

**DETECCIÓN DE CARACTERÍSTICAS GLOBALES SIMBÓLICAS EN
IMÁGENES DIGITALES**

TESIS

QUE PARA OBTENER EL GRADO ACADÉMICO DE

**DOCTOR EN CIENCIA Y TECNOLOGÍA EN LA ESPECIALIDAD DE
MECATRÓNICA**

PRESENTA

M.CyT. LEONARDO BARRIGA RODRÍGUEZ

DIRECTOR DE TESIS

DR. HUGO JIMÉNEZ HERNÁNDEZ

SANTIAGO DE QUERÉTARO, DICIEMBRE DEL 2014



CIENCIA Y TECNOLOGÍA

Director de Posgrado
PICYT – CIDESI
Querétaro

Los abajo firmantes, miembros del Comité Tutorial del alumno **Leonardo Barriga Rodríguez**, una vez leída y revisada la Tesis titulada “**Detección de características globales simbólicas en imágenes digitales**”, aceptamos que la referida tesis revisada y corregida sea presentada por el alumno para aspirar al grado de **Doctor(a) en Ciencia y Tecnología** en la opción terminal de **Mecatrónica** durante el Examen de Grado correspondiente.

Y para que así conste firmo la presente a los 16 días del mes de Diciembre del año dos mil catorce.

Dr. Hugo Jiménez Hernández

Tutor Académico



CIENCIA Y TECNOLOGÍA

Director de Posgrado
PICYT – CIDESI
Querétaro

Los abajo firmantes, miembros del Jurado del Examen de Grado del alumno **Leonardo Barriga Rodríguez**, una vez leída y revisada la Tesis titulada “**Detección de características globales simbólicas en imágenes digitales**”, aceptamos que la referida tesis revisada y corregida sea presentada por el alumno para aspirar al grado de **Doctor(a) en Ciencia y Tecnología** en la opción terminal de Mecatrónica durante el Examen de Grado correspondiente.

Y para que así conste firmamos la presente a los 16 días del mes de Diciembre del año dos mil catorce.

Dr. Jesús Carlos Pedraza Ortega

Presidente

Dr. Juan Manuel Ramos Arreguín

Secretario

Dra. Ana Marcela Herrera Navarro

Vocal

Dr. Jorge Alberto Soto Cajiga

Vocal

Dr. Hugo Jiménez Hernández

Vocal

Le dedico este trabajo a mi esposa Karla y a mis
hijos Sergio y Juan Pablo.

AGRADECIMIENTOS.

Agradezco ante todo a Dios por todas las cosas buenas que me ha dado en mi vida sobre todo la salud, algo tan importante para poder llegar a concluir cualquier proyecto de vida. A mi familia por estar a mi lado ofreciéndome su apoyo incondicional. A mi esposa Karla, por todo ese apoyo incalculable. A mis hijos Sergio y Juan Pablo, por ser mi motor para ser mejor cada día.

Agradezco al Dr. Carlos Rubio González, al Dr. José Cruz Pineda, al Dr. Fernando Hernández, al Dr. Alfonso Gómez, en su momento, mis autoridades inmediatas en el CIDESI, que me ofrecieron el apoyo para que pudiera continuar con mis estudios.

Agradezco ante todo, el apoyo de mi asesor el Dr. Hugo Jiménez Hernández, sin el cual no hubiera llegado de manera satisfactoria a terminar este trabajo.

Agradezco a mis sinodales, Dr. Jesús Carlos Pedraza, Dr. Juan Manuel Ramos, Dra. Ana Marcela Herrera, Dr. Jorge Alberto Soto, por sus comentarios y revisiones al trabajo.

A todos mis compañeros del Laboratorio de Modelación y Simulación, muchas gracias por su amistad.

A CIDESI por su apoyo con su infraestructura para el desarrollo de este trabajo.

RESUMEN

En este trabajo se desarrolla un método para determinar zonas que resulten invariantes a escala y rotación en una imagen el cual permite mapear a un espacio simbólico donde se define una métrica de similitud. Se propone una forma de ecualizar las imágenes con el objetivo de generar características con este método que sean consistentes.

Debido a que el método propuesto se basa en la cantidad de información en la imagen a través de intensidades de los píxeles y sus variaciones, se busca un canal de color que ofrezca por su naturaleza la mayor información posible.

Se muestra el fundamento matemático del uso de este tipo de descriptor visual, tal que, el espacio de codificación, contiene propiedades que ayuda a sí mismo a clasificar un conjunto de datos y descubrir cuáles son los más importantes.

Se desarrolló un método para llevar a cabo pruebas orientadas a verificar el funcionamiento del descriptor ante situaciones de cambio de rotación, de escala o de perspectiva, con imágenes reales e imágenes artificiales.

ABSTRACT

In this work we develop a method for determining areas that are invariant to scale and rotation in an image which can be mapped to a symbolic space where a similarity metric is defined. We proposed a way to equalize the images in order to generate with this method features that are consistent.

Because the proposed method is based on the amount of information in the image using pixel intensities and its variations, a color channel that by their nature offer much information as possible is sought.

The mathematical basis for the use of this type of visual descriptor is shown, such that the coding space, contains properties that help themselves to classify a data set and find out what are the most important.

A method was developed to conduct aimed at verifying the operation of the descriptor in situations of change of rotation, scale or perspective, with real images and artificial images tests.

CONTENIDO

RESUMEN.....	6
ABSTRACT.....	7
ÍNDICE DE FIGURAS.....	10
ÍNDICE DE TABLAS.....	11
1 INTRODUCCIÓN.....	12
1.1 OBJETIVO GENERAL.....	14
1.2 OBJETIVOS PARTICULARES.....	14
1.3 HIPÓTESIS.....	14
1.4 JUSTIFICACIÓN.....	15
1.5 PLANTEAMIENTO DEL PROBLEMA.....	17
1.6 LIMITACIONES Y ACOTACIONES DE LA TESIS.....	18
1.7 ORGANIZACIÓN DE LA TESIS.....	19
2 ESTADO DEL ARTE.....	21
2.1 ANÁLISIS DE LOS DESCRIPTORES.....	21
2.2 HARRIS.....	21
2.3 SIFT.....	23
2.4 SURF.....	26
2.4.1 Detección de puntos de interés.....	26
2.5 BRISK.....	28
3 FUNDAMENTO TEÓRICO.....	30
3.1 ALGORITMOS PARA LA SELECCIÓN DE COLOR.....	30
3.1.1. Color y representación.....	32
3.1.2. Espacios de color.....	33
3.1.3. Espacio de color RGB.....	34
3.1.4. Espacio de colores opuestos.....	35
3.2. ALGORITMOS PARA ECUALIZACIÓN.....	36
3.2.1. Ecualización de imágenes con una variación de TopHat.....	36
3.2.2. Reconstrucción morfológica.....	39

3.2	ESPACIOS DE BINARIOS.....	41
3.2.3.	Fundamentos Matemáticos	41
4	PROPUESTA.....	49
4.1.	CARACTERÍSTICAS DE LAS IMÁGENES.....	49
5	MODELO EXPERIMENTAL Y RESULTADOS.....	52
5.1	MODELO EXPERIMENTAL	52
5.2	DISCUSIÓN DE RESULTADOS Y ANÁLISIS DE DATOS.	53
6	CONCLUSIONES.....	59
7	BIBLIOGRAFÍA.....	60

ÍNDICE DE FIGURAS.

Figura 2-1 Clasificación de imagen usando los valores propios con Harris.	23
Figura 2-2. Por cada Octava en diferentes escalas se aplica una Diferencia Gaussiana	25
Figura 3-1. Ecuación por tophat.....	37
Figura 3-2. Reconstrucción morfológica.....	40
Figura 3-3. Cubo distancia Hamming.	44
Figura 3-4. Espacio Hamming tridimensional	45
Figura 3-5. Distribución de distancia Hamming en espacio binario y de distancia Euclidiana.....	48
Figura 4-1. Histograma sobre un vecindario w	50
Figura 4-2. Representación de $f(w)$ en forma de cadena binaria.....	51
Figura 5-1. Representación de distancia de una característica tomada desde diferente perspectiva.	56
Figura 5-2. Distancia de una característica tomada desde diferente perspectiva.	57
Figura 5-3. Región de una imagen obteniendo superficie de distancia de características binarias y tensores de Harris.....	58

ÍNDICE DE TABLAS

Tabla 3-1. Ejemplo de diferencia entre distancia de Hamming y distancia Euclidiana.....	42
Tabla 3-2. Distribución del espacio binario.....	47
Tabla 5-1. Pruebas de imágenes con rotación	54
Tabla 5-2. Pruebas de imágenes con escalado	55

1 INTRODUCCIÓN

En visión por computadora y procesamiento de imágenes el concepto de **detección e identificación de características** se refiere a métodos que se dirigen a abstracción de la información de la imagen y realizar decisiones a nivel local en cada punto de la imagen. Algunos ejemplos de características incluyen: Bordes, Esquinas o puntos de interés, Blobs o regiones de interés, ridges o cordilleras, entre otros. Típicamente estos descriptores están asociados a alguna estructura geométrica de interés o a algún comportamiento intrínseco en la información codificada por el arreglo de sensores en la imagen.

La tarea de definir un descriptor está asociada directamente con el tipo de análisis a efectuar y muchas veces se predispone la existencia de estas estructuras en las imágenes de interés (Raoui, Bouyakhf, Devy, & Regragui, 2011). Sin embargo, en sistemas de visión un punto que retoma gran importancia, es cuando no se sabe explícitamente qué se quiere caracterizar o sólo se busca un descriptor global que permita en análisis posteriores para discriminar la información de una forma eficiente. En este contexto los modelos de características globales retoman una importancia grande, porque sirven como aproximaciones para extraer puntos de interés de forma robusta, e inherente al contenido de la información codificada en la imagen. En conjunción a esto, es necesario también establecer un criterio eficiente para medir la similitud de una característica, para poder discernir cuando dos regiones locales de información representan información relacionada. En este contexto, una de las tareas que requiere de buenos descriptores es la reconstrucción tridimensional. Pues es necesario establecer criterios para encontrar información distintiva en distintas proyecciones bidimensionales (imágenes) para alinear y empalmar distintas vistas.

La reconstrucción tridimensional consiste en encontrar una transformación F de una imagen I_k a otra I_{k+1} , de forma que las áreas visibles comunes en las imágenes puedan empatarse (si las tiene) y en el caso que no se mapeen a un espacio global de forma que cada imagen que se posee es una proyección parcial de este espacio. Entonces las coordenadas globales y el mapeo continuo de imágenes permite referenciar y determinar la geometría de los objetos involucrados en cada I_k ; para esto, el uso de características es importante porque estas permiten definir las regiones que bajo alguna medida de información (Goshtasby, 2012) permiten crear la función de transformación entre cada imagen. De lo anterior se tiene que la ubicación fiable de características de la imagen es un problema fundamental en las aplicaciones de visión por computadora, como la reconstrucción 3D a partir de un par de imágenes (Tuytelaars & Van Gool, 2000) , la estructura y estimación de movimiento (Yan Cui, 2009), la generación de panorámicas (Lowe M. B., 2007), o reconocimiento de objetos (Sivic & Zisserman, 2004). Especialmente, si el cambio de punto de vista en 3D entre las imágenes es grande, la adaptación de las características de la imagen debe ser invariante a transformaciones de imagen y cambios de iluminación.

Cuando se realiza reconstrucción tridimensional a través de imágenes es necesario contar con características que sean invariantes a rotación y escalado. Algunas de las características más importantes de este tipo son:

- Harris corner (Detector de esquinas de Harris), Fue propuesto por Chris Harris y Mike Stephens en 1988 y consiste en determinar las zonas con dos componentes en textura.
- SIFT (Scale Invariant Transform Features), David Lowe 1999, consiste en encontrar características que resulte invariante ante transformaciones de rotación y escala.
- SURF (Speeded Up Robust Features) presentado en 2008 por Herbert Bay et al (Bay, Ess, Tuytelaars, & Gool, 2008), y consiste en optimizar el banco de filtros de forma que resulte eficiente e su implementación.

