algún dispositivo con la capacidad de moverse, la traslación tendrá que considerarse.

4. Se tomarán N puntos p_i de la imagen en 2-D, y sus correspondientes N puntos P_i en 3-D de las coordenadas del modelo. La relación entre estos dos conjuntos de puntos son dados por la proyección perspectiva de la forma:

$$[x_i, y_i]^T = \left[f \frac{X_i}{Z_i}, f \frac{Y_i}{Z_i} \right]^T \quad (4.6)$$

5. Juntando las dos ecuaciones (4.4), (4.5) y (4.6), cada correspondencia de puntos generan dos ecuaciones no lineales.

$$x_i = f \frac{r_{11}X_i + r_{12}Y_i + r_{13}Z_i + T_1}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_3} \quad (4.7)$$

$$y_i = f \frac{r_{21}X_i + r_{22}Y_i + r_{23}Z_i + T_2}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_3}$$

Las incógnitas de R y T pueden obtenerse de un determinado número de correspondencias, cada una trayendo dos ecuaciones como las anteriores. Así, el sistema resultante se compone de N correspondencias teniendo seis incógnitas, las componentes del vector de traslación T_1 , T_2 y T_3 y los ángulos de rotación sobre los tres ejes de la cámara ϕ_1 , ϕ_2 y ϕ_3 .

6. Se resuelva el sistema usando el método de posición 3-D que consiste en hacer una estimación inicial de la posición de la cámara al momento de detectar el objeto, es decir un vector de traslación inicial T_0 y los posibles ángulos ϕ_1 y ϕ_2 . Con ello una estimación inicial de la matriz de rotación que llamaremos R_0 .

7. Usar R_o , T_o y la ecuación. (4.4) para calcular los puntos de ubicación 3-D en el marco de la cámara que llamaremos P_c a partir de los P_i de las coordenadas del modelo del objeto.
8. Se proyectan los N puntos P_c en el plano imagen usando (4.6) y obtener los N puntos $p_{io} = [x_{io}, y_{io}]$ y ahora se calculan los residuales:

$$\begin{aligned}\delta x_i &= x_{io} (R_o) - x_i \\ \delta y_i &= y_{io} (R_o) - y_i\end{aligned}\quad (4.8)$$

posteriormente, resolver el sistema de las N correspondencias dado por:

$$\begin{aligned}\sum_{j=1}^3 [(dx_i/dT_j) * \Delta T + (dx_i/d\phi_j) * \Delta \phi_j] &= \delta x_i \\ \sum_{j=1}^3 [(dy_i/dT_j) * \Delta T + (dy_i/d\phi_j) * \Delta \phi_j] &= \delta y_i\end{aligned}\quad (4.9)$$

para las incógnitas de corrección $\Delta \phi_1, \Delta \phi_2, \Delta \phi_3, \Delta T_1, \Delta T_2$ y ΔT_3

9. Actualizar las estimaciones del vector T , la matriz R y por tanto de ϕ_1, ϕ_2 y ϕ_3 .
10. Si los residuos son lo suficientemente pequeños o cero, el proceso se repite para cada una de las imágenes de la secuencia mientras la cámara detecte el objeto. Si no, regresar al paso 6.
11. Cuando el objeto sale del campo de visión, se considera la historia del movimiento para determinar la próxima dirección de rotación de la cámara. Esto se realiza de acuerdo a los últimos ángulos de rotación registrados, por lo tanto la cámara se mantiene con esa orientación un tiempo adecuado y analizando en la imagen

siguiente de la secuencia la última región donde se captó al objeto. En el caso que después del tiempo definido no se capte objeto alguno, la cámara vuelve al patrón de movimiento inicial.

4.2.1. Algoritmo de calibración.

1. La entrada son los puntos de la imagen obtenidos de la detección, los puntos del modelo de objeto, los parámetros intrínsecos de la cámara y las estimaciones iniciales de la matriz de rotación R y el vector de traslación T .
2. Obtener los puntos del modelo en las coordenadas de la cámara usando las estimaciones iniciales de R y T mediante la ecuación 4.4.
3. Proyectar los puntos 3-D de las coordenadas de la cámara en el plano imagen por medio de la ecuación 4.6.
4. Calcular los residuos entre los puntos imagen del paso 3 con los puntos de la imagen originales por medio de la ecuación 4.8.
5. Resolver el sistema de ecuaciones 4.9 resolviendo para las incógnitas de corrección de R y T .
6. Actualizar las estimaciones de R y T .
7. Si el valor de los residuos es mínimo o cero termina el proceso y se toman R y T como el resultado.
8. Si los residuos o grados de error son considerables regresar al paso 1, tomando ahora los valores actualizados de R y T .

4.3. CONCLUSIONES.

En este capítulo hemos descrito en que consiste nuestro proyecto, presentamos el modelo de detección en base al algoritmo de clasificación *k-medias* utilizado en la clasificación de color, vimos como se realizó la obtención de datos e información para la alimentación del modelo matemático del algoritmo. Estos métodos dan con eficiencia las

probabilidades de pertenencia de que determinado color en una imagen sea o no, el que forma parte del conjunto de color que buscamos.

Se han definido los pasos a seguir en el proceso de calibración, empezando por describir la situación inicial en la que se encuentra la cámara, la obtención de puntos de referencia en la imagen y su relación con los puntos modelo predefinidos, el comportamiento a partir de que se le manda la orden de orientarse en la dirección que el modelo de detección le diga, la actualización constante de sus parámetros y el seguimiento del objeto en movimiento hasta donde el campo de visión le alcanza. Es decir, damos todo el desarrollo teórico de nuestra propuesta y cuyos resultados obtenidos los mostramos en el siguiente capítulo.

5. RESULTADOS EXPERIMENTALES DEL SEGUIMIENTO DE UN CUBO.

Se implementó el sistema de visión con una cámara pan-tilt marca logitech Pro-4000 (Figura 5.1). Esta cámara tiene la capacidad de rotar su foco en forma horizontal y vertical. Es por tanto, la responsable de suministrar al sistema las imágenes necesarias para realizar la detección y calibración. Una vez hecho lo anterior, se lleva a cabo el seguimiento del objeto que se persigue.



Figura 5.1. Cámara con pan-tilt usada en la implementación.

