

Capítulo 2. Recuperación de información visual

Existen varias maneras para realizar la recuperación de imágenes. Una de ellas es basándose en descripciones de texto, que consiste en indexar las imágenes partiendo de anotaciones textuales y ejecutando consultas basadas en texto; otra es basándose en consultas ejemplos de imágenes, recuperando las imágenes que mejor concuerdan con la imagen a consultar. La similitud entre imágenes se calcula basándose en color, textura o forma. En este tipo de recuperación el usuario alimenta al sistema con una imagen que representa lo que quiere recuperar de la base de datos.

Durante el proceso de entrada, las imágenes son procesadas para extraer las características seleccionadas y así representar el contenido de la imagen. Este proceso se conoce como *indexación* y asigna a cada imagen un conjunto de descriptores de identificación ó índices. Estos índices son utilizados por el sistema para hacer la fase de correspondencia y así recuperar imágenes relevantes, rechazando las que no concuerden. Los índices, que se almacenan en una estructura ó base de datos, son diseñados para hacer una recuperación eficiente.

2.1 Análisis de sistemas comerciales existentes.

A continuación se muestran algunos sistemas existentes de recuperación de imágenes y las características sobre la forma en que realizan la recuperación.

2.1.1 A-LIP (Automatic Linguistic Indexing of Pictures)

Sistema para anotar (clasificar) automáticamente imágenes en el Web [13]. Se basa en la segmentación y extracción de características de imágenes entrenadas, extrayendo su color promedio. Cada categoría de imágenes corresponde a un concepto que es manejado por un modelo estadístico, en este caso es el modelo oculto de Markov de multiresolución 2D (2D MHMM). El 2D MHMM representa 2 tipos de información: clústers de vectores de características y relaciones espaciales entre cada clúster. Un 2D MHMM se calcula para cada categoría por separado y se almacena en un diccionario de conceptos. Cada categoría de imagen en el conjunto de prueba, es manualmente comentada por lo que existe un mapeo entre el perfil del 2D MHMM y el conjunto de palabras. En una imagen de prueba el sistema consta de 3 componentes: el proceso de extracción de características, el proceso de modelado estadístico de multiresolución y el proceso estadístico de indexación. Una ventaja que tiene este sistema es que si la imagen representa un nuevo concepto, sólo los modelos estadísticos que se refieren a ese concepto necesitan ser re-entrenados. La desventaja que presenta es que las imágenes son 2D, por lo que se limita la habilidad de aprender conceptos de forma acertada [13].

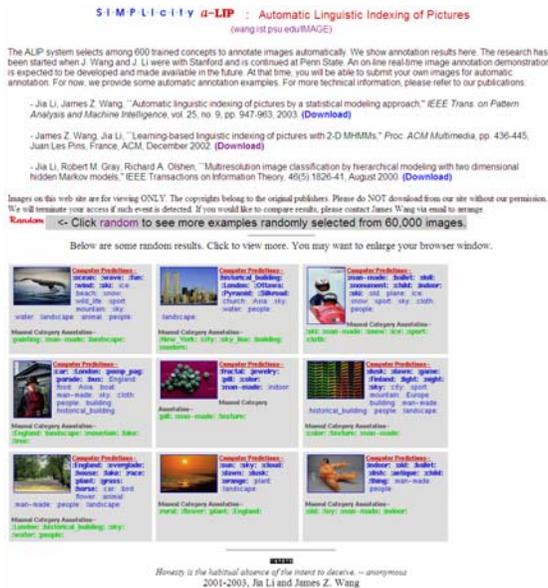


Figura 2.1: Sistema a-LIP [13].

La recuperación se basa en:

- Promedio de color.
- Textura, energía en bandas de alta frecuencia.

2.1.2 AMORE (Advanced Multimedia Oriented Retrieval Engine)

Sistema que permite la recuperación de imágenes, por su similitud [4]. El usuario puede elegir la categoría de imágenes en donde realizar la búsqueda. La imagen es segmentada en 8 regiones de color homogéneo con un tamaño de 24x24 píxeles. Las regiones se utilizan directamente para hacer la correspondencia. El usuario puede indicar la relevancia del color y forma. Primero se encuentra una correspondencia entre las regiones de la imagen buscada y la imagen considerada la correcta. La similitud entre 2 regiones se basa en el número de píxeles que se traslapan (correspondencia de plantilla). La similitud de colores es la distancia entre dos regiones uniformes de colores. Este sistema presenta la ventaja de refinar la consulta agrupando los resultados de la búsqueda que puede ser de 3 formas: Agrupación por composición de imagen, formando clusters de imágenes basándose objetos similares en formas de diferentes tamaños; agrupación por claves, imágenes que son semánticamente similares forman un cluster; agrupación de URL, el URL de un sitio web puede dar indicaciones del tipo de imagen que contiene, la agrupación es con keywords (claves) del contenido de los párrafos asociados.