1.1 OBJETIVO GENERAL.

Desarrollar un método para determinar zonas que resulten invariantes a escala y rotación en una imagen, que puedan ser usados para la reconstrucción de objetos 3D utilizando una cámara convencional.

1.2 OBJETIVOS PARTICULARES

1. Determinar qué efectos influyen en la detección de características globales.
2. Desarrollar un método para normalizar las condiciones lumínicas globales en una escena.
3. Determinar qué componente de color, a partir de un grupo de espacios de color, aporta más información para determinar las características globales en la imagen.
4. Crear un método para encontrar características globales en una imagen donde la topología del espacio ayude a desarrollar un esquema eficiente para definir un espacio métrico.

1.3 HIPÓTESIS

Si un proceso de representación y codificación a partir de un conjunto de sensores ópticos, permite mapear a un espacio simbólico donde se define una métrica de similitud de forma que la topología dada por esta métrica ayuda a definir regiones identificables; entonces es posible generar criterios que ayuden a discriminar las zonas con información que resulte invariante ante un conjunto de transformaciones (cambios de perspectiva, escala, y condiciones lumínicas).

1.4 JUSTIFICACIÓN

En el desarrollo de algoritmos robustos y automáticos, se necesitan buenos criterios para codificar datos. Estos criterios deben ser lo suficientemente robustos para que sean repetibles a pesar que las circunstancias en que los datos han cambiado. Un caso muy concreto es en el área de visión y análisis de imágenes; debido a que a pesar existan dos fotografías de un mismo lugar, en diferente temporalidad, o con diferente infraestructura (cámara) existen discrepancias que dificulta definir con un mismo criterio sobre las regiones (zonas) de mayor información. Esto ocasiona que en la búsqueda de criterios globales para determinar zonas con un contenido mayor de información se han desarrollado métodos que ayudan a discriminar características globales (Shi, 1994). Los métodos desarrollados utilizan un criterio de disparidad y medición de información que permite discernir sobre las zonas relevantes en la imagen. De entre los criterios más conocidos se utilizan criterios lineales para discriminar y medir la cantidad de información, bajo el supuesto que mayor información está representado por una mayor independencia lineal sobre una región determinada (Harris & Stephens, 1988), otros, por ejemplo utilizan los primeros momentos estadísticos y medidas derivadas (Kurtosis y Skew) para medir la distribución sobre un vecindario dado (Mendenhall, 2008); otras parten de una previa binarización de objetos analizando la topología de los objetos conexos, como es el caso del Numero de Euler en Imágenes (Horn, 1986); aunque tradicionalmente el mapeo entre espacios permite representar en forma de transformadas una imagen e igual calcular un conjunto de descriptores, tal es el caso de transformadas de Fourier (Chávez, 2012) o en su forma más general Transformada ondeleta (Meyer, 1993) que mapean en espacios bajo alguna estructura dada (conocida como kernel) la periodicidad sobre los datos contenidos en una imagen, cabe mencionar que en este tipo de aproximaciones análisis por SVD también es usado como una forma de reducción de información al proyectarlo sobre un espacio dimensionalmente

más compacto y fácil de caracterizar (Aase, Husoy, & Waldemar, 1999); de igual forma, con el afán de cubrir una cantidad mayor de invariancias, plantean la combinatoria de varias situaciones de interés, proponiendo metodologías para extraer información relevante a partir de los datos en la imagen (Lowe D. G., Object recognition from local scale-invariant features, 1999), otros más intentan proponer alternativas de muestreo sobre la información, con el objetivo de tener una mejor representación de la información y por lo tanto un método más efectivo para su aplicación (Leutenegger, Chli, & Siegwart, 2011).

En forma paralela al desarrollo de diferentes formas de analizar los datos, se tiene el problema físico de la invariancia y respetabilidad del sensor sobre un escenario para generar un mismo conjunto de datos. Esto es, para un mismo escenario, dos mediciones sobre la misma o diferente infraestructura, debe de generar datos con cierta respetabilidad. Por ejemplo, el caso de la resolución del sensor para denotar la zona de interés, pues, una imagen no es más que un muestreo discreto espaciado uniformemente, sobre una proyección bidimensional de un espacio 3D. Esto conlleva que cuando la proyección bidimensional sobre un objeto no garantiza las condiciones mínimas suficientes de representatividad por el muestreo la característica se pierda o altere su naturaleza. Esta limitante sin importar el tipo de característica siempre limitará cualquier método propuesto y se puede encontrar en proyecciones que resulten

Como se puede apreciar actualmente, la mayoría de las aproximaciones trabaja sobre una estructura previa que intenta definir la búsqueda de características globales. El número de invariancias resulta de las condiciones que utiliza. Si bien los trabajos resultan prácticos, en condiciones reales resulta un tanto complicado mantener un criterio certero debido a la cantidad de variables libres en el escenario y las condiciones no-controlables.

1.5 PLANTEAMIENTO DEL PROBLEMA

En este trabajo se presenta una propuesta distinta de buscar descriptores. Esta propuesta consiste en determinar un método de codificación simbólico, tal que el espacio de codificación, contenga propiedades que ayude a sí mismo a clasificar un conjunto de datos y descubrir cuáles son los más importantes. La importancia de las zonas de interés está basado en una métrica tal que para el espacio resulte adecuada; esto es, la distribución de la distancia de los elementos ayude a agrupar en cúmulos, y evitando una distribución plana sobre el espacio.

Esta propuesta está denotado por una metodología que incluye tres aspectos básicos:

- Selección del canal con la mejor información.
- Homogenización de las condiciones lumínicas globales.
- Determinación y selección de las zonas que son las mejores características de la imagen.

Por razones de extensión y claridad, las dos primeras etapas se consideran como etapas complementarias del trabajo y sólo se focalizará en el criterio y métrica de selección de forma eficiente de las zonas de interés. La representación y codificación de la información en un espacio de alta dimensionalidad proporcionan las invariancias morfológicas de la topología de los datos. Mientras que las invariancias lumínicas son cubiertas por las etapas previas (que no se discuten en el trabajo).

La búsqueda de descriptores efectivos en la imagen es una tarea abierta, porque las variables que se ven involucradas no son controlables directamente; es decir, son totalmente dependientes al escenario, la cámara y las condiciones.

Esto propicia que para una correcta reconstrucción sea necesario identificar características en forma automática que sean altamente confiables y distinguibles en las imágenes.

El interés de esta investigación es encontrar una metodología semi-automática que permita identificar características en imágenes de objetos o escenarios, que permitan extraer la información geométrica para su posterior reconstrucción tridimensional.

1.6 LIMITACIONES Y ACOTACIONES DE LA TESIS

1. Se utiliza siempre la misma cámara para evaluar el desempeño de repetibilidad de la propuesta.
2. La aproximación es automática para determinar las mejores zonas en una imagen, y sólo utiliza la información contenida en ella.
3. A pesar que es complicado tener una superficie lambertiana (Chávez, 2012), se controlan las situaciones lumínicas para incrementar repetibilidad.
4. Se tienen imágenes con sombras difusas.
5. Las fuentes de luz difusas afectan globalmente a la imagen.
6. El propósito del trabajo es demostrar la utilidad del método, por lo tanto se provee una complejidad teórica sobre una implementación ideal; la implementación se efectúa sobre software de prototipado.
7. La cámara utilizada contiene las siguientes características:
 - a. Resolución. 12.2 Megapíxeles
 - b. Sensibilidad. Rango de ISO de 100 a 1600

- c. Lente. Lente con estabilizador óptico de imagen EF-S 18-55 mm f/3.5 - 5.6 IS

1.7 ORGANIZACIÓN DE LA TESIS

Esta tesis se estructura de la siguiente manera:

En el capítulo 2 se muestra un análisis de los principales descriptores invariantes a rotación y escalado mencionados ampliamente en la literatura y los más comúnmente utilizados en la reconstrucción de escenas 3D. Se desarrolla el estado del arte de detección de esquinas de Harris, características SIFT y características SURF.

En el capítulo 3 se desarrolla el fundamento teórico sobre los diferentes espacios de color más utilizados en el procesamiento de imágenes, además de mostrar el método de ecualización de imágenes a través del método morfológico denominado TopHat. En este capítulo también es desarrollado los fundamentos matemáticos de un espacio binario, con todas sus propiedades y como esto es tomado como ventajas para el descriptor propuesto.

En el capítulo 4 se presenta la propuesta del descriptor simbólico, se muestra el proceso de cómo son extraídas las características en las imágenes y como son representadas.

En el capítulo 5 muestra el desarrollo de la experimentación y los resultados obtenidos, principalmente al realizar pruebas de rotación, escalado y en diferente perspectiva con el método propuesto.

Posteriormente se muestran las conclusiones finales de este trabajo.

2 ESTADO DEL ARTE

Los descriptores son criterios para denotar información útil F bajo una estructura S en un conjunto de sensores luminicos S_I representados por un arreglo I , bajo un proceso de muestreo L . La naturaleza de cada uno de los elementos es lo que determina el tipo, limitante y complejidad de cada descriptor.

2.1 ANÁLISIS DE LOS DESCRIPTORES

La detección de objetos es crucial para la mayoría de tareas de visión por ordenador; particularmente, en aplicaciones que requieren el posterior reconocimiento de estos. Los primeros enfoques para la solución del reconocimiento de objetos por computadora se basan en la búsqueda de correspondencia entre modelos geométricos del objeto y características en la imagen. Para evitar la necesidad de poseer tales modelos a priori, surgió en las últimas dos décadas el paradigma del reconocimiento de objetos basado en apariencia, usando técnicas de reducción de dimensionalidad tales como el análisis de componentes principales. Desafortunadamente, la correspondencia basada en apariencia es propensa a fallar en situaciones con pequeñas oclusiones o bajo distintas condiciones de iluminación o del fondo. Recientemente, el reconocimiento de objetos se ha beneficiado de estrategias que combinan la combinación de características geométricas locales, así como de apariencia. El más popular de estos métodos, es quizá, el uso de descriptores SIFT (Shi, 1994).

2.2 HARRIS

Fue definido por los investigadores Chris Harris y Mike Stephens en 1988. Los puntos detectados por este procedimiento permanecen invariantes ante

cambios en la escala, rotación, iluminación y ruido de la imagen. Este detector define la siguiente matriz $C(x, y)$, que se calcula sobre una sub-ventana de tamaño $n \times n$ para cada punto de interés en la posición (x, y) .

$$C(x, y) = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix}$$

Donde I_x e I_y son los gradientes vertical y horizontal de la imagen. Si definimos a λ_1 y λ_2 como los valores propios de la matriz $C(x, y)$, podemos definir entonces la función de autocorrelación R como:

$$R = \lambda_1 \lambda_2 - K(\lambda_1 + \lambda_2)$$

Donde K es un parámetro experimental. En esta función R vamos a distinguir tres casos posibles:

1. La función tendrá un máximo si ambos valores propios son altos, lo que quiere decir que un desplazamiento en cualquier dirección va a producir un incremento importante, indicando, por tanto, que se trata de una esquina.
2. La función será casi nula si ambos valores propios son bajos, es decir que un desplazamiento en cualquier dirección va a producir cambios muy pequeños, por tanto, la región que engloba la subventana de tamaño $n \times n$ es de intensidad constante (pertenece al mismo objeto).
3. Si un valor propio es alto y el otro bajo, entonces, la función tendrá forma de cresta. Por tanto, sólo los desplazamientos en una dirección van a producir pequeños cambios en $C(x, y)$ y cambios significativos en la dirección perpendicular. Esto indicará la presencia de un borde.

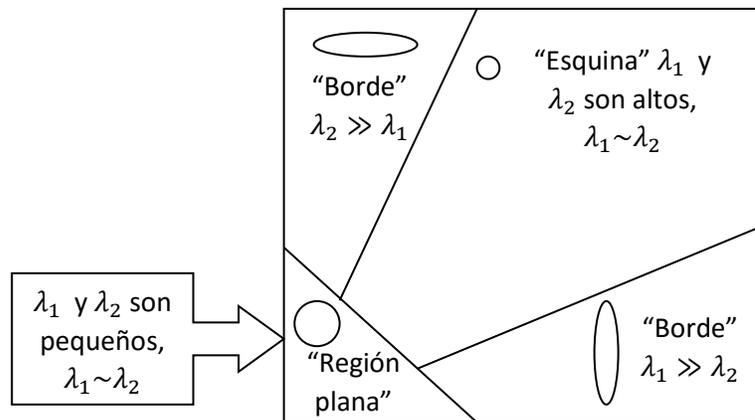


Figura 2-1 Clasificación de imagen usando los valores propios con Harris.