Para realizar los experimentos para el proceso de detección se han recopilado 1,250,000 píxeles de información en imágenes de muestras de piel como la observada en la figura 4.1, en las cuales, se calcula su histograma para obtener los datos necesarios para alimentar el algoritmo de detección. En este Capítulo se muestran las pruebas realizadas con este algoritmo de detección, así como también el comportamiento del método de calibración y seguimiento que realiza la cámara.

El proceso de detección de objetos tiene dos objetivos principales y básicos en cualquier algoritmo de esta naturaleza: la correcta segmentación de cada imagen de la secuencia de video en la que se detecte un objeto de interés y la suficiente rapidez en tiempo real del procesamiento de la imagen. Es preciso recalcar que el entrenamiento del algoritmo se ejecuta previo a la detección como se explicó ampliamente en el Capítulo 4. Así pues, los resultados obtenidos en el entrenamiento son los que se utilizan de referencia para llevar a cabo el método de detección con el algoritmo *k-medias*.

5.1. DETECCION DEL OBJETO.

En principio se muestran cuales son los criterios utilizados para la inicialización y fijación de los parámetros utilizados por el *k-medias* para el proceso de detección:

1. Los centros para cada una de las clases se inicializan en base a la información aportada por los histogramas para cada canal del espacio *HSI* de las imágenes de recopilación de muestras de piel. El algoritmo únicamente contempla dos clases (*piel* y *no-piel*). En principio para la clase *no-piel* se probaron diferentes centros de inicio en el modelo *HSI* y cuyo rendimiento medido en base al número de iteraciones promedio en las que se llegaba a una correcta clasificación se muestra en la tabla 5.1. Al observar que la velocidad de convergencia del algoritmo no era la esperada se prefirió definir los centros con la información proporcionada por histogramas de una recopilación de muestras de alrededor de 4,528,970 píxeles de imágenes que no contuvieran piel en diferentes ambientes naturales y urbanos.

Tabla 5.1. Diferentes puntos de inicio de los centros para la clase No-Piel.

Centros Piel	Centros No Piel	Número de Iteraciones It.
(H,S,I)=(11.2, 122, 146)	(H,S,I)=(0,0,0)	It=35
(H,S,I)=(11.2, 122, 146)	(H,S,I)=(360,255,255)	It=42
(H,S,I)=(11.2, 122, 146)	(H,S,I)=(180,128,128)	It=29
(H,S,I)=(11.2, 122, 146)	(H,S,I)=(180, 155, 150)	It=16

NOTA: Los centros de clase piel son los mismos para todas las pruebas.

Según la información de la tabla 5.1 elaborada en base a diferentes imágenes en diversos ambientes vemos que el promedio de iteraciones es menor cuando se toman los centros en $(H,S,I)=(180, 155, 150)$. Estos fueron determinados con los histogramas mencionados, a diferencia de los otros puntos de inicio que fueron elegidos arbitrariamente. Los centros de la clase piel son los mismos en todo el análisis.

Los parámetros H , S e I fueron tomados dentro de los rangos de valores siguientes:
 $H=0-360$. $I=0-255$. $S=0-255$.

De acuerdo a lo anterior definimos los centros para las dos clases piel (p) y no-piel (np) de la forma:

$$C_{piel}=[H_{po}, S_{po}, I_{po}] = [11.21, 122, 146]$$

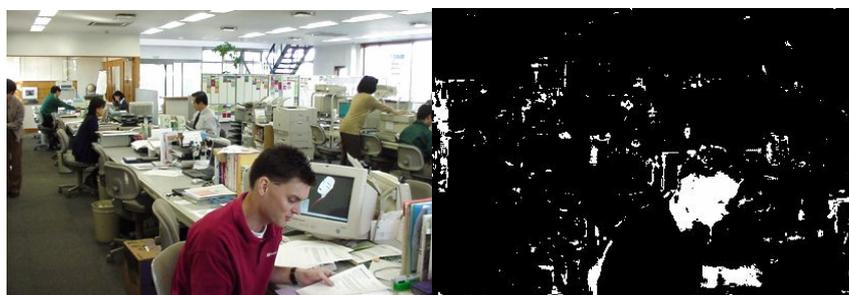
$$C_{nopiel}=[H_{npo}, S_{npo}, I_{npo}] = [180, 155, 150]$$

2. El valor de peso exponencial escalar b que define el porcentaje de contribución de las muestras a la función objetivo de las estimaciones de los centros de clases m_i y las probabilidades $P(c_i/x_j)$ se fijó en 3. Este valor se obtuvo después de realizar diferentes pruebas y constatar que es el de mejores resultados, aunque en otros trabajos se recomienda fijar en 4 , [Pajares y De la Cruz, 2002].
3. Se decidió fijar el número de iteraciones en $It=15$ en base a pruebas de rapidez y comportamiento de convergencia del algoritmo, ya que no tendría caso aumentar el número de iteraciones haciendo más lento el algoritmo cuando el cambio en la estimación de los nuevos centros posteriores a la iteración 15 o incluso desde antes es mínima o despreciable.

Tomamos inicialmente pruebas con imágenes en las que se aplicó el algoritmo de clasificación de color y las segmentamos de acuerdo a píxeles de clase piel en blanco y clase no-piel en negro obteniendo los siguientes resultados:



(a)



(b)



(c)



(d)

Figura 5.2. Pruebas del algoritmo de detección basado en el *k-medias*.

Podemos apreciar una aceptable detección de piel, se nota en las imágenes una mayor tendencia a errores cuando hay menos luz o mucha luz en la escena. En la figura 5.2(a) una imagen a la luz del día con sombras en áreas de piel, se ve una considerable

clasificación de píxeles de piel. En la figura 5.2 (b) se ve una imagen de un ambiente normal con una cantidad importante de personas, y se aprecia que se agrupan en su totalidad los píxeles de color piel, aunque también se toman diferentes objetos con colores próximos a piel. En la imagen 5.2(c), se nota una buena detección de la mano, a pesar de la oscuridad de la imagen y en la figura 5.2(d) se nota una casi nula detección al no haber tonos de piel en la imagen, es decir, se hace una buena clasificación. Por errores en el procesamiento se quiere decir que se nota en las imágenes segmentadas pequeñas regiones negras que se toman como piel y que son más propensas a suceder cuando se presentan escenas con muy poca luz o escenas con demasiada luz. Para mejorar esta situación se realiza durante el procesamiento de la imagen el paso de depuración de errores explicado en la sección 4.1.3, el cual, es eficiente para nuestros propósitos.