Las pruebas de evaluación de la técnica de este sistema se realizaron con 400 imágenes, recuperando las imágenes, primero por similitud semántica y después aplicando la similitud visual a

las imágenes recuperadas. Obteniendo la siguiente gráfica donde se indica la similitud semántica y visual de 50 imágenes, que corresponden a una consulta por imagen. [4]

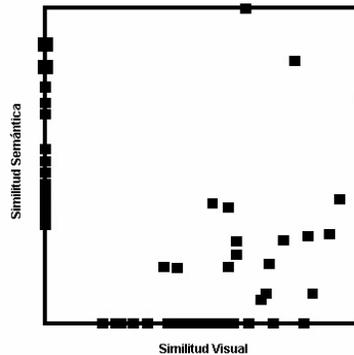


Figura 2.2: Similitud de 50 imágenes en una búsqueda por imagen. Eje X representa la similitud visual, eje Y representa similitud semántica [14].

La recuperación se basa en:

- Segmentación de la imagen en regiones homogéneas de color.
- Keywords para recuperación semántica usando un modelo de vector de espacio

2.1.3 C-BIRD (Content-Based Image Retrieval from Digital libraries)

Sistema de recuperación de información basado en contenido, tanto de video como de imágenes de Internet para librerías digitales [15]. La imagen es dividida en 64 celdas y para cada celda, se determinan los colores más frecuentes y el número de bordes para cada orientación. En cada imagen recolectada se calcula un descriptor de características y un descriptor de trazo. Un descriptor de características es un conjunto de 4 vectores: un vector de color, un vector de color más frecuente, un vector de la orientación de los bordes más frecuentes y un vector de cromaticidad. Los histogramas de orientación de texturas y de color, se comparan a través de la intersección de los histogramas. Para cada región se extrae la localización de color, así como datos geométricos: el área, el centro de masa y la excentricidad. Se filtran las imágenes en la base de datos que comparten el número de colores de las regiones con la imagen consultada y se producen vectores que miden la distancia del centro de masa de la primera región con los demás centros de masa. El análisis de estos vectores producidos y los ángulos entre ellos, produce una hipótesis de que existe un objeto de cierta escala y orientación, la cual es probada comparando el histograma de textura de cada par de regiones correspondientes en las 2 imágenes, finalmente se verifica la forma basándose en la Transformada de Houg [15].

La recuperación se basa en:

- Densidad de color.
- Textura.
- Formas.



Figura 2.3: Sistema C-BIRD [15].

2.1.4 CIRES. (Content Based Image REtrieval System)

Sistema de recuperación de imágenes basado en contenido, que combina características de bajo nivel y de alto nivel [16]. Este sistema de recuperación se basa en diferentes especificaciones de importancia que define el usuario, se consideran más de una consulta, tomando las características de, sólo estructura, sólo color, sólo textura, color y textura y estructura, color y textura. Se toman las distancias de estructura y color entre la imagen de consulta y la imagen de la base de datos, intersectando los histogramas de color de ambas imágenes.

La ventaja que presenta este mecanismo es que integra varias características, como color, textura y forma, la ponderación de cada característica se puede modificar, además la imagen de salida se reutiliza para hacer la consulta de nuevo, modificando la ponderación automáticamente.

Las pruebas realizadas a la técnica aplicada se realizaron con 10,221 imágenes, de 6 clases, en la Figura 1.4, se muestra la eficiencia de la recuperación de imágenes. La precisión aumenta al integrar más características. Se puede observar que los mejores resultados, se obtuvieron con la integración de las 3 características: estructura, color y textura [17].