2.3 SIFT

Este detector (Brown, 2002) fue presentado por el investigador y profesor de la Universidad British Columbia, David G. Lowe, en el año 1999. Los puntos de interés encontrados con este método se caracterizan por permanecer invariantes ante los posibles cambios producidos en los siguientes parámetros de nuestro sistema:

- Iluminaciones globales que no afecten la forma de la textura y no saturen o no sean medibles por el sensor.
- Ruido de la imagen, siempre y cuando la intensidad del mismo no comprometa la relación de la textura mínima a definir.
- Rotación es frontoparalelas al sensor.
- Escalado de la imagen.
- Pequeños cambios de orientación y proyección siempre y cuando la perspectiva no afecte o cambie la forma de la textura.

En cuanto a las características de este detector, éste se caracteriza principalmente por generar una gran cantidad de descriptores muy estables

y por presentar un buen rendimiento en relación a la velocidad de cálculo y precisión (Brown, 2002).

El algoritmo se realiza mediante 4 pasos:

1- Construcción de Pirámides de Scale-Space: Se representa la imagen en diferentes escalas y tamaños. Se lleva a cabo de manera eficiente mediante el uso de la función de diferencia gaussiana, para identificar los posibles puntos de interés que son invariables a escala y orientación.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$
$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} * \left(-\frac{x^2 + y^2}{2\sigma^2} \right)$$

Para lograr una detección eficiente de puntos de interés estables, este detector sugiere la utilización en el espacio-escala de la convolución de una diferencia de Gaussianas con la imagen. De modo que esto queda de la siguiente manera:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)$$
$$= L(x, y, k\sigma) - L(x, y, \sigma)$$

Donde $D(x, y, \sigma)$ es la diferencia de los resultados de difuminar la imagen de entrada con una Gaussiana de escala $k\sigma$ y σ y el vecindario sobre el cual se tiene que reconstruir la gaussiana.

Los motivos que llevan a la utilización de esta función D (también llamada Diferencia de Gaussianas) son por un lado su mayor eficiencia al tratarse simplemente de la resta de dos imágenes, y por otro lado debido a que nos

da una aproximación más precisa del Laplaciano normalizado de una Gaussiana.

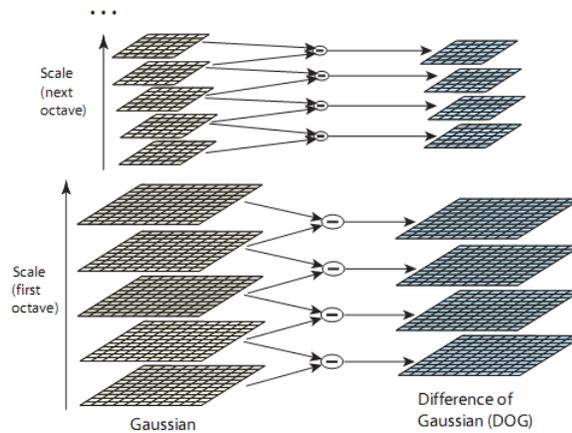


Figura 2-2. Por cada Octava en diferentes escalas se aplica una Diferencia Gaussiana

2- Localización de puntos clave: Se buscan aquellos puntos que se mantienen en cuanto a cambios de escala, para ello se debe estudiar cada píxel y realizar una comparación con los píxeles vecinos. Los puntos clave son seleccionados en base a las medidas de su estabilidad.

3- Asignación de orientación: en este paso se asigna a cada punto clave una dirección de acuerdo a las direcciones del gradiente y a la zona que rodea dicho punto.

4- Descriptor de Puntos clave: consiste en calcular un descriptor para la región de la imagen local que sea fácilmente identificable, sin embargo tan invariable como sea posible a las variaciones restantes, tales como el cambio en la iluminación o el punto de vista 3D.

2.4 SURF

Este detector fue presentado en 2008 por Herbert Bay et al (Bay, Ess, Tuytelaars, & Gool, 2008). Aunque guarda cierta similitud con el detector SIFT, el detector SURF presenta notables diferencias con éste. En comparación con SIFT, sus autores afirman que este detector presenta principalmente dos mejoras:

- Velocidad de cálculo considerablemente superior.
- Mayor robustez ante posibles transformaciones de la imagen.

Para conseguir estas mejoras, el detector SURF reduce la dimensión y la complejidad de los descriptores obtenidos, pero lo hace de modo que éstos continúen siendo suficientemente característicos e igualmente repetitivos.

A continuación, vamos a describir los pasos necesarios para obtener los descriptores SURF.

2.4.1 Detección de puntos de interés

Este detector está basado principalmente en la matriz Hessiana, ya que usa una aproximación muy básica de ésta. El motivo de usar la matriz Hessiana se encuentra en su buen rendimiento en cuanto a la velocidad de cálculo y la precisión. Sin embargo, este detector, al contrario del método empleado por otros detectores, en vez de usar una medida diferente tanto para elegir la posición como la escala, lo que hace es emplear el determinante de la matriz Hessiana en ambos casos. Por tanto, dado un punto $p = (x, y)$ de la imagen I , la matriz Hessiana $H(p, \sigma)$ en el punto p a la escala σ se define como:

$$H(p, \sigma) = \begin{bmatrix} L_{xx}(p, \sigma) & L_{xy}(p, \sigma) \\ L_{xy}(p, \sigma) & L_{yy}(p, \sigma) \end{bmatrix}$$

Donde $L_{xx}(p, \sigma)$ es la convolución de la derivada parcial de segundo orden de la Gaussiana $\frac{\partial^2}{\partial x^2} g(\sigma)$ con la imagen I en el punto p . Lo mismo ocurre para $L_{xy}(p, \sigma)$ y $L_{yy}(p, \sigma)$.

A pesar de que los filtros Gaussianos son óptimos para el análisis del espacio escala, debido a una serie de limitaciones de estos filtros (como la necesidad de ser discretizados, la no prevención totalmente del efecto aliasing, etc.), se ha probado en el detector SURF con una alternativa a los filtros Gaussianos: los filtros de caja. Éstos aproximan las derivadas parciales de segundo orden de las Gaussianas y pueden ser evaluados de manera muy rápida usando imágenes integrales, independientemente del tamaño de éstas. La escala más pequeña que vamos a tener corresponde a un filtro de caja de dimensión 9×9 , correspondiente a las aproximaciones de la derivada parcial de segundo orden de una Gaussiana con $\sigma = 1,2$. Las aproximaciones de las derivadas parciales se denotan como D_{xx} , D_{xy} y D_{yy} . En cuanto al determinante de la matriz Hessiana, éste queda definido de la siguiente manera:

$$\det(H_{approx}) = D_{xx}D_{yy} - (0,9D_{xy})^2$$

Los espacios de escala son a menudo implementados como pirámides de imágenes, en las que éstas son suavizadas repetidamente con un filtro Gaussiano y posteriormente submuestreadas para alcanzar un nivel más alto en la pirámide. En el detector SURF, debido al uso de filtros de caja e imágenes integrales, no se tiene que aplicar iterativamente el mismo filtro a la salida de una capa filtrada previamente, sino que se pueden aplicar dichos filtros de cualquier tamaño a la misma velocidad directamente en la imagen original. De modo que el espacio escala es analizado por medio de ir elevando el tamaño del filtro, en vez de ir reduciendo el tamaño de la imagen (tal y como se hacía con el detector SIFT). La salida obtenida de aplicar el filtro de dimensión 9×9 es la considerada como escala inicial ($s = 1,2$,

correspondiente a una Gaussiana con $\sigma = 1,2$). Las sucesivas capas se van obteniendo como consecuencia de aplicar gradualmente filtros mayores. En cada nueva octava, el incremento de tamaño del filtro es el resultado de doblar el incremento realizado en la octava anterior:

Octava inicial: $9 \times 9 \xrightarrow{6} 15 \times 15 \xrightarrow{6} 21 \times 21 \xrightarrow{6} 27 \times 27$

Siguiente octava: $15 \times 15 \xrightarrow{12} 27 \times 27 \xrightarrow{12} 39 \times 39 \xrightarrow{12} 51 \times 51$

Siguiente octava: $27 \times 27 \xrightarrow{24} 51 \times 51 \xrightarrow{24} 75 \times 75 \xrightarrow{24} 99 \times 99$

Y así sucesivamente...

Simultáneamente, los intervalos de muestra para la extracción de puntos de interés pueden ser también doblados.

Finalmente, para localizar los puntos de interés en la imagen y en todas las escalas, se procede a la eliminación de los puntos que no sean máximos en una región de vecindad $3 \times 3 \times 3$. Entonces, el máximo determinante de la matriz Hessiana es interpolado en la escala y el espacio de la imagen. Con esto ya estaría realizada la detección de los puntos de interés.

2.5 BRISK

Dado un conjunto de puntos clave (que consisten en sub-píxel refinado ubicación de las imágenes y valores de la escala de coma flotante asociados), el descriptor BRISK se compone como una cadena binaria concatenando los resultados de las pruebas de comparación de brillo simples. Esta idea se ha demostrado en (Leutenegger, Chli, & Siegwart, 2011) para ser muy eficiente, sin embargo aquí se emplea de manera mucho más cualitativa.

Con BRISK, se identifica la dirección de la característica de cada punto clave para permitir descriptores de orientación – normalizado y por lo tanto lograr invariancia rotación que es clave para robustez general. El concepto clave del descriptor BRISK es que hace uso de un patrón utilizado para el muestreo de la vecindad del punto significativo.

3 FUNDAMENTO TEÓRICO.

En este capítulo se muestra el fundamento teórico usado en la propuesta. Este fundamento se estructura de la siguiente manera: En la primera parte se retoma el concepto de selección de mejor componente de color basado en un algoritmo que busca el subespacio ortogonal formado de la información de varias componentes de color de diferentes espacios de color; luego se muestra un método y finalmente se presentan los fundamentos teóricos que tienen los espacios simbólicos de alta dimensionalidad.

3.1 ALGORITMOS PARA LA SELECCIÓN DE COLOR

Una cámara es un arreglo de foto-sensores, que codifica una cierta longitud de onda como una intensidad o medida. Por lo general, está constituida por varios foto-sensores pre-configurados a diferentes longitudes de onda del campo electromagnético. Las intensidades de una serie de foto-sensores están relacionados con la tonalidad y longitud de onda del color percibido. Típicamente, el color se aprecia como una mezcla de ciertas longitudes de onda reflejadas sobre un sensor dando la sensación de la percepción de color. El número de foto-sensores y el umbral de longitud de onda de cada uno describen las tonalidades de color del escenario percibido.

El análisis de color consiste en determinar cuáles longitudes de onda resultan útiles para denotar ciertas zonas con información de la imagen. Un espacio de color es apreciado como un conjunto de longitudes de onda básicos utilizados para crear los colores de la mezcla entre ellos. La naturaleza de las longitudes de onda utilizada como base y las limitaciones físicas definen las propiedades del espacio, tales como el número de colores que se pueden representar, o las operaciones que se pueden representar en el espacio. Los espacios de color más comunes son RGB (Susstrunk, Buckley, & Swen,

1999), CMYK (Gatter, 2005), HLS (Levkowitz & Herman, 1993), por mencionar algunos. La aplicación de cada uno de ellos se puede apreciar en diferentes trabajos (Vartak & Mankar, 2013), (Ramanath, S., Yoo, & Drew, 2005), (Skarbek & Andreas Koschan, 1994), (Bhattacharayya, 2011). Usan las propiedades del espacio de color enfocándose en una tarea en particular, la explotación de las características se pueden utilizar para fines propios. El espacio de color RGB representa un estándar que se utiliza en varios dispositivos. Físicamente se utilizan sensores calibrados para responder a rojo, azul y verde de la longitud de onda, incluso algunos incluyen un color adicional para incrementar la sensibilidad de la longitud de onda.