Una vez mostrada la eficiencia del clasificador de color en diferentes imágenes, prueba el mismo proceso pero ya con el propósito de detectar cierto objeto de interés en un ambiente general y lograr obtener la información para desarrollar la calibración de la cámara y el seguimiento del objeto.

Como se comentó al inicio de este trabajo, al no haber modelos precisos de partes de cuerpo humano, para hacer la detección y calibración en forma conjunta es necesario tomar un objeto geométrico con características de color similares al las de la piel humana. El objeto que se usa para las pruebas fue un cubo de madera de dimensiones conocidas. El color del cubo (ver figura 5.3) fue elegido muy próximo al del color de piel humana para probar el algoritmo de detección pasando por los diferentes aspectos detallados en el Capítulo 4. De esa manera no se tiene que modificar la inicialización de los centroides.

Dada una imagen de la secuencia de video (Figura 5.3) en la que se encuentra el objeto de interés (cubo de madera), describimos a continuación los resultados obtenidos paso por paso.



Figura 5.3. Primera imagen original de video.

Modelo de color

- Se obtiene la imagen y la se almacena en nuestro algoritmo como la variable temporal de entrada que consiste de tres matrices de valores en el espacio RGB . Cada matriz corresponde a cada uno de los canales R , G y B .
- Se cambia el tipo de formato de las matrices a doble.
- Según el tipo de valores de la imagen en el espacio RGB se hace la conversión al espacio de color HSI en la imagen de las ecuaciones (3.13) de la sección 3.1.3. Esta conversión da otras tres matrices de valores representativos de la imagen pero ahora en valores H , S e I .
- Se ubican los valores H , S , e I de la imagen en los rangos arriba mencionados. ($H=0-360$. $I=0-255$. $S=0-255$).

Clasificación del color.

- Se toman los valores H , S e I como patrones de datos y se aplica el algoritmo de clasificación k -medias, el cual, clasifica según los valores de los canales del espacio de color y los centroides de inicio de clases en color de piel y no piel. Los centroides se tomaron como en las pruebas realizadas en las imágenes de la figura 5.2. La salida de este paso es una imagen segmentada con los píxeles clasificados en piel y no-piel (Figura 5.4). La segmentación se realizó como sigue:
Blanco-píxeles de color piel y Negro-píxeles color no piel.



Figura 5.4. Imagen segmentada de salida del algoritmo *k-medias*

Este paso se muestra como resultado del procesamiento de la imagen; por supuesto la imagen segmentada no se despliega, solo sirve como muestra de prueba de la detección.

Depuración.

- A continuación se realiza la depuración de la imagen segmentada. Se toman los píxeles en blanco de la imagen de la figura 5.4 y se comparan con sus correspondientes píxeles de la imagen original en valores de *HSI*. Los píxeles cuyos valores no se encuentren en los rangos $H_{piel}(295-35)$, $S_{piel}(52-125)$ e $I_{piel}(60-125)$ se toman como píxeles erróneos y se desechan, poniéndolos en negro en la segmentación de la imagen. Los valores en el canal *H* se toman de acuerdo al modelo de triángulo de color *HSI* (figura 3.3), donde los ángulos representativos del valor *H* para el color de la piel están entre los ángulos de 295 a 35 grados. La imagen de salida después de la depuración de errores se muestra como sigue (Figura 5.5):

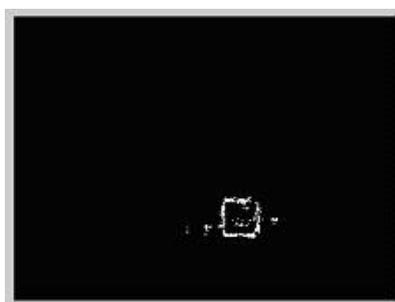


Figura 5.5. Imagen segmentada de salida del proceso de depuración de errores.

- Se determina la región de interés más grande. Una vez encontrada piel en la imagen, se determina el tamaño de las regiones, en caso de que hubiera más de una, tomando el número de píxeles que abarquen cada una. La región más grande será la que se mantendrá para el procedimiento siguiente, la o las regiones más pequeñas se desecharán y no se tomarán en cuenta.

La salida de este paso da una imagen con una sola región de interés, que se muestra como una imagen segmentada con la región de interés (piel) en blanco y el fondo de la imagen en negro (Figura 5.6).

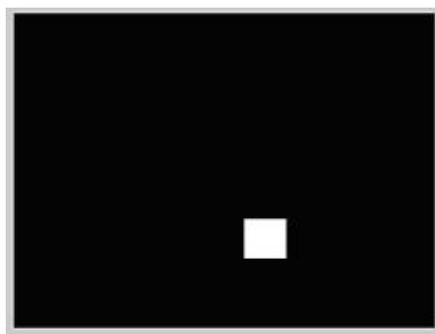


Figura 5.6. Imagen segmentada con la región de interés captada.

- Se obtienen las coordenadas de las esquinas superior izquierda e inferior derecha y se calcula el tamaño rectangular de la región de piel encontrada. De la misma información se obtienen las coordenadas del objeto en la imagen, las que se utilizan para el método de calibración desarrollado más adelante.
- En este paso termina el proceso de detección. Se muestra la región de piel captada como un rectángulo rojo bordeando la región de interés en la imagen original tomada. La imagen original captando y encerrando la mayor región de piel se despliega como sigue:



Figura 5.7. Imagen original con la región del objeto detectado en rojo.

5.2. CALIBRACIÓN DE LA CÁMARA.

Para el desarrollo de la calibración de la cámara explicado en la sección 4.2 se tomaron diversas consideraciones que se han adaptado según el tipo de cámara que se ocupa. Se planteó el problema de la calibración para obtener solamente los parámetros extrínsecos de la cámara R y T . Entonces se tuvo que encontrar los parámetros intrínsecos según el tamaño de la imagen y la configuración de la cámara. Primeramente, se fijó el tamaño de la imagen en 360 píxeles a lo ancho y 240 a lo alto según las opciones de configuración de la cámara. Así, se obtienen el tamaño efectivo del píxel y las coordenadas del centro de la imagen en la mitad de lo alto y de lo ancho. Además, se deshabilitó la opción del zoom de la cámara para obtener un valor fijo del foco. De lo anterior se determinaron los siguientes valores:

Longitud focal $f= 5.0$ cm.