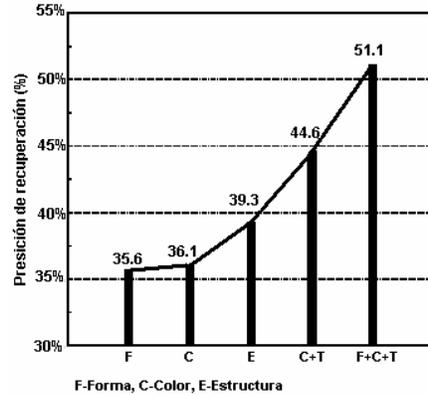


Figura 2.4: Precisión en la recuperación usando 5 metodologías: Solo estructura (E), sólo color (C), sólo textura (T), color y textura (C+T), y estructura, color y textura (E+C+T) [17].

La recuperación se realiza a partir de:

- Textura.
- Inferencias de consultas.
- Agrupamiento de regiones.
- Histogramas de colores

Query Image

Relevance feedback type: Cluster. Weights: Perceptual Grouping = 0.7, Color = 0.2, Texture = 0.1, L, A, B channels. For relevance feedback, please select the check boxes below each image, and then select the feedback type. Note that NS = "Not Sure".

Retrieved Images

Home Papers Research

Land: Cars

man_tsp_lnd_car_freefoto_cabse1.jpg

View Image Submit Query Reset Form

Weights (should sum to 1):

Perceptual grouping: 0.33 Color: 0.33 Texture: 0.33

Texture:

L, A and B channels
 L channel only (~Grayscale texture)

Figura 2.5: Sistema CIRES [16].

2.1.5 Excalibur Visual RetrievalWare.

Software para desarrollar aplicaciones de manipulación de archivos de imágenes digitales y su contenido visual, extracción de características, indexación y recuperación basada en contenido. La consulta se realiza especificando la importancia de los atributos con que se realiza la recuperación (Color, forma, textura, brillantez, estructura de color) [18].

La recuperación se hace a partir de:

- Curvaturas y contrastes de líneas en la imagen.
- Histogramas de colores HSV.
- Orientación relativa.
- Atributos de texturas.
- Formas.
- Cantidad de brillo.

2.1.6 PhotoBook.

Herramienta para responder a consultas de imágenes basada en contenido [19]. Compara características asociadas con las imágenes, implementando 3 estrategias para construir representaciones de imágenes para las consultas, cada una para un tipo específico de contenido: rostros, figuras 2D e imágenes de texturas. La descripción de la textura, se expresa como la suma de 3 componentes correspondiendo a la periodicidad, dirección y aleatoriedad. Para descripción de forma, se extrae el contorno y se eligen un número de puntos de éste, los cuales se utilizan como nodos para construir un modelo de la forma. Después se toman como puntos de correspondencia entre la nueva forma y el promedio de las demás formas. Las dos formas se comparan calculando la cantidad de esfuerzo necesario para deformar una forma para corresponder a las otras. [19].

La recuperación es de acuerdo a:

- Color
- Textura
- Formas.

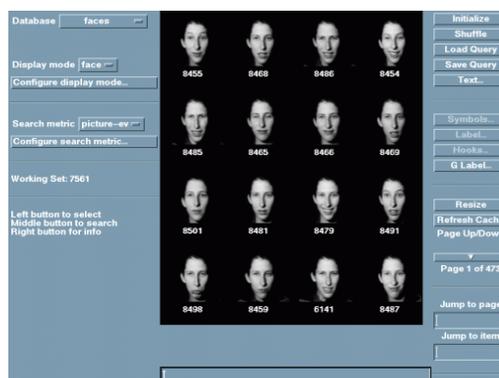


Figura 2.6: Sistema PhotoBook: base de datos para reconocimiento de rostros [19].

2.1.7 QBIC (Query by Image Content).

Sistema de recuperación de imágenes basado en contenido (porcentajes de colores, distribución del color y textura.) [3]. Se calculan las siguientes características de color: el vector de color 3D promedio de un objeto, histogramas de color RGB. En las características de textura se obtienen la aspereza, contraste y dirección, en las características de la forma se obtienen el área de la forma, la circularidad, la excentricidad, orientación y un conjunto de momentos invariantes. La correspondencia de dos histogramas de colores, se obtiene calculando la distancia del color promedio y la distancia entre colores promedios en el espacio RGB. La distancia de textura se obtiene con la Distancia Euclidiana entre vectores de características de forma. La ventaja que presenta QBIC es que implementa un método para recuperar imágenes basándose en un diseño de esquema del usuario, por lo que las imágenes en la base de datos son representadas por un mapa binario reducido de puntos de bordes. Para la correspondencia de esquemas, se realiza una correspondencia de plantilla, partiendo el esquema del usuario en 8 bloques de 8x8 píxeles y encontrando la máxima correlación de cada bloque con el esquema. El valor de la correspondencia es la suma de todas las correlaciones de todos los bloques locales. El promedio del color y la textura se indexan utilizando árboles-R [3].