En otro contexto, uno de los principales inconvenientes en el análisis del color es que no hay una manera de ordenar un espacio de color. Esto es, cada espacio se basa en componentes básicos relacionados con una base física o un criterio determinado; pero estrictamente no define una relación de orden sobre los datos que ayuden a segmentar diferentes mezclas de los componentes. Además de este problema, no existe un criterio para determinar los escenarios no controlados, que canal o cuál espacio se vuelve más confiable para analizarlo.

Un enfoque para identificar qué componente de color proporciona más información sobre la linealidad expresada en los datos, será muy útil a encontrar características en una imagen. Este proceso de discriminación se basa en cuantificar la contribución de cada uno de los componentes de color basado en componentes principales ortogonales. La información se mide como importancia de los componentes ortogonales compuestos por varios componentes de color. Después, un método experimental que se realiza en escenarios interiores / exteriores en los que se identifican los componentes principales.

3.1.1. Color y representación

El espectro electromagnético está constituido por todos los posibles niveles de energía de luz. El espectro visible consiste en la longitud de onda que puede ser apreciado por el ser humano. Incluso la percepción del color depende del proceso cerebral del observador, es posible asociar un color a una determinada longitud de onda. El color es la forma en que el sistema visual humano mide una parte del espectro electromagnético, aproximadamente entre 300 y 830 nm. Debido a ciertas propiedades del HVS no somos capaces de ver todas las combinaciones posibles del espectro visible, pero tendemos a agrupar varios espectros en colores. Los colores representan una longitud de onda particular de espectros visibles.

Los colores del arco iris incluyen todos los colores que se pueden producir por la luz visible de sólo una única longitud de onda, los colores espectrales puros o monocromáticos. La longitud de onda es medida en nanómetros y su frecuencia en Tera-hercios para los varios colores espectrales puros. Las longitudes de onda se miden en el aire o en vacío. El color no debe ser interpretado como un número definitivo de combinaciones de colores perceptibles, los colores espectrales puros forman un espectro continuo. La intensidad de un color espectral, en relación con el contexto en el que se ve, puede alterar su percepción considerablemente, por ejemplo, una baja intensidad de color naranja-amarillo es de color marrón, y una baja intensidad de color amarillo-verde es de color verde oliva.

Para gestionar diferentes tonalidades de color, se han propuesto diferentes enfoques (Palus, 1998). Se ajustan cada color como una mezcla de componentes básicos. Esos componentes se definen por consideraciones teóricas o físicas. La variedad de color se realiza por la mezcla de cada uno de los componentes básicos. Sin embargo, las limitaciones consideran para conformar cada color algunas veces definen la cantidad de diferentes mezclas realizadas. Para los casos, el espacio RGB, puede representar un

menor número de tonalidades que el espacio HSVI, como se comenta en la sección siguiente.

3.1.2. Espacios de color

Los componentes básicos usados para representar diferentes tonalidades conforman una base de un espacio de color. Un espacio de color es una anotación en la que se especifican los colores posibles. Diferentes bases se mezclan como tres o cuatro tuplas. Cada componente representa cualquier valor de un componente básico. Esto es, un espacio de color S se denota por un conjunto de componentes básicos $C_1 \dots C_n$, tal como n es el número de componentes y por lo general toma 3 o 4 valores. Cada C_i representa un posible valor de un componente particular. Típicamente, cada componente se normaliza entre el intervalo $[0, 1]$ o $[-1, +1]$.

Como se ha comentado anteriormente, dominio componente es continuo, pero las implementaciones físicas se limitan a las representaciones resolución. Incluso cada color se representa como una mezcla de los componentes básicos, que no garantiza que los componentes básicos están correlacionados o sean independientes. Esta situación, complican en algunos casos diferenciar cierta mezcla de colores. Además, los componentes básicos no están relacionados con una manera perceptiva para representar cada intensidad, que hace que no hay una relación de semi-orden en el espacio generado por la mezcla de los componentes. Ambas situaciones hacen complicado de analizar y definir los operadores útiles para ser implementado usando componentes básicos. Varios trabajos han estado trabajando para tratar con él, sin embargo este enfoque están lejos de ser general.

Los espacios de color se pueden clasificar en tres grupos principales (Poynton, 1995), (D'O., 1996), (Palus, 1998), (Wyszecki & Stiles, 2000).

- Espacios basados en HSV, que incluyen el espacio de color RGB, los colores oponente teoría de espacios de color base y los espacios de color fenomenales. Estos espacios de color están motivados por las propiedades de los HVS.
- Espacios de color de aplicación específica. Esta clasificación incluye los espacios de color adoptados de los sistemas de televisión (YUV, YIQ), sistemas de imagen (Kodak PhotoYCC) y sistemas de impresión (CMY (K)).
- Espacios de color CIE son espacios propuestos por la CIE (International Commission on Illumination) y tienen algunas propiedades de gran importancia como el dispositivo-independencia y linealidad perceptual (CIE XYZ, Lab y Mimos) Desarrollos tecnológicos recientes incluyen ciertas variaciones de colores RGB utilizados en la norma como HDMI para los formatos de vídeo de alta definición.

3.1.3. Espacio de color RGB

La mayoría de los dispositivos de captura de imágenes tienen un detector de luz LMS. Por lo general se refiere a estos dispositivos como RGB. El color se describe con tres componentes básicos: R, G y B. El valor de estos componentes es la suma de las respectivas funciones de sensibilidad y la luz entrante:

$$R = \int_{300}^{830} S(\lambda)R(\lambda)d\lambda$$

$$G = \int_{300}^{830} S(\lambda)G(\lambda)d\lambda$$

$$B = \int_{300}^{830} S(\lambda)B(\lambda)d\lambda$$

donde $S(\lambda)$ es el espectro de luz, $R(\lambda)$, $G(\lambda)$ y $B(\lambda)$, son las funciones de sensibilidad de los sensores de R, G y B, respectivamente. Esta transformación de la distribución de potencia espectral de un vector tridimensional es una técnica de compresión de gran alcance con una relación de compresión de más de 10: 1. Como podemos ver en las ecuaciones anteriores, los valores de RGB dependen de la función de sensibilidad específica del dispositivo de captura. Esto hace que el espacio de color RGB de un espacio de color dependiente del dispositivo. La principal desventaja del espacio de color RGB en aplicaciones con imágenes naturales es una alta correlación entre sus componentes. Su psicología no intuitiva es otro problema porque un ser humano tiene problemas con la visualización de un color definido con los atributos de R, G y B. Otro problema es la falta de uniformidad perceptual, es decir, la baja correlación entre la diferencia percibida de dos colores y la distancia euclidiana en el espacio RGB.

3.1.4. Espacio de colores opuestos

Un modelo simple de esta transformación es:

$$RG = R - G$$

$$YeB = 2G - R - G$$

$$WhBl = R + G + B$$

Hay también una versión logarítmica de la transformación anterior (después de la respuesta logarítmica de los conos) propuesto por Fleck, Forsyth y Bregler (Fleck, Forsyth, & Bregler, 1996)

$$RG = \log R - \log G$$

$$YeB = \log B - \frac{(\log R + \log G)}{2}$$

$$WhBl = \log G$$

Un espacio de color útil ha sido propuesto por Ohta (Ohta, Kanade, & Sakai, 1980). Este espacio de color es una aproximación de la transformación de Karhunen-Loeve del RGB (descorrelación de los componentes RGB)

$$I1 = \frac{R + G + B}{3}$$

$$I2 = \frac{R - B}{2}$$

$$I3 = \frac{2G - R - B}{4}$$

3.2. ALGORITMOS PARA ECUALIZACIÓN

En esta sección se consideran los algoritmos que se utilizan para homogenizar las variabilidades percibidas por el sensor. La forma en como homogeniza los cambios y las suposiciones sobre la fuente de luz, es lo que determina la naturaleza del método de ecualización

3.2.1. Ecualización de imágenes con una variación de TopHat

La elección de un filtro morfológico dado es impulsada por el conocimiento disponible sobre la forma, el tamaño y orientación de las estructuras que se desea filtrar. Por ejemplo, podríamos escoger una apertura de un elemento estructurante cuadrado de 2 x 2 para eliminar ruido de impulso o un cuadrado grande para suavizar bordes de los objetos, o aperturas en imágenes de escala de grises pueden ser utilizadas para compensar iluminación no uniforme en el fondo.

Restando una apertura a una imagen de la original es llamada transformación top-hat, la cual es denotada como:

$$TH_B(A) = A - (A \circ B)$$

Donde A es la imagen de entrada original, B es el elemento estructurante.

En efecto, el enfoque adoptado con top-hat consiste en utilizar el conocimiento sobre las características de las formas que no son compartidas por las estructuras relevantes de la imagen. Una apertura con un SE que no se ajusta a las estructuras relevantes de la imagen se usa para eliminarlos de la misma. Estas estructuras se recuperan a través de la diferencia aritmética entre la imagen y su apertura. El éxito de este enfoque se debe al hecho de que no existe necesariamente una relación uno a uno entre el conocimiento sobre lo que es un objeto de la imagen y lo que no lo es. Por otra parte, a veces es más fácil eliminar los objetos relevantes de la imagen que tratar de suprimir los irrelevantes.

Si todos los objetos de la imagen tienen el mismo contraste local, es decir, si son o todo más oscuro o más brillante que el fondo, las transformaciones top-hat se pueden utilizar para mitigar los gradientes de iluminación. En efecto, un top-hat con un gran elemento estructurante isotrópico actúa como un filtro pasa altas. A medida que el gradiente de iluminación se encuentra dentro de las bajas frecuencias de la imagen, se elimina por la transformación top-hat.

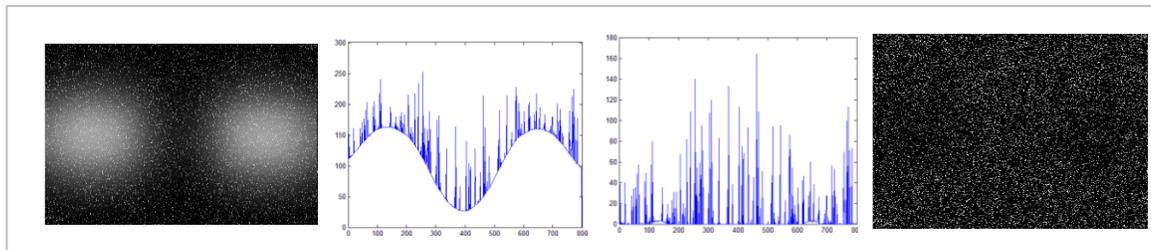


Figura 3-1. Ecuación por tophat.

En contraste a la transformación top-hat, una transformación bottom-hat se define como la cerradura de la imagen menos la imagen original. Tanto la transformación top-hat como bottom-hat se pueden utilizar juntos para mejorar el contraste de la imagen.

Restando la imagen original de la cerradura a la imagen se denomina transformación bottom-hat, la cual es denotada por:

$$BH_B(A) = (A \bullet B) - A$$

La Morfología Matemática es una técnica aplicada de procesamiento de imágenes y visión por computadora basada en la topología, la teoría de grafos, el álgebra y la geometría integral. El objetivo de las transformaciones morfológicas es la extracción de estructuras mediante la utilización de una estructura morfológica, conocida como elemento estructural (EE). Las propiedades del EE se seleccionan previamente, de acuerdo a las formas que se desean obtener (Serra, 1986). Los filtros morfológicos básicos son la apertura $\gamma_{\mu B}$ y la cerradura $\varphi_{\mu B}$, con un elemento EE dado μB . La apertura y cerradura están dadas por:

$$\gamma_{\mu B}(x) = \delta_{\mu B}(\varepsilon_{\mu B}(f))$$

$$\varphi_{\mu B}(x) = \varepsilon_{\mu B}(\delta_{\mu B}(f))$$

Donde estas operaciones están expresadas a partir de la erosión $\varepsilon_{\mu B}(f)(x) = \text{Min}\{f(y); y \in \mu B_x\}$ y la dilatación $\delta_{\mu B}(f)(x) = \text{Max}\{f(y); y \in \mu B_x\}$ donde *Min* es el operador de valor mínimo y *Max* es el operador valor máximo.