Tamaño efectivo del píxel horizontal y vertical $S_x=0.0303$ cm y $S_y=0.0303$ cm.

Coordenadas del centro de imagen $(O_x, O_y) =(160,120)$ en píxeles.

También se necesita para los propósitos del trabajo hacer una estimación inicial de R y T . Se opta en este sentido, partir de una matriz de rotación R igual a la matriz identidad y el vector traslación T en base a diferentes pruebas de posibles posiciones de inicio del objeto. Así quedan definidos como:

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad T = [0 \ 0 \ 20.54 \ 1]$$

Una vez conocidos estos parámetros se procede a explicar los resultados de la calibración obtenidos de la información proporcionada por el proceso de detección realizado en la sección anterior. La situación en la que nos encontramos es que ya encontramos un objeto de interés y sabemos donde se ubica en nuestra imagen, ahora tenemos que encontrar la posición y orientación de la cámara con respecto al objeto encontrado y por tanto ubicar nuestro objeto en el espacio. Para tal efecto se realiza lo siguiente:

- De la posición del objeto se obtienen sus puntos 2-D en la imagen de la figura 5.7 y de acuerdo al último paso de la detección obtuvimos los siguientes puntos imagen:

$$p1 = (15, 85), p2 = (15, 60), p3 = (45, 60), p4 = (45, 85),$$

$$p5 = (15, 85), p6 = (15, 60), p7 = (45, 60) \text{ y } p8 = (45, 85)$$

Algunos puntos del objeto en la imagen se determinan según las dimensiones de la región encontrada, es decir, si la longitud de todos los lados de la región de piel son aproximadamente iguales, se asume que estamos viendo solo una cara del cubo y se toman cuatro vértices como referencia. Pero, si no son iguales, se considera que el cubo esta girado hacia algún lado y se hace una simple relación en base a la geometría del cubo para encontrar esos puntos. De la imagen 5.7, la región de piel es aproximadamente igual en

todos sus lados, por lo tanto el cubo esta de frente, y es por ello que se toman los puntos como se muestran arriba.

- Ahora recurrimos al modelo matemático 3-D explicado en el capítulo anterior y en el cual definimos el modelo del cubo de acuerdo a sus dimensiones como se muestra en la figura 5.8.

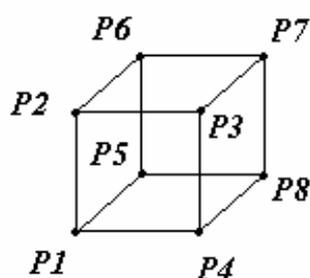


Figura 5.8. Configuración del modelo del cubo.

Cada vértice del cubo es punto de referencia del objeto y según las dimensiones del cubo real los definimos como:

$$P1=(-7.5,-7.5,20), P2=(-7.5,7.5,20), P3=(7.5,7.5,20), P4=(7.5,-7.5,20), \\ P5=(-7.5,-7.5,12.5), P6=(-7.5,7.5,12.5), P7=(7.5,7.5,12.5) \text{ y } P8=(7.5,-7.5,12.5).$$

Entonces, de acuerdo al las dimensiones de la región de piel encontrada, se toman el número de puntos de referencia del cubo en la imagen y se relacionan con sus correspondientes puntos del modelo del objeto. Así, se cuenta ya con la información necesaria para hacer la calibración de la cámara siguiendo los pasos explicados en el capítulo 4 en la sección de calibración.

- Se ejecuta el proceso descrito haciendo la calibración de la cámara y se obtienen los ángulos de rotación que alinean el eje óptico de la cámara con el centro del objeto,

es decir, se obtiene la dirección hacia la que se orienta la cámara para mantener el objeto en el centro de su campo de visión.

- En el caso de los puntos imagen y puntos 3-D del modelo obtenidos en los pasos anteriores se obtuvieron los ángulos de rotación y el vector de traslación siguientes:
Elevación (El): -7.5649 y Apertura (Ap): 1.0816
 $T=(0,0,123)$

Los ángulos de rotación se han definido tomando como origen el eje óptico de la cámara, por lo tanto la cuadrantes de los ángulos en grados se dan de la forma:

$Ap -$ $El +$	$Ap +$ $El +$
	$(0, 0)$
$Ap -$ $El -$	$Ap +$ $El -$

Figura 5.9. Cuadrantes de los ángulos de Apertura (Ap) y Elevación (El)

Estos ángulos de elevación y apertura son con los que se alimenta el programa del motor que maneja la base sobre la que está sentada la cámara y que permite llevar el foco de la cámara desde la posición actual hacia el centro del objeto. El vector traslación se considera como la información que nos dice cual es la posición que guarda el objeto con respecto a la cámara. Entonces la cámara se rota según los ángulos obtenidos y logra poner el objeto en el centro de su campo de visión.

5.3. SEGUIMIENTO DEL OBJETO.

En este momento se ha demostrado que el estudio es capaz de detectar objetos y ubicarlos en el espacio por medio de la calibración. Ahora toca hacer el seguimiento del objeto a medida que se mueva y hacer una calibración dinámica de la cámara cada determinado tiempo. El seguimiento del objeto se realiza aplicando el algoritmo de detección a cada imagen recibida, solo que con la particularidad de que en cada una de ellas se toma la información proveniente de la imagen anterior. También es necesario para el seguimiento hacer la calibración de la cámara cada determinado lapso de tiempo para saber que posición y orientación guarda el objeto con respecto a nuestra cámara. El tiempo estimado para hacer cada calibración de la cámara es cada 15 imágenes capturadas. A continuación describimos los pasos para llevar a cabo estos objetivos:

1. Lectura del archivo de video.
2. Se captura la secuencia de imágenes que son tomadas por la cámara. El software de la cámara está configurado para desplegar 15 imágenes por segundo.
3. Se toma la primera imagen de la secuencia (Figura 5.3) y le aplicamos el proceso de detección completo.
4. Si no se encontró un objeto de color piel regresar a tomar la siguiente imagen de la secuencia, si se encontró un objeto seguir adelante.
5. Se toma la última información de la detección como la mostrada en las Figuras 5.6 y 5.7. Hacemos más grande la región de interés en 20 píxeles a lo largo y a lo ancho con la finalidad de definir una región mayor para próximas búsquedas (Figura 5.10).
6. Se obtienen las dimensiones del objeto encontrado así como las coordenadas imagen del mismo.
7. Se ejecuta ahora el proceso de calibración de la cámara y obtenemos los ángulos de elevación y apertura de la cámara.
8. Se mueve la orientación de la cámara con el algoritmo que maneja el motor.