La recuperación es a partir de:

- Histogramas de color.
- Especificación de regiones basadas en color.

2.1.8 Retrievr.

Herramienta que permite buscar imágenes por medio del trazo de esquemas (bosquejos realizados por el usuario) [20]. Su algoritmo considera las formas básicas y la información de color de la imagen de consulta. Previamente aplican una transformada a cada imagen en la base de datos para coleccionar sólo algunos coeficientes más grandes de esta transformada. Se obtiene una marca para cada imagen, finalmente se almacenan estas marcas para después compararlas con una nueva marca. Este proceso se aplica a la imagen de consulta, la marca obtenida se compara con las marcas de las imágenes de la base de datos y se elige la mejor correspondencia. Esta técnica, ya que requiere almacenar una cantidad pequeña de datos, hace más rápido el proceso de correspondencia. Se obtienen muy buenos resultados al realizar como consulta esquemas de dibujos a mano o imágenes escaneadas que pudieran tener baja calidad. [20].

Su recuperación se basa en:

- Trazado de esquemas de objetos dentro de la imagen.

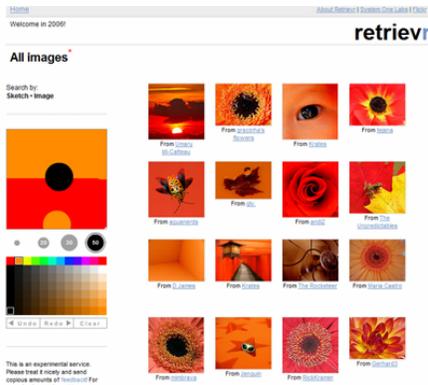


Figura 2.7: Sistema Retrievr [20].

2.1.9 SIMPLicity (Semantics-sensitive Integrated Matching for Picture Libraries).

Sistema de recuperación de imágenes que hace uso de la clasificación de semántica (con textura/sin textura y gráfico/fotografía) [10]. La extracción de características se realiza basándose en transformadas Wavelets, así como en la correspondencia basada en la segmentación de imagen. Implementa una medida de similitud basada en regiones, incorporando propiedades de todas las regiones segmentadas, para que la información sobre una imagen pueda ser utilizada. En la etapa de indexación, el sistema parte una imagen en bloques de 4x4 y extrae un vector de características por cada bloque. Un algoritmo estadístico de formación de clusters se utiliza para segmentar de forma rápida la imagen en regiones. La segmentación se alimenta a un clasificador que decide el tipo de semántica de la imagen. Las características que reflejan color, textura, forma e información de localización son extraídas para cada región en la imagen. La marca que distingue a cada imagen es la colección de características para todas sus regiones, las marcas de las imágenes con varios tipos de semántica, se almacenan en bases de datos separadas. Si la consulta no se encuentra en la base de datos, ésta pasa al mismo proceso de extracción de características utilizado en el indexado. Para una imagen en la base de datos, se revisa primero su tipo de semántica y después se extrae su marca, después se muestran las imágenes que se acerquen al mismo tipo de semántica de la imagen de consulta. El sistema presenta la ventaja de que ser robusto a las alteraciones de la imagen, tales como variación de intensidad, distorsión de color, traslaciones, rotaciones y recortes de márgenes. La recuperación se basa en:

- Correspondencia de regiones de imágenes segmentadas.
- Regiones caracterizadas por color, textura, forma y localización.



Figura 2.8: Sistema SIMPLiCity [10].

2.1.10 SQUID System (Shape Queries Using Image Databases).

Sistema para la recuperación de imágenes en el Web [5]. El contorno de una imagen se representa por tres parámetros globales de la forma: excentricidad, figura circular y la proporción entre la altura y la anchura de la imagen. Para la correspondencia, se eligen de la base de datos, aquellos modelos cuyos parámetros globales son suficientemente cercanos a los parámetros globales de la consulta. Utilizando la distancia Euclidiana entre parejas de correspondencias.

La recuperación de imágenes se basa en:

- Especificación de formas de objetos (contorno).