Otra clase de filtros está formada por la apertura y la cerradura por reconstrucción (Soille, 2003). Estos filtros morfológicos se construyen

utilizando las transformaciones conocidas como dilatación y erosión geodésicas. Esas transformaciones están dadas por:

$$\delta_f^1(g)(x) = \text{Min}\{f(x), \delta_\varepsilon(g)(x)\} \text{ con } f \geq g, \text{ para la dilatación geodésica y}$$

$\varepsilon_f^1(g)(x) = \text{Max}\{f(x), \varepsilon_\delta(g)(x)\} \text{ con } f \leq g, \text{ para la erosión. A partir de estas transformaciones geodésicas básicas se construyen las transformaciones por reconstrucción, iterando dichas transformaciones hasta alcanzar un punto de estabilidad (Vincent, 1993).}$

$$R(f, g) = \lim_{n \rightarrow \infty} \delta_f^n(g) = \underbrace{\delta_f^1 \delta_f^1 \dots \delta_f^1(g)}_{\text{Hasta estabilidad}}$$

$$R^*(f, g) = \lim_{n \rightarrow \infty} \varepsilon_f^n(g) = \underbrace{\varepsilon_f^1 \varepsilon_f^1 \dots \varepsilon_f^1(g)}_{\text{Hasta estabilidad}}$$

En particular, cuando la función g está dada por la erosión o la dilatación, se obtienen la apertura y la cerradura por reconstrucción.

$$\tilde{\gamma}_\lambda(f) = R(f, \varepsilon_\lambda(f))$$

$$\tilde{\phi}_\lambda(f) = R^*(f, \delta_\lambda(f))$$

3.2.2. Reconstrucción morfológica

La reconstrucción morfológica puede ser descrita conceptualmente como repetidas dilataciones de una imagen, llamada imagen del marcador, hasta que el contorno de la imagen del marcador quede por debajo de una segunda imagen, llama la imagen de máscara. En la reconstrucción morfológica, los picos en la imagen del marcador se "extienden", o se dilatan.

Esta figura ilustra este proceso en 1-D. Cada dilatación sucesiva se ve limitada a estar por debajo de la máscara. Cuando la posterior dilatación deja

de cambiar la imagen, el proceso se detiene. La dilatación final es la imagen reconstruida. La figura muestra las dilataciones sucesivas del marcador.

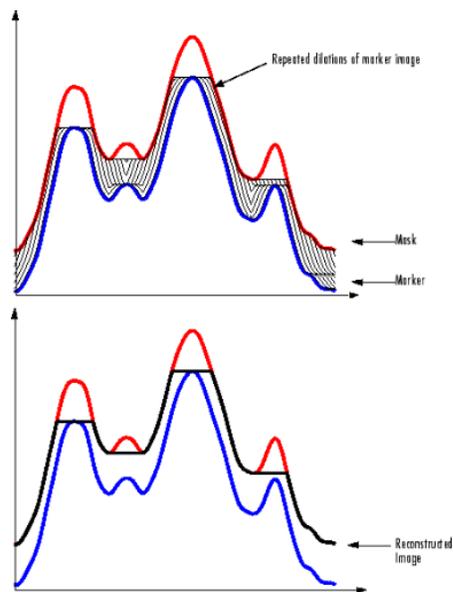


Figura 3-2. Reconstrucción morfológica.

La reconstrucción morfológica se basa en la dilatación morfológica, teniendo en cuenta las siguientes características:

- El proceso se basa en dos imágenes, un marcador y una máscara, en lugar de una imagen y un elemento estructurante.
- El procesamiento se basa en el concepto de la conectividad, en lugar de un elemento de estructuración.
- Se repite hasta que la estabilidad de proceso, es decir, la imagen ya no cambia.

La reconstrucción morfológica procesa una imagen, llamada el marcador, basándose en las características de otra imagen, llamada la máscara. Los puntos altos o picos, en la imagen del marcador especifican dónde comienza

el procesamiento. El proceso continúa hasta que los valores de la imagen dejan de cambiar.

3.2 ESPACIOS DE BINARIOS

3.2.3. Fundamentos Matemáticos

Espacio $\{0, 1\}^n$

Los principales objetos de estudio son vectores de dimensión n con componentes binarios. Dependiendo del contexto, los vectores son llamados puntos, patrones, direcciones, palabras, elementos de memoria, datos o eventos (Kanerva, 2003).

Sea n el número de dimensiones del espacio. El número de puntos, o posibles elementos de memoria, es entonces 2^n . Este número se denota por N .

Puntos de N son naturalmente representados por n -tuplas de ceros y unos. Por Ej. $\langle 0,1,0,1,1 \rangle$ y puede ser escrita como n -bits enteros en una representación binaria (01011). Pensando en ellos como enteros, es engañoso, ya que los números enteros están clasificados, como un conjunto unidimensional, mientras que en el presente contexto no hay razón para decir que 00001, por ejemplo, viene o no antes que 00100. Lo útil de las propiedades del espacio $\{0,1\}^n$ resulta de su largo número de dimensiones.

Conceptos relacionados con el Espacio $\{0, 1\}^n$

Origen, 0. El punto con todas sus coordenadas 0 es llamado el origen, $0 = 000\dots 00$.

Complemento, \bar{x} . El complemento, u opuesto, de un punto x son las n -tuplas que tienen unos hacerlas ceros y viceversa.

Norma, $|x|$. La norma de un punto x es el número de unos en su representación binaria.

Diferencia, $x - y$. La diferencia de dos puntos x y y es la n -tupla que tiene unos donde x y y son diferentes y ceros en cualquier otro lugar. Esto es el 'or exclusivo' bit a bit de x y y : $x - y = x \oplus y$. Esta diferencia tiene propiedad conmutativa $x - y = y - x$.

Distancia, $d(x, y)$. La distancia entre dos puntos x y y es el número de dimensiones en el cual x y y difieren. Es llamado *distancia de Hamming* (su raíz cuadrada es la *distancia Euclidiana*) y es expresada en bits. La distancia es la norma de la diferencia:

$$d(x, y) = |x - y|$$

$$= |x_0 - y_0| + \dots + |x_{n-1} - y_{n-1}|$$

Dos puntos de N que son cercanos uno al otro, se dice que son similares; ese término se utiliza sobre todo en un sentido relativo, como en “ x es el elemento de X más similar a y ”.

Ejemplo:

Punto inicial 100101 = 37		
Punto	Distancia de Hamming	Distancia Euclidiana
110101 = 53	1	16
100111 = 39	1	2
101101 = 45	1	8
100001 = 33	1	4

Tabla 3-1. Ejemplo de diferencia entre distancia de Hamming y distancia Euclidiana

Intermediario, $x: y: z$. El punto y esta entre los puntos x y z sí y solo sí la distancia desde x a z es la suma de las distancias desde x a y y desde y a z ; esto es:

$$d(x, z) = d(x, y) + d(y, z)$$

Es fácil ver que cada bit de un punto en el medio es una copia del bit correspondiente de un punto final:

$$x: y: z \text{ iff } y_i = x_i \text{ or } y_i = z_i \text{ for } i = 0, \dots, n - 1$$

La relación $w: x: y: z: \dots$ esta definida recursivamente como $w: x: y$ & $x: y: z: \dots$. La intermediación es simétrica ($x: y: z: \dots$ iff $\dots: z: y: x$) pero no transitiva (en el sentido que $w: x: y: z$ no implica $w: x: z$). El conjunto Y esta entre los puntos x y z ($x: Y > z$) sí y solo sí todos sus puntos están entre x y z . Todo N esta entre un punto y su complemento: $x: N: 'x$.

Ortogonalidad, $x \perp y$. El punto x es ortogonal al punto y , o los dos son perpendiculares o indiferentes, sí y solo sí la distancia entre los dos es la mitad del número de dimensiones:

$$d(x, y) = n/2$$

La distancia $n/2$ es llamada la distancia indiferente de $\{0,1\}^n$. Si x es ortogonal a y , también es ortogonal a su complemento $'y$ (x esta a la mitad de y y $'y$): $x \perp y$ implica $x \perp 'y$.

Círculo, $O(r, x)$. Un círculo con radio r y centro x es el conjunto de puntos que se encuentran en la mayoría de los r bits de x :

$$O(r, x) = \{y | d(x, y) \leq r\}$$

Cualquier círculo de radio r es el espacio entero N :

$$O(r, x) = N$$

Ejemplos. Un pequeño número de dimensiones basta para demostrar estos conceptos. Sea $n = 5$, $x = 00111$, y $y = 01010$. Entonces

Complemento: $'x = 11000, y 'y = 10101$

Norma: $|x| = 3, y |y| = 2$

Diferencia: $x - y = y - x = x \oplus y = 01101 (= 'x - 'y)$

Distancia: $d(x, y) = |x - y| = |01101| = 3$

Intermediación: Cualquier $z = 0 ** 1 *$, donde cada $*$ puede ser 0 o 1, está entre x y y . Por ejemplo, $x: 01110: y$

Ortogonalidad: Ningún punto puede ser (exactamente) ortogonal a x , porque n es impar. $0101 \perp 0011$

Círculo: $O(1, x) = \{00111, 10111, 01111, 00011, 00101, 00110\}$

El espacio N puede ser representado por los vértices del cubo unitario en un espacio Euclidiano de n dimensiones. La distancia entre dos puntos de N es entonces la longitud de la ruta más corta a lo largo de los bordes del cubo que une los dos puntos, también es llamado distancia city-block. Cuando dos puntos están más de un bit uno del otro, hay múltiples rutas más cortas entre ellos.

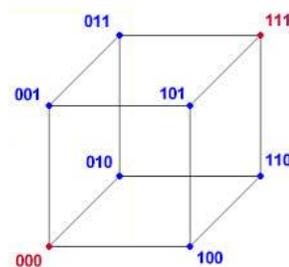


Figura 3-3. Cubo distancia Hamming.

La analogía con la esfera. Los vértices de un cubo unitario de n dimensiones parecen la superficie de una esfera de n dimensiones con (métrica Euclidiana) radio $\sqrt{n/2}$. Esto da lugar a la analogía de la esfera. Se llamará un espacio esférico si (1) cualquier punto x tiene un único opuesto $'x$, (2) el

espacio entero está entre cualquier punto x y su opuesto \bar{x} , y (3) todos los puntos son “iguales” (significa que para cualquier par de puntos x y y existe una distancia de preservación auto mórfica del espacio que mapea x a y , de manera que desde cualquiera de sus puntos el espacio “parece” el mismo). La superficie de una esfera (en Euclidiano de 3 espacios) claramente es esférica. De acuerdo a la definición, N es también esférico, desde $y \oplus x \oplus (\dots)$ es un automorfismo que mapea x a y (de hecho, para cada una de las $n!$ permutaciones de las coordenadas, hay un automorfismo que mapea x a y).

Ya que N es esférica, ayuda en pensar en ello como una superficie de esfera con una circunferencia de $2n$. Todos los puntos de N están clasificados igualmente como puntos de origen, y un punto y su complementado son como dos polos de distancia n uno del otro, con todo el espacio en el medio. Los puntos a mitad de camino entre los polos y perpendiculares a ellos son como el ecuador. Fiel a la analogía de esfera, esto se refiere a un conjunto de puntos a una distancia dada de un punto como un círculo en N .