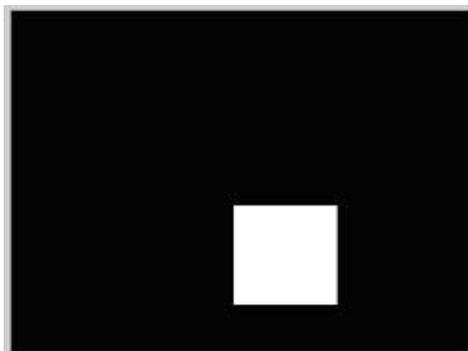


Figura 5.10. Imagen segmentada con regiones vecinas al objeto para el seguimiento.

9. El proceso se repite para cada una de las imágenes posteriores de la secuencia tomada, con la particularidad de que la búsqueda de regiones de piel se inicia en las vecindades de la región de piel encontrada en la imagen anterior.
10. La calibración se ejecuta cada 15 imágenes siempre que el proceso de detección nos diga que el objeto se encuentra en el campo de visión, de forma que a medida que el objeto este en movimiento los ángulos de rotación se van actualizando. Realizando el proceso completo nos permite realizar un buen seguimiento del objeto manteniéndolo en el centro de campo de visión la mayor parte del tiempo.

Un aspecto importante mencionado al inicio de este trabajo, es poder conjuntar los procesos de detección y calibración para realizar un buen seguimiento. Por ello, mostramos en la figura 5.11 un aproximado de los lapsos de tiempo del desarrollo de cada uno de ellos. En ella se toman una secuencia de 15 imágenes capturadas por la cámara, en la primera se aplica la detección y de la información de la detección se hace la calibración en el caso que hubiera un objeto detectado. A partir de la segunda imagen inicia el seguimiento del objeto en base a la detección de la imagen tomando la información de la imagen anterior. El proceso se repite para las siguientes imágenes siempre que el objeto se siga captando, pero, sin hacer la calibración hasta que llegamos a la imagen 15. En esta última imagen se detecta el objeto y se hace la calibración y el seguimiento. Como se ve, el proceso se retrasa un poco al hacer la calibración de la cámara y se agiliza al momento

de hacer el seguimiento en las imágenes intermedias de la secuencia. Es por ello que se inclinó por hacer la calibración de la cámara cada 15 imágenes, tomando en cuenta que no hay gran pérdida de información en ese lapso.

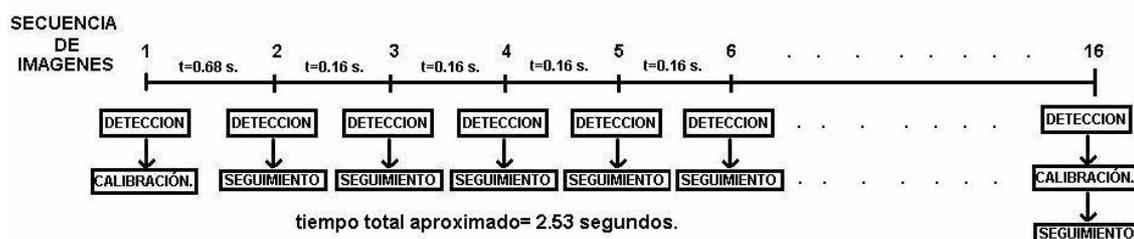


Figura 5.11. Tiempo de ejecución de la detección y calibración.

A continuación se muestran dos diferentes pruebas realizadas en secuencias de video. En la primera se muestran 6 de las 15 imágenes (Figura 5.12) tomados por la cámara en un segundo captando el cubo en posición fija. Se hace notar que en la última imagen se actualizan los ángulos de rotación tal y como se precisó en las pruebas experimentales de este capítulo.

La imagen No.1 es la imagen original recibida por la cámara. La imagen no.2 es el resultado del proceso de detección. De esta misma imagen se hace la calibración de la cámara y se obtienen los ángulos de rotación y vector traslación siguientes:

Elevación(El): 2.7581 y Apertura(Ap): -7.01

$T = (0.45, 0.1, 145)$

Las imágenes No. 3 a la No.6 son las resultantes del seguimiento basado en la búsqueda de objetos en regiones de la imagen donde se encontró el objeto en la imagen anterior. En la imagen No.6 que es la imagen 15 de la secuencia se obtienen nuevamente los ángulos de rotación y vector de traslación que al tomar el objeto fijo casi no cambiaron.

Elevación (El): 2.8697 y Apertura (Ap): -6.8907

$T = (0.5, 0.19, 145)$.



Imagen No.1.

$Ap = 0 \quad El = 0$



Imagen No.2.



Imagen No.3.



Imagen No.4.



Imagen No.5.



Imagen No.6.

$Ap = -6.89 \quad El = 2.869$

Figura 5.12. Secuencia de video en un segundo haciendo el seguimiento del objeto.

En la segunda prueba se muestran 9 imágenes de una secuencia de 80 (Figura 5.13) tomadas durante 2 segundos siguiendo al cubo en movimiento. En cada una de las

imágenes mostradas se iban actualizando los ángulos de rotación. Se aprecia que durante toda la secuencia se pudo localizar al objeto por medio de la detección, además de que la velocidad del movimiento del objeto es considerable.



Imagen No.11
 $Ap = 13.85$ $El = -0.4849$



Imagen No.12
 $Ap = 12.65$ $El = -1.342$



Imagen No.13
 $Ap = 12.08$ $El = -0.9324$



Imagen No.14
 $Ap = 11.1586$ $El = -0.8951$



Imagen No.15
 $Ap = 9.06$ $El = -1.417$



Imagen No.16
 $Ap = 8.30$ $El = -1.33$



Imagen No.17
 $Ap = 4.43$ $El = -2.01$



Imagen No.18
 $Ap = -0.85$ $El = -2.06$



Imagen No.19
 $Ap = -3.09$ $El = 2.05$

Figura 5.13. Secuencia de video en un segundo haciendo el seguimiento del objeto.