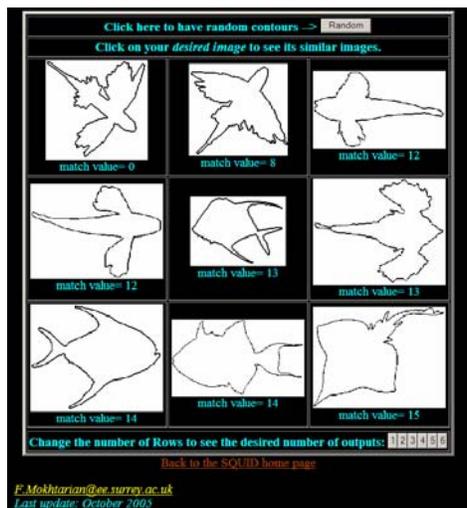


Figura 2.9: Sistema SQUID System [5].

2.1.11 VisualSEEK.

Sistema para la búsqueda de imágenes, por sus características visuales [2]. Cada imagen se descompone en regiones equitativas de colores dominantes. Para cada región, las propiedades de características espaciales se retienen para las siguientes consultas. Primero se encuentran las imágenes que contienen arreglos más parecidos de regiones similares. En la extracción de color, se define un vector binario de dimensión 166, correspondientes a los colores en el espacio HSV. Para establecer una consulta, el usuario diseña un número de regiones, posiciones y sus dimensiones en una cuadrícula y selecciona un color para cada región. También puede indicar bordes o relaciones espaciales entre regiones. Para la correspondencia de imágenes, se realizan independientemente las consultas de color, localización de regiones y área.

La técnica fue evaluada con 500 imágenes, los resultados obtenidos se muestran en la Figura 1.10. Se observa que la estrategia de espacio de colores (Q2), tiene mayor desempeño de recuperación, que la estrategia de histograma de color (Q3). La estrategia de región de colores (Q1) tiene un mayor desempeño que el espacio de colores. La diferencia representa la pérdida de información en el proceso de extracción de regiones [21].

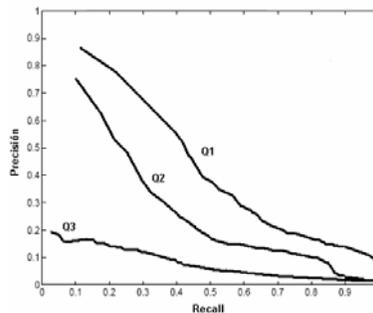


Figura 2.10: Recuperación de 100 consultas aleatorias a la base de datos. Q1 - región de colores, Q2 – espacio de colores, Q3 – histograma de color. [21]

La recuperación es a partir de:

- Localización en un espacio.
- Regiones en la imagen basadas en color.

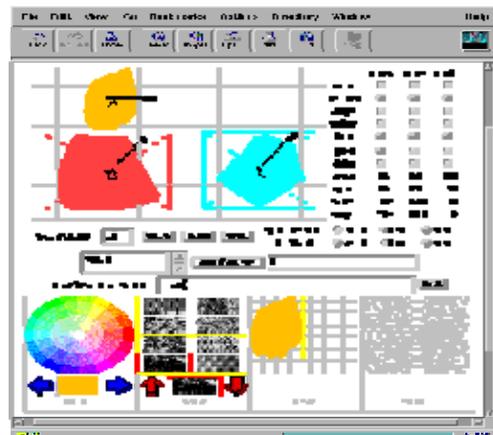


Figura 2.11: Sistema VisualSEEK [2].

2.1.12 WebSEEK.

Es una herramienta de búsqueda de imágenes y video para el World Wide Web, que se basa en su contenido. Los resultados de la consulta usan una comparación, basada en color a toda la base de datos y seleccionan la imagen con mayor similitud de color. El sistema presenta la ventaja de que el usuario puede seleccionar ejemplos positivos y negativos partiendo del resultado de la consulta para reformularla [22].

La recuperación se realiza partiendo de:

- Color y textura.
- Posición espacial.