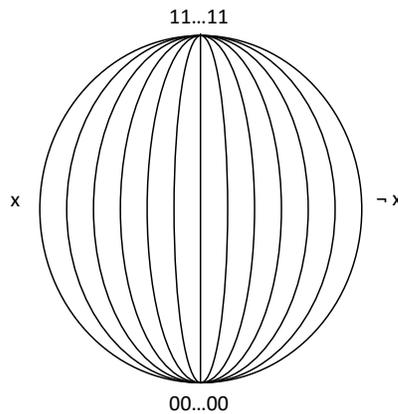


Figura 3-4. Espacio Hamming tridimensional

Algunas propiedades del espacio $\{0, 1\}^n$

Debido a que N es esférica en el sentido definido anteriormente, puede ser "girada" como para mover cualquiera de sus puntos al origen. Por lo tanto, al hablar de un punto arbitrario x , a menudo es conveniente pensar en x como el origen, $x = 0$, porque la distancia desde 0 a algún otro punto y es simplemente el número de unos en y :

$$d(0, y) = |y - 0| = |y|$$

Distribución del espacio N El número de puntos que son exactamente d bits desde un punto arbitrario x (dicho, desde el punto 0) es el número de caminos a escoger d coordenadas de un total de n coordenadas, y es por lo tanto dado por el coeficiente binomial

$$(n : d) = \binom{n}{d} = \frac{n!}{d!(n-d)!}$$

La distribución de N por lo tanto es la distribución binomial con parámetros n y p , donde $p = \frac{1}{2}$. La media del binomial es $n/2$, y la varianza es $n/4$. Esta función de distribución es denotada por $N(d)$. La distribución normal F con media $n/2$ y desviación estándar $\sqrt{n/4}$ es una buena aproximación a ella:

$$\begin{aligned} N(d) &= Pr\{d(x, y) \leq d\} \\ &\cong F\{(d - n/2)/\sqrt{n/4}\} \end{aligned}$$

Tendencia a la Ortogonalidad: Una propiedad excepcional de N es que la mayor parte se encuentra aproximadamente a una distancia promedio de $n/2$ desde un punto (y su complemento), con sólo una fracción de minuto más cerca o más lejos. En otras palabras, la mayor parte del espacio es casi ortogonal a cualquier punto dado, y el más grande es n , más pronunciado es este efecto. En cuanto a una analogía con una esfera, con cualquier punto y su complemento tomándolos como polos, casi todo el espacio se encuentra

sobre o cerca del ecuador (al igual que los puntos de una esfera euclidiana n-dimensional).

La matemática en la que funciona es de la siguiente manera: Si dividimos la distancia media $n/2$ por la desviación estándar de la distancia $\sqrt{n/2}$, se obtiene que la distancia de un punto a la mayor parte del espacio (a partir de un polo al ecuador) es \sqrt{n} desviaciones estándar. Para $n = 1,000$ es 31,6 desviaciones estándar. De acuerdo con la distribución normal, sobre 0.999999 del espacio se encuentra dentro de 5 desviaciones estándar de la media o dentro de $(\sqrt{n} \pm 5)$ desviaciones estándar desde un punto de N . Con $n = 1,000$, la distancia media es 500 bits, pero sólo alrededor de un millonésima del espacio está más cerca del punto de 422 bits de o más lejos de él que 578 bits (5 desviaciones estándar vienen a 78 bits). Algunos datos básicos sobre la distribución de N se resumen en la Tabla 3-2 para $n = 100$, 1,000 y 10,000 y también aparecen en la Figura 3-5.

Porción del espacio N en el rango R sobre la distancia media.	Número de dimensiones, n					
	n=100		n=1000		n=10000	
	Rango R Bits	% de n	Rango R Bits	% de n	Rango R Bits	% de n
0.5	50 ± 3	6	500 ± 11	2	5,000 ± 34	0.6
0.98	50 ± 12	24	500 ± 37	7	5,000 ± 116	2.3
0.998	50 ± 15	30	500 ± 49	10	5,000 ± 154	3.1
0.9998	50 ± 19	38	500 ± 59	12	5,000 ± 186	3.7
0.999998	50 ± 24	48	500 ± 75	15	5,000 ± 238	4.8
0.999999998	50 ± 30	60	500 ± 95	19	5,000 ± 300	6.0
Distancia media.		50		500		5,000
Desviación estándar de la distancia		5		15.8		50

Tabla 3-2. Distribución del espacio binario

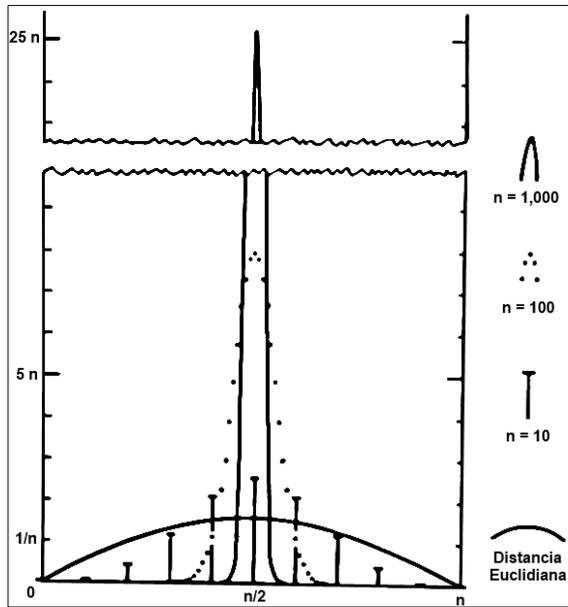


Figura 3-5. Distribución de distancia Hamming en espacio binario y de distancia Euclidiana

4 PROPUESTA

La cantidad de información en la imagen se denota por el conjunto de intensidades de los píxeles y sus variaciones; es decir, las áreas de la imagen con regiones planas tienen menos información que imágenes texturizadas. La medición de la cantidad de información en una imagen representa una tarea difícil.

4.1. CARACTERÍSTICAS DE LAS IMÁGENES

Sea w el vecindario de una posición de un píxel x de una imagen $I(x)$. La función de densidad de probabilidad (fdp) de las intensidades locales del vecindario w describen las variaciones locales de intensidad del píxel. Se denota la distribución de densidad de probabilidad de la vecindad por $w \sim f(w)$. La naturaleza de $f(w)$ se vuelve difícil de caracterizar, porque no hay manera de conocer previamente la forma de la fdp, pero podría ser aproximada mediante un enfoque discreto basado en su histograma. El enfoque del histograma debe tener suficientes píxeles de información para estimar la fdp de acuerdo a la ley de los grandes números (Goñi, 2011). Entonces, para describir el comportamiento local de w , se debe extraer la información más significativa en $f(w)$. Para caracterizar $f(w)$, utilizamos la información proporcionada por sus derivadas, que se centra sobre los cambios locales y los resultados son fáciles de calcular, haciendo factibles de utilizar la información como descriptor. La información más significativa usada de la derivada se toma de los puntos críticos de $f^{(n)}(w) = 0$, que corresponden a derivada de orden n ; es decir, para describir un vecindario w definir el rango de orden de derivadas que se usa para describir la región. El uso de derivadas de orden superior proporciona un buen criterio para medir y caracterizar el comportamiento de $f(w)$, siendo las características más útiles las que no son dependientes de magnitud, sólo de variación. Como ejemplo tenemos el caso de la **primera derivada y la función original**. La

primera derivada se convierte a invariante a la posición inicial y sólo caracteriza el comportamiento local, en lugar de la función original que es totalmente dependiente de los valores iniciales. En resumen, los descriptores alrededor de un vecindario w para una posición particular x de la imagen se denotan por

$$\Phi = \{\phi_1, \phi_2, \dots, \phi_n\}$$

tal que $f^j(\phi_i) = 0$ para la derivada j de la fdp sobre w donde $j \in [k, m]$, donde $f(x)$ es la fdp sobre el vecindario w , k y m son el orden de las derivadas en el intervalo usado. El conjunto de descriptores sobre w mantiene suficiente información sobre la fdp que podría usar para caracterizar.

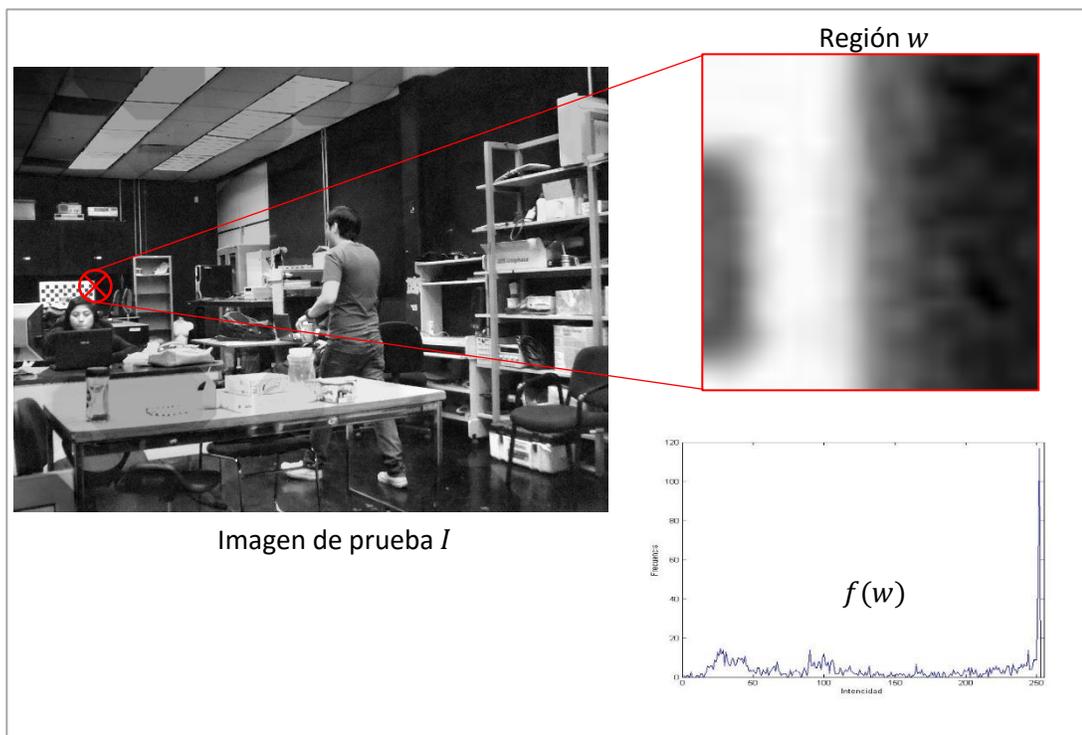


Figura 4-1. Histograma sobre un vecindario w

En la imagen se observa una región w de la imagen I en la cual se obtiene su histograma, el cual nos servirá para la codificación de su característica.

Sin embargo, observamos que la deformación geométrica causada por la rotación diferente al plano podría afectar a $f^j(x)$, especialmente cuando se convierten en una derivada de orden superior, en cambio, las primeras derivadas son menos afectadas con pequeñas rotación al plano. Finalmente, el conjunto de características de los descriptores sobre una región w , deben ser representadas como único valor expresado de la siguiente manera

$$\Phi = S^k$$

donde el conjunto de descriptores se codifican como un elemento en el espacio S^k a través de la función $c(\Phi)$. La existencia de $c(\Phi)$ no implica la existencia de $c^{-1}(\Phi)$.

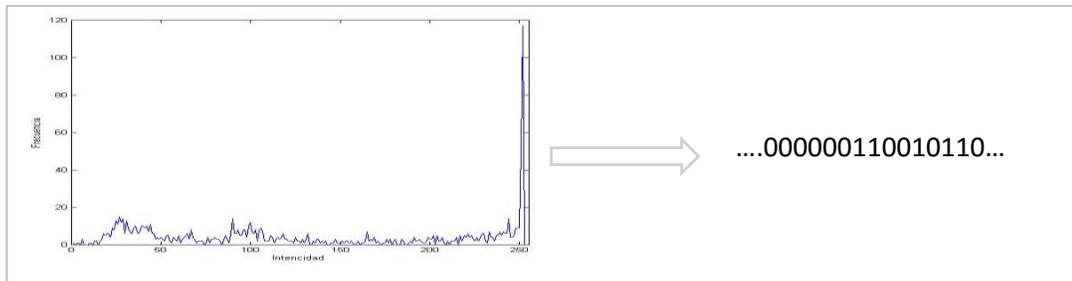


Figura 4-2. Representación de $f(w)$ en forma de cadena binaria.

La imagen $I(x)$ con elementos discretos, proporciona información para ofrecer una aproximación discreta a la fpd del vecindario. Para tratar con esta característica de la imagen, proporcionamos una versión discreta.