5.4. ALGORITMO DE SEGUIMIENTO.

El diagrama de flujo del proceso se puede apreciar en el Apéndice A, así como la aplicación completa del proceso en el Apéndice B. A continuación se presenta el algoritmo completo de la detección y seguimiento:

1. Lectura de la secuencia de imágenes captadas por la cámara que mantiene un patrón de movimiento de derecha a izquierda e izquierda a derecha.
2. Se toma la primera imagen y se aplica el método de detección de color.
3. Si se encuentra en la imagen regiones de color piel se procede a la depuración de errores y el patrón de movimiento de la cámara cambia, si no se regresa al paso 2.
4. Se hace la depuración de errores de la imagen.
5. En caso de haber más de un objeto de interés se opta por eliminar los más pequeños.
6. Se obtiene la región rectangular de la región captada como piel, en este caso del objeto y se delimita dentro de la imagen como rectángulo rojo.
7. Se obtiene una región de 20 píxeles a lo ancho y a lo alto de la del objeto como región de búsqueda en la siguiente imagen a procesar.
8. De acuerdo al objeto captado se obtienen N puntos imagen del objeto (coordenadas en píxeles en el plano de la imagen) y se toman como puntos de referencia.
9. Se toman las coordenadas del modelo del objeto previamente identificado y deben ser el mismo número de N puntos.
10. Se ejecuta el algoritmo de calibración usando los puntos imagen y los puntos del modelo.
11. La salida son los ángulos de rotación y vector de traslación.
12. El proceso de detección y depuración se realiza para cada imagen recibida. El algoritmo de calibración se ejecuta cada 15 imágenes a partir de que se detectó el objeto y hasta que se encuentre en el campo de visión.

5.5. CONCLUSIONES.

En este Capítulo se ha llegado a la ejecución de nuestro proyecto. Hemos descrito y analizado paso por paso como se lleva acabo la detección de un objeto, su seguimiento y la calibración de la cámara. Se ocupó como objeto de práctica un cubo por su geometría y dimensiones conocidas que nos permitan tener un modelo de objeto con la mayor confiabilidad en la información. Se ha elaborado un algoritmo completo pero fácil de entender de todo el proceso. Este comprendió nuestro estudio y resume en gran medida la información agrupada en los primeros cuatro capítulos de este documento. No hay mejor forma de apreciar la eficacia de un método, que su rendimiento en la práctica; tal es el caso del proceso de detección que agrupa de buena forma el color del objeto que estamos buscando, y el cual, usando el espacio de color *HSI* con todas las implicaciones y condiciones también mencionadas nos da una mejora sustancial en la detección del objeto.

6. CONCLUSIONES.

En este proyecto se estudió, trabajó e implementó un método para calibrar una cámara en base al seguimiento de características en una secuencia de imágenes. Se puso en práctica un método de detección de objetos en base a clasificación de color usando como herramienta principal el algoritmo *k-medias* en su modalidad difusa.

Se analizaron los métodos de calibración de una cámara y en base a ello se decidió utilizar un proceso iterativo de la estimación de los parámetros extrínsecos de la cámara que consiste en la solución de un sistema de ecuaciones lineales. El objetivo de resolver el sistema de ecuaciones es eliminar el grado de error de una estimación inicial de la rotación y traslación. Cada iteración realizada nos da una mayor aproximación a la posición y orientación de la cámara con respecto a un objeto.

Al iniciar este proyecto se plantearon objetivos generales y particulares entre los que podemos mencionar:

- Hacer una detección efectiva de objetos en base a un color predefinido.
- Extender la detección a objetos en movimiento.
- Obtener con precisión puntos de referencia de las imágenes.
- Realizar una efectiva calibración de la cámara.
- Ubicar con precisión objetos de interés en el espacio.

Al término del trabajo se puede decir que se ha logrado hacer una efectiva detección de un objeto por clasificación de píxeles en base color usando el *k-medias*. Aplicando este método de detección a cada una de las imágenes que componen una secuencia de video, se ha podido detectar objetos en movimiento. Se han obtenido con buena precisión puntos de referencia del objeto, aunque para realizar una efectividad mayor en la calibración aún hay detalles por mejorar. Tomar estimaciones iniciales más precisas y disminuir el tiempo de ejecución son parte de esos detalles. De la misma forma ubicar los objetos en el espacio ha

dado buenos resultados, pero también faltan aspectos de considerar para hacerlo un método más completo. Se han seguido objetos de manera efectiva, pero aún se puede agilizar más este proceso para hacer el seguimiento de objetos que se muevan muy rápido. La ejecución de los procesos en tiempo real de la detección es buena, no tanto del proceso de calibración.

6.1. APORTACIONES.

Dentro de las principales propuestas del trabajo se consideró usar el modelo de color *HSI* para determinar diferentes aspectos de color en las imágenes que componen una secuencia de video que no podrían ubicarse con el tradicional modelo de color *RGB*. Los canales del *HSI* nos sirvieron como patrones de información para alimentar el *k-medias*. La correcta elección de los centros de inicio del *k-medias* fue crucial para la buena ejecución del mismo y obtener los resultados deseados a la mayor velocidad posible. Además se complementó la detección usando procesos de depuración de la imagen, tanto para la eliminación de errores como para la correcta elección del objeto.

Probamos la eficiencia del proceso sometiendo imágenes en diferentes condiciones de ambiente a la detección, y se identificó bajo que condiciones es más o menos efectivo.

Se analizaron los métodos de calibración de una cámara y en base a ello decidimos utilizar un proceso iterativo de la estimación de los parámetros extrínsecos de la cámara que consiste en la solución de un sistema de ecuaciones lineales.

Para todo método de calibración es necesario definir puntos de referencia en la escena que nos permitan ubicar ciertos objetos con respecto a la cámara. A diferencia de la mayoría de trabajos anteriores que usan patrones de calibración ya establecidos, nosotros optamos por el uso de modelo de objeto. Conociendo las características del objeto que vamos a encontrar y seguir, definimos dicho modelo, que nos permite a partir de él, definir los puntos de referencia usados para la calibración de la cámara. En las pruebas

experimentales el modelo consistió de un cubo, ya que el objeto que detectamos era un cubo de madera de color y textura muy aproximados al de la piel humana.

Para la calibración se tomaron en cuenta las condiciones actuales de las cámaras, y por tanto planteamos inicialmente que los parámetros intrínsecos de la cámara eran conocidos y simplificamos el problema a encontrar los parámetros extrínsecos compuestos por la matriz de rotación y el vector de traslación.

En general, los experimentos realizados mostraron buenos resultados; muchas de las secuencias de imágenes de pruebas incluyeron movimiento de un objeto, en las que el seguimiento del objeto, toma una importancia fundamental.

Se ha elaborado este trabajo con la idea de realizar la calibración de la cámara en base a la detección de características en una secuencia de imágenes y a la definición de modelo de un objeto.

Además se propuso el método de calibración tomando puntos de referencia variables. En la situación de movimiento del objeto en la escena, los puntos de referencia del objeto en la imagen van cambiando. Por ello, se implementó la calibración de la cámara para poder ubicar el objeto en el espacio en cada nueva posición en la que el objeto se encontraba.

6.2. LIMITACIONES.

Durante el análisis de las pruebas experimentales notamos que nuestro método de detección es sensible aún, en escenas con alto contraste en sombras, lo que dificulta en buena medida la detección completa de algún objeto o persona. De igual forma, pequeñas variaciones en la luz durante la toma de la secuencia de imágenes cambia en pequeña escala las dimensiones en las regiones de piel detectadas durante las pruebas en las que un

objeto no está en movimiento. Esta situación dio como resultado ligeros cambios en las actualizaciones de la rotación y traslación.

Como es de esperarse, la precisión del método de calibración esta limitado por la precisión con que se defina el modelo del objeto y la efectividad con que se realicé la detección de características en la imagen.

El trabajo no está capacitado aún para realizar la detección y seguimiento de más de un objeto en movimiento. Además, el seguimiento es sensible a errores en caso de movimiento demasiado rápido.

6.3. TRABAJO A FUTURO.

Así que un paso adelante a nuestro trabajo, es lograr la definición de modelos más complicados. La idea de utilizar el color de la piel para hacer la detección de objetos, es que una vez que definamos modelos precisos de las partes del cuerpo humano, seremos capaces de hacer reconocimientos y seguimientos muy eficaces de cualquier ser humano. También es importante poder realizar seguimiento de uno o más objetos con movimientos usando las características de la imagen aquí detalladas. Un reto importante, es también mejorar aspectos de seguimiento es casos de movimientos muy rápidos o de varios objetos en movimiento..

REFERENCIAS.

[Chiori y Parvin, 1996], Chiori T. y Parvin B., “Active Calibration and Tracking in Video Streams”, <http://vision.lbl.gov/Publications/Calib>, April 9, 1996

[Cortijo, 2001], Cortijo F. J. Técnicas no supervisadas: métodos de agrupamiento. http://www-etsi2.ugr.es/depar/ccia/rtf/www/tema4_00-01_www/node11.html, Octubre, 2001.

[Dron, 1993], Dron, L. 1993. “Dynamic Camera self-calibration from controlled motion”, In Proc. IEEE. Conference on Computer Vision and Pattern Recognition, pp 501--506, New York, NY, 1993.

[Duda et al, 2001], Duda, R.O., Hart P. E. y Store, D. G., Pattern Classification, Wiley, New York, 2001.

[Eisert, 2002], Eisert P., “Model-based camera using analysis by synthesis techniques”. Proc. 7th International Workshop Vision, Modeling, and Visualization 2002, Erlangen, Germany, pp. 307-314, Noviembre 20-22, 2002.

[Fremont y Chellali, 2002], Fremont, V. Y Chellali, R., “Direct Camera Calibration using two concentric circles from a single view”, Presented at ICAT (International Conference on Artificial Reality and Telexistence) Tokyo, Dec. 2002.

[Fu et al, 1987], Fu, K. S., González, R.C. y Lee, C.S. G., 1987. Robotics: Control, Sensing, Vision and intelligence, McGraw-Hill, New York.

[Garcia G., et al, 2002], García M. Ginés, Vicente C. Cristina y García M. Andrés, “Localización y seguimiento de caras usando búsqueda en rejilla”, I Conferencia Iberoamericana en sistemas, cibernética e informática. Orlando, Florida (USA), 19-21 de Julio de 2002.

[González R.C. y Woods R. E. 1987], González R.C. y Woods R.E., 1992 Digital Image Processing. Adison-Wesley Publishing Company, 1992.

[Jones y Rehg, 1999], Jones J. Michel y Rehg James M., “Statistical colors models with application to skin detection”, Computer Vision and Pattern Recognition”, Fl. Collins co.pp. 274-280 Junio, 1999.

[Pajares y De la Cruz, 2002], Pajares G. y De la Cruz J. M., “Clasificación de texturas naturales mediante k-means”, Revista Electronica de visión por computador, No. 6, Marzo, 2002