De hecho, la gama de las intensidades de los píxeles es discreto, el histograma normalizado proporciona una aproximación de la fpd $f(x)$ con k elementos de ancho, donde cada elemento corresponde a una valor discreto de intensidad de los píxeles, y el ε error asociado se define como $\varepsilon = x_{j+1} - x_j$ para dos elementos consecutivos x_j y x_{j+1} ; como ejemplo, imágenes en escala de grises con 8 bits de resolución, tenemos 256 elementos. Pero, la naturaleza discreta de $f(x)$ hacen que las derivadas $f^j(x)$ debe ser estimado considerando la naturaleza discreta de la aproximación de la fpd.

5 MODELO EXPERIMENTAL Y RESULTADOS

En esta sección mostramos un modelo de validación experimental de la propuesta. La validación experimental se efectúa en dos sentidos, en imágenes reales y en imágenes artificiales. Las imágenes artificiales se utilizan para mostrar la eficiencia del algoritmo en condiciones controladas y las imágenes reales para verificar el impacto que se tienen cuando las afectaciones de la imagen son difíciles de controlar.

5.1 MODELO EXPERIMENTAL

Para desarrollar el modelo experimental se utiliza la siguiente infraestructura

- Cámara Canon con 12.2 Megapíxeles de resolución, lente con estabilizador óptico de imagen EF-S 18-55 mm f/3.5 -5.6 IS.
- Computadora Lenovo Procesador Core i5 a 2.3 GHz con 8 GB de memoria RAM.
- Matlab v2014a.

Del cual se desprenden dos experimentos: En ambientes controlado con objetos artificiales; y en ambientes exteriores

En el primer experimento, se definen objetos con textura de forma artificial y se efectúan transformaciones básicas de rotación escalada midiendo la distancia de hamming. En el segundo experimento, se definen en imágenes en exteriores pruebas de perspectiva y selección de características en escenarios externos.

5.2 DISCUSIÓN DE RESULTADOS Y ANÁLISIS DE DATOS.

En esta sección evaluaremos las distancias Hamming desde una característica original hacia una característica de la misma imagen que ha sido rotada en 30° hasta 180° .

Evaluamos la eficiencia del detector ante rotación de imágenes, para lo cual utilizamos una imagen I de 16×16 píxeles a la cual se le calcula su función de distribución discreta con 256 elementos que representan las variaciones de escala de grises. Se mide la distancia de Hamming entre la imagen original y la imagen resultante con rotación.

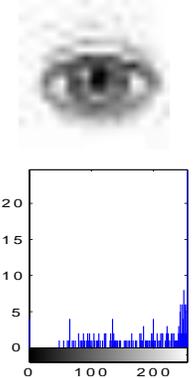
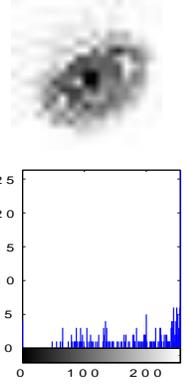
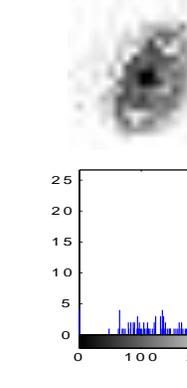
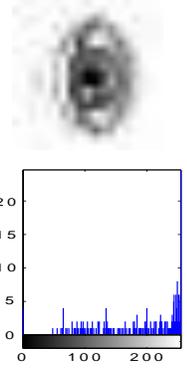
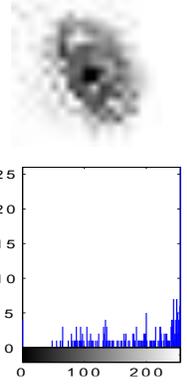
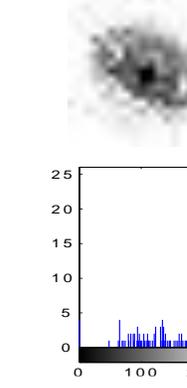
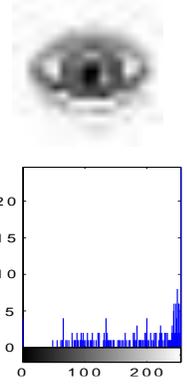
<p>Distancia con respecto a la original. 0 Rotación 0°</p> 	Distancia con respecto a la original. 16	Distancia con respecto a la original. 21	Distancia con respecto a la original. 0
	Rotación 30°	Rotación 60°	Rotación 90°
			
	Distancia con respecto a la original. 17	Distancia con respecto a la original. 20	Distancia con respecto a la original. 0
	Rotación 120°	Rotación 150°	Rotación 180°
			

Tabla 5-1. Pruebas de imágenes con rotación

Para evaluar su comportamiento ante escalado de imágenes (zoom), la imagen I (de tamaño de 16×16), es escalada desde un 25% hasta 100%, y posteriormente se obtiene de nuevo su característica binaria, de la cual se calcula su distancia de Hamming con respecto a la original.

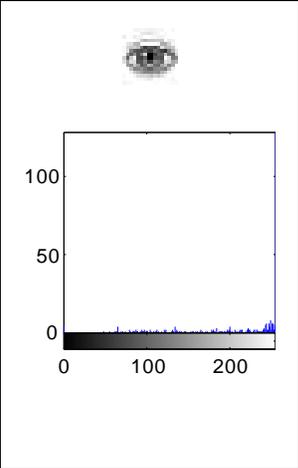
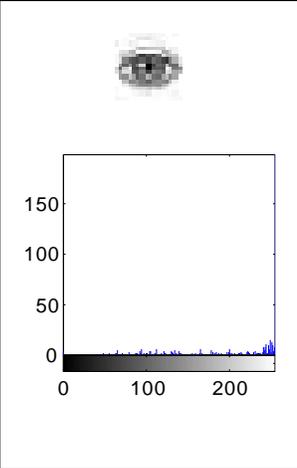
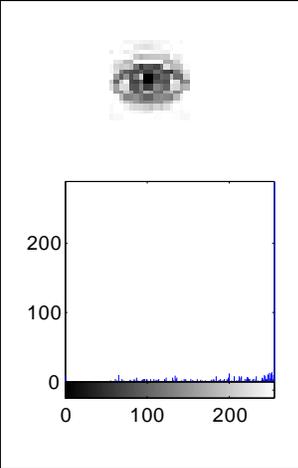
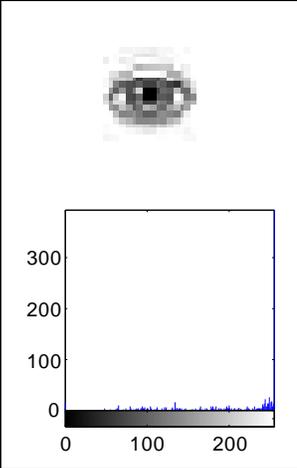
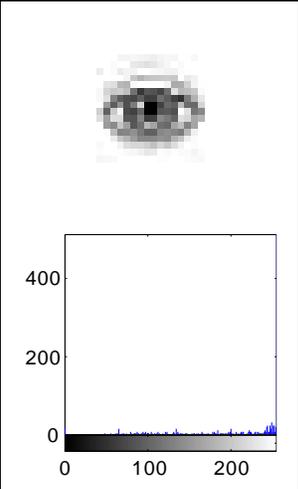
Distancia con respecto a la original. 0	Distancia con respecto a la original. 20	Distancia con respecto a la original. 20	Distancia con respecto a la original. 18
Escala 1	Escala 1.25	Escala 1.5	Escala 1.75
			
Distancia con respecto a la original. 0			
Escala 2			
			

Tabla 5-2. Pruebas de imágenes con escalado

Siguiendo con este tipo de pruebas, son adquiridas varias imágenes desde diferente perspectiva. Para cada imagen se selecciona el mismo punto para obtener su correspondiente característica binaria y posteriormente se calcula la distancia entre la característica original.

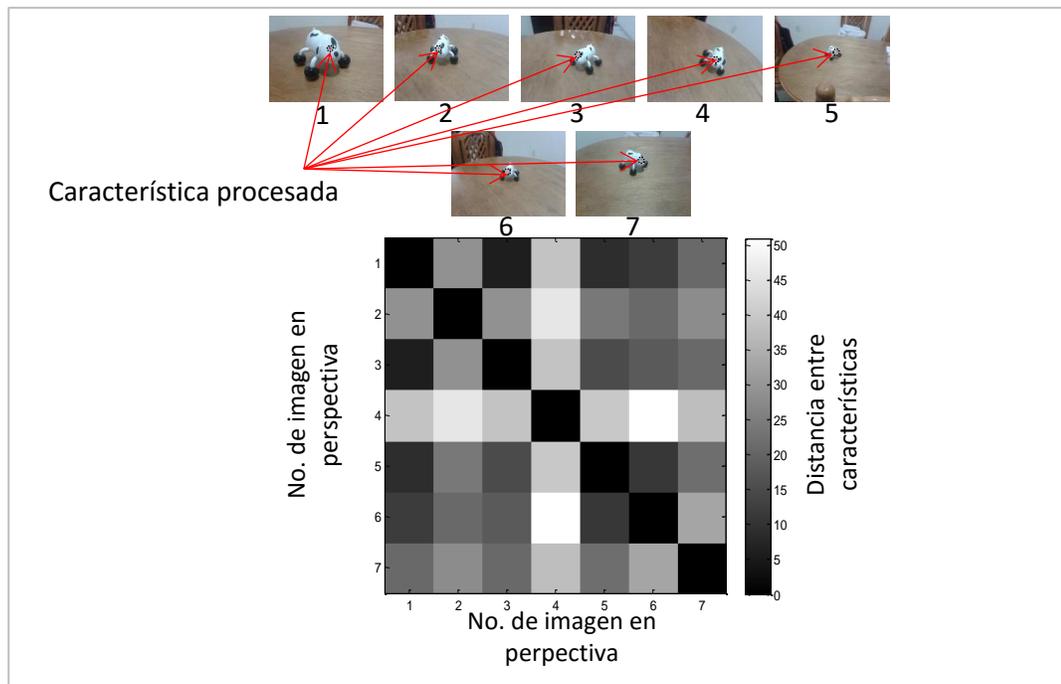


Figura 5-1. Representación de distancia de una característica tomada desde diferente perspectiva.

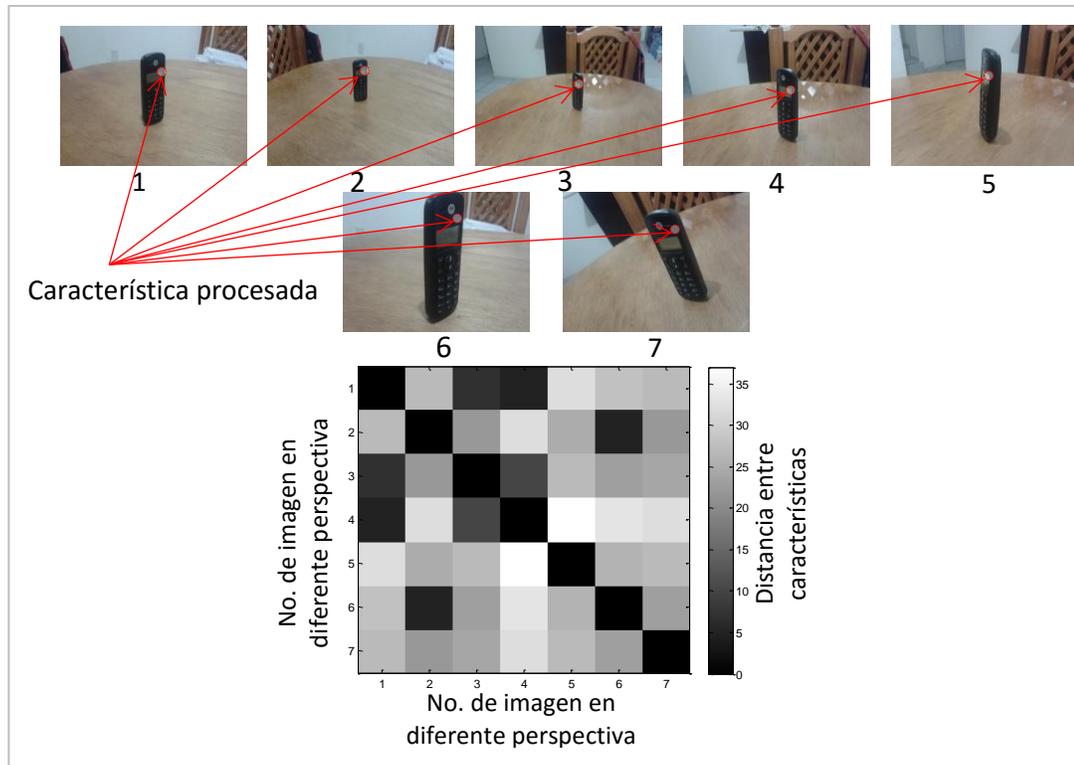


Figura 5-2. Distancia de una característica tomada desde diferente perspectiva.