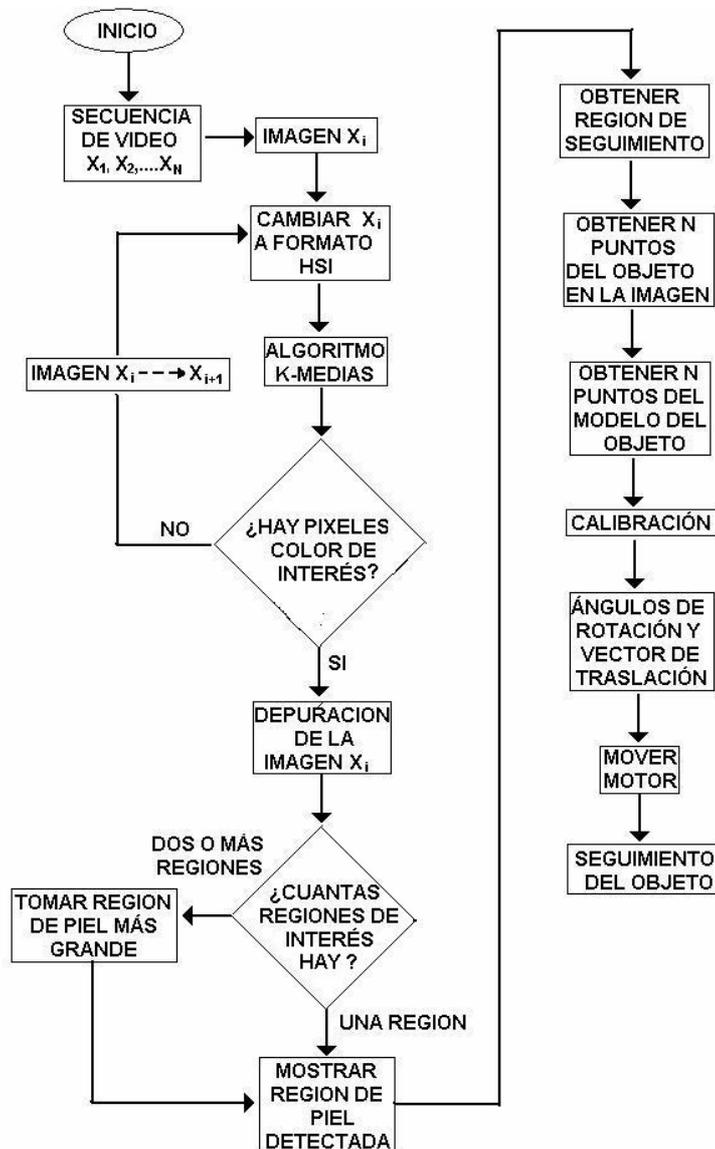
[Tommaselli y Tozzi, 1999], Tommaselli A.M.G. y Tozzi C.L, “Line based camera calibration in machine vision dynamic applications”, SBA Controle y Automacao Vol.10 No. 02, Agosto, 1999.

[Trucco y Verri, 1998], Trucco E. y Verri A., 1998. Introductory Techniques for 3-D Computer Vision. Prentice, Mayo, 1998.

[Zhang y Wang, 2002], Zhang C. Y Wang P., “A New Method of Color Image Segmentation Based on Intensity and Hue Clustering”, Proc. of ICPR, 2, 617-620, 2000.

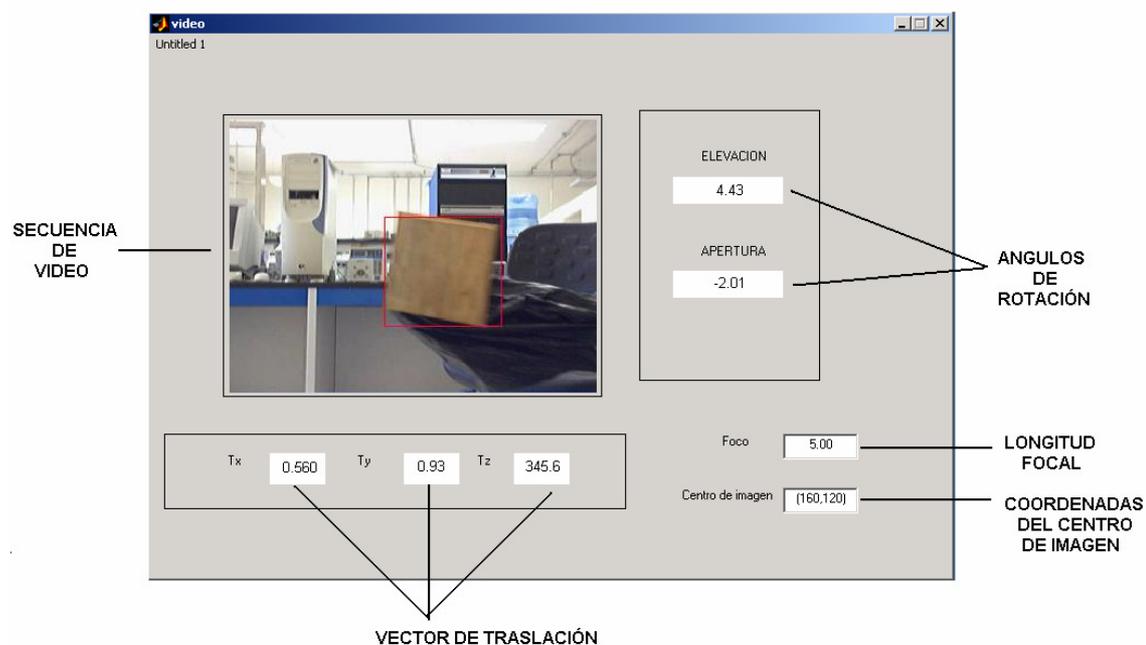
[Zhang y Schenk, 1997], Zhang, Z y Schenk V., “Self-Maintaining camera calibration over time”, Lb.CVPR Conference on Computer Vision and Pattern Recognition, 231-236.Puerto Rico, 1997.

APENDICE A. DIAGRAMA DE FLUJO DEL SEGUIMIENTO DE UN OBJETO.



CICLO DE SEGUIMIENTO. El proceso inicia tomando una imagen de la secuencia de video, se cambia a formato *HSI* y se aplica el algoritmo de clasificación. Después se depura la imagen y se determina cual es el objeto buscado. Se toman los puntos de referencia de la imagen y se calibra la cámara. Se actualizan los ángulos de rotación y se manda la información al motor para que lleve el campo de visión de la cámara a ubicarlo frente al objeto.

APENDICE B. INTERFAZ GRAFICA DEL PROGRAMA DE CALIBRACIÓN.



La interfaz está diseñada para apreciar la actualización dinámica de los ángulos de rotación y el vector de traslación de la cámara con respecto a algún objeto detectado en la escena. En el recuadro se puede apreciar la secuencia de imágenes capturada por la cámara después de su procesamiento, es decir, vemos un objeto con las características de color buscadas, encerrado en un rectángulo rojo que nos dice que es la región de la imagen donde se encuentra el objeto detectado. En el extremo izquierdo, están situados los valores de los ángulos de elevación y apertura que nos ubican el eje óptico de la cámara con el centro del objeto encontrado. El valor de los ángulos de rotación se toman de acuerdo a los cuadrantes como se muestran en la figura 5.9.

En la parte inferior izquierda, se dan los valores de la traslación en X , Y y Z . Estos nos dicen la posición que guarda el objeto con respecto a la cámara.

Y en la parte inferior derecha están los valores de los parámetros intrínsecos de la cámara como se han configurado según las posibilidades de la misma.