Se muestra en la imagen a continuación (Figura 5-3), la topología del espacio métrico de características binarias en comparación con la topología de la matriz de tensores de Harris (Harris & Stephens, 1988). Se observa que los dos espacios son diferentes porque en donde la matriz de tensores es máxima solo corresponde a zonas que pueden representar esquinas mientras que el espacio métrico binario tiene regiones máximas en lugares donde la textura es distintiva de forma que aparte de incluir las esquinas contiene los bordes y líneas bien definidas en la imagen. De esta comparativa se concluye que ambos descriptores son diferentes. El espacio métrico binario resulta útil para encontrar zonas de alta textura (que pueden llegar a ser referencias entre imágenes), lo que le da la versatilidad para usarse como

descriptores globales. Mientras que en la matriz de tensores solo se focaliza en lugares que corresponden a esquinas (líneas que convergen en forma ortogonal), sin importar su orientación.

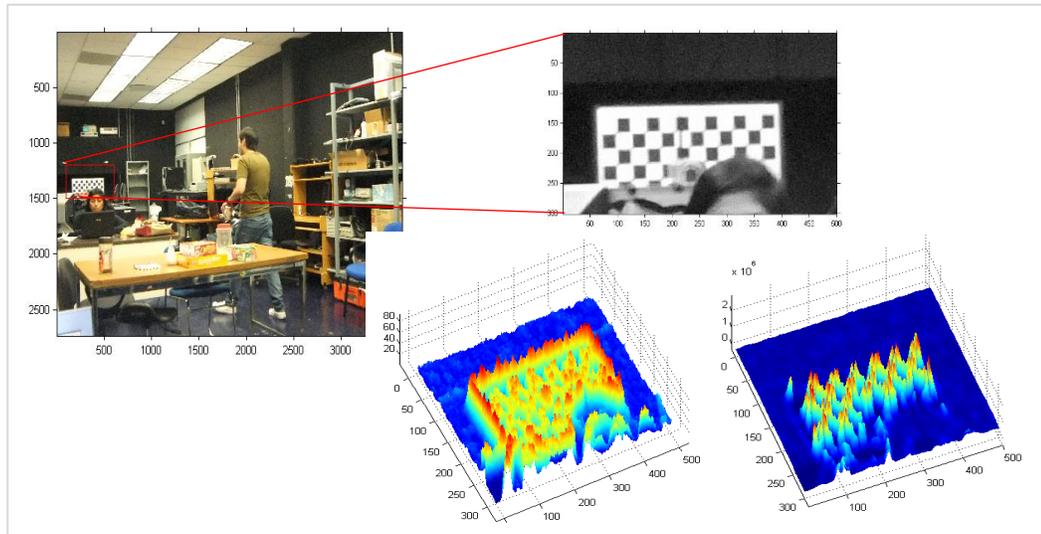


Figura 5-3. Región de una imagen obteniendo superficie de distancia de características binarias y tensores de Harris

6 CONCLUSIONES

- El espacio binario, bajo las propiedades descritas hacen que se comporte en altas dimensionalidades como una esfera, permite que sea ubicada más fácilmente dos características similares, ya sea que provengan de la misma imagen o de imágenes diferentes incluso con diferente perspectiva.
- La generación de descriptores con características binarias utilizando las variaciones de intensidad y la textura, es muy confiable siempre y cuando exista una etapa de ecualización que permita eliminar los efectos de luz sobre la imagen.
- La selección del mejor canal en diferentes espacios de color toma gran importancia cuando se tomará como información para el descriptor las variaciones de intensidad.
- En este contexto la idea de utilizar cadenas binarias que permitan la localización de regiones en imágenes toma gran relevancia.
- Una vez que ha sido identificada una imagen o ciertas regiones de ella, mediante características simbólicas (binarias), es posible identificar aquellas que sean similares de acuerdo a una métrica de similitud dependiente de la distancia de Hamming.
- Los motores más populares y poderosos (como Google) están basados en texto, pero la investigación sobre el desarrollo de los motores de búsqueda se ocupan de la información visual está recibiendo una atención cada vez más (T.H. Bui, 2005).

7 BIBLIOGRAFÍA.

- Aase, S., Husoy, J., & Waldemar, P. (1999). A critique of SVD-based image coding systems. *Circuits and Systems, Proceedings of the 1999 IEEE International Symposium on*. IEEE.
- Bay, H., Ess, A., Tuytelaars, T., & Gool, L. V. (2008). SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding*, 346--359.
- Bhattacharayya, S. E. (2011). Survey of color image preprocessing and segmentation techniques. *Journal of Patter Recognition Research*, 120-129.
- Brown, M. &. (2002). Invariant Features from Interest Point Groups. *Proceedings of the 13th British Machine Vision Conference* (págs. 253–262). Cardiff.
- Chávez, J. L. (2012). *Tratamiento Digital de Imágenes Multiespectrales*. Lulu.com.
- Cipolla, R., Fletcher, G., & Giblin, P. (1995). Surface geometry from cusps of apparent contours. *Proceedings., Fifth International Conference on Computer Vision* (págs. 858 - 863). IEEE.
- Criminisi, A., Reid, I., & Zisserman, A. (2000). Single view metrology. *International Journal of Computer Vision*, 123-148.
- D'O., C. S. (1996). *Colour, colour spaces and the human visual systems*. England: University of Birmingham.
- Fischler, M. A. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 381-395.
- Fleck, M. M., Forsyth, D. A., & Bregler, C. (1996). Finding naked people. *ECCV in Computer Vision* (págs. 593-602). Springer Berlin Heidelberg.
- Gatter, M. (2005). *Getting it Right in Print: Digital Pre-press for Graphic Designers*. Laurence King Publishing.
- Gonzalez, R. C. (2007). *Digital Image Processing (3rd Edition)*. Prentice Hall.

- Goñi, J. M. (2011). *Matemáticas. Complementos de formación disciplinar*. Grao.
- Goshtasby, A. (2012). Similarity and Dissimilarity Measures. En *Advances in Computer Vision and Pattern Recognition* (págs. 7-66). Springer-Verlag.
- Han, F., & Zhu., S.-C. (2003). Bayesian Reconstruction of 3D Shapes and Scenes From A Single Image. *Proceedings of the First IEEE International Workshop on Higher-Level Knowledge in 3D Modeling and Motion Analysis* (pág. 12). IEEE Computer Society.
- Harris, C., & Stephens, M. (1988). A Combined Corner and Edge Detector. *Proceedings of the Fourth Alvey Vision Conference*, (págs. 147-151).
- Horn, B. P. (1986). *Robot Vision*. McGraw-Hill.
- Horn., I. (1975). Obtaining Shape from Shading Information. En *The Psychology of Computer Vision* (págs. 115-155). McGraw-Hill.
- Hugo Jiménez, J. S. (2008). Dynamic Selection of Characteristics for Feature Based Image Sequence Stabilization. En *Advanced Concepts for Intelligent Vision Systems* (págs. 290-297). Springer Berlin Heidelberg.
- Kanerva, P. (2003). *Sparse Distributed Memory*. Bradford Books.
- Leutenegger, S., Chli, M., & Siegwart, R. (2011). BRISK: Binary Robust invariant scalable keypoints. *International Conference on Computer Vision* (págs. 2548 - 2555). IEEE.
- Levkowitz, H., & Herman, G. (1993). GLHS: A Generalized Lightness, Hue, and Saturation Color Model. *Graphical Models and Image Processing*, 271–285.
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. *International Conference on Computer Vision*, (págs. 1150-1157).
- Lowe, D. G. (2004). Distinctive Image Features from Scale Invariant Keypoints. *International Journal of Computer Vision*, 91-110.
- Lowe, M. B. (2007). Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*,, 59-73.

- Matthew Brown, D. G. (2007). Automatic Panoramic Image Stitching using Invariant Features. *International Journal of Computer Vision* , 59-73.
- Mendenhall, W. (2008). *Introducción A la Probabilidad Y Estadística*. Cengage Learning Latin America.
- Meyer, Y. (1993). Wavelets-algorithms and applications. *Wavelets-Algorithms and applications Society for Industrial and Applied Mathematics Translation.*, 142.
- Ohta, Y.-I., Kanade, T., & Sakai, T. (1980). Color information for region segmentation. *Computer Graphics and Image Processing*, 222–241.
- Palus, H. (1998). Representations of colour images in different colour spaces. En *The Colour Image Processing Handbook* (págs. 67-90). Springer US.
- Poynton, C. (1995). A guided tour of color space. *Advanced Television Electronic Imaging Conference*, (págs. 167-180). San Francisco.
- Ramanath, R., S., W. E., Yoo, Y., & Drew, M. S. (2005). Color image processing pipeline: A general survey of digital still camera processing. *IEEE Signal Processing Magazine*, 34–45.
- Raoui, Y., Bouyakhf, E. H., Devy, M., & Regragui, F. (2011). Global and Local Image Descriptors for Content Based Image Retrieval and Object Recognition. En *Applied Mathematical Sciences* (págs. 2109-2136.).
- Sameer Agarwal, Y. F. (2010). Reconstructing Rome. *IEEE Computer*, 40-47.
- Serra, J. (1986). Introduction to mathematical morphology. En *Computer vision, graphics, and image processing* (págs. 283-305).
- Shi, J. T. (1994). Good Features to Track. *Computer Vision and Pattern Recognition* (págs. 593 - 600). IEEE.
- Siebert, B. C. (2009). *An introduccion to 3D computer vision techniques and algorithms*. John Wiley & Sons, Ltd.

- Sivic, J., & Zisserman, A. (2004). Video data mining using configurations of viewpoint invariant regions. *Computer Vision and Pattern Recognition* (págs. 488–495). IEEE.
- Skarbek, W., & Andreas Koschan. (1994). Colour image segmentation: A survey. *tech. rep., Polish Academy of Sciences.*
- Snavely, K. N. (2008). *Scene Reconstruction and Visualization from Internet Photo Collections*. ProQuest.
- Stephen Se, D. G. (2002). Global Localization using Distinctive Visual Features. *Intl. Conference on Intelligent Robots and Systems*. IEEE/RSJ.
- Susstrunk, S., Buckley, R., & Swen, S. (1999). Standard RGB color spaces. *Color Imaging Conference* (págs. 127-134). Laboratory of Audio-Visual Communication.
- T.H. Bui, R. L. (2005). Color and Texture Induced Structures of Keyword Space. *Proc. of 10th Congress of the International Colour Association*, (págs. 1059-1062).
- Thormählen, T., Hasler, N., Wand, M., & Seidel, H. (2008). Merging of feature tracks for camera motion estimation from video. *In. 5th European Conference on Visual Media Production*. CVMP.
- Torr, P. Z. (1998). Robust computation and parametrization of multiple view relations. *Sixth International Conference on Computer Vision* (págs. 727 - 732). IEEE.
- Tuytelaars, T., & Van Gool, L. (2000). Wide baseline stereo matching based on local, affinely invariant regions. *In. Proc. British Machine Vision Conference*. (págs. 412–425). BMVC.
- Vartak, A. P., & Mankar, V. (2013). Colour image segmentation- a survey. *International Journal of Emergin Technology and Advanced Engineering*, 681–688.

- Witkin, A. P. (2003). Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17–45.
- Wyszecki, G., & Stiles, W. S. (2000). *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley & Sons, Inc.
- Yan Cui, N. H.-P. (2009). Scale Invariant Feature Transform with Irregular Orientation Histogram Binning. En *Image Analysis and Recognition* (págs. 258-267). Springer Berlin Heidelberg.