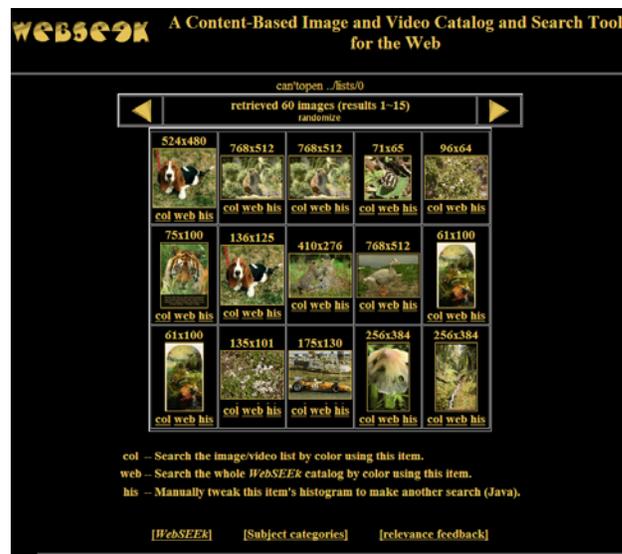


Figura 2.12: Sistema WebSEEK [22].

2.2 Comparación de métodos utilizados en sistemas comerciales.

Tabla 2.1: Tabla comparativa de los sistemas anteriores de recuperación de imágenes.

Sistema de recuperación de información visual	Características de búsqueda	Organismo
a-LISP (Automatic linguistic indexing of pictures)	- Color promedio de bloques. - Textura, a través de transformadas Wavelets.	The Pennsylvania State University
Amore (Advanced Multimedia Oriented Retrieval Engine)	- Regiones homogéneas de colores. - Keywords para recuperación semántica usando un modelo de vector de espacio.	NEC USA Inc. C&C Research Laboratories.
CIRES (Content-Based Image Retrieval System)	- Histograma de color (espacio de colores). - Textura.	University of Texas at Austin.
C-Bird (Content-Based Image Retrieval from Digital Libraries)	- Color. - Textura. - Color más frecuente. - Vector de orientación más frecuente. - Vector de cromaticidad.	Escuela de Ciencias de la Computación de la Universidad de Simon Fraser.
Excalibur Visual RetrievalWare	- Aspecto de proporción. - Brillo. - Color local y global. - Forma. - Textura. - Orientación relativa. - Curvatura. - Histograma de color.	Excalibur Technologies ahora Convera Corporation.
PhotoBook	- Forma. - Textura. - Descripción textual.	MIT MediaLab.
QBIC (Query By Image Content)	- Promedio de color. - Texturas. - Contraste. - Area. - Circularidad. - Excentricidad. - Momentos invariantes.	IBM
Retrievr	- Transformadas Wavelet para obtener coeficientes de color	University of Washington System One Inc.
SIMPLiCity (Semantics-sensitive Integrated Matching for Picture Libraries)	- Histograma de color. - Capas de colores. - Regiones de colores. - Texturas. - Forma y espacios.	The Pennsylvania State University
SQUID (Shape Queries Using Image Databases)	- Excentricidad. - Circularidad. - Forma. - Proporción de curvatura.	CVSSP (Centre for Vision, Speech and Signal Processing).
VisualSEEK	- Color y textura. - Regiones de colores.	Columbia University
WebSEEK	- Keywords de tags de imágenes web. - Histograma de color.	Columbia University

2.3 Representación de imágenes

2.3.1 Descripción de color

El color es una característica que utilizan la gran mayoría de los sistemas de recuperación basada en contenido, es independiente al tamaño de la imagen y a su orientación. Existen muchos trabajos que utilizan diferentes métodos para la comparación de color, por ejemplo [23]:

- Correspondencia de color, para encontrar imágenes que contiene colores específicos en proporciones asignadas.
- Búsquedas en similitud, para encontrar una lista clasificada de imágenes similares a una imagen proporcionada o un esquema trazado por el usuario.
- Búsquedas de regiones, para encontrar imágenes que contienen regiones de color como las que contiene la imagen especificada en la consulta.
- Búsquedas que cumplan con los objetivos, para encontrar una lista de imágenes en las que aparece algún objeto que fue especificado por el usuario.
- Categorías semánticas, para agrupar imágenes en categorías significativas, como un gráfico versus fotografías, ó imágenes de interiores versus exteriores.
- Recuperación de imágenes con efectos de colores.

Estas tareas dependen de la eficiencia de las características de color que se puedan representar en el contenido de imágenes. No existe una representación simple de color que sea la más eficiente de todas, solamente utilizando múltiples representaciones que definen la característica de color desde diferentes perspectivas. Dependiendo del contexto, es decir las características seleccionadas, las propiedades básicas que deberían ser presentadas son[23]:

1. **Similitud de percepción:** la característica de distancia entre dos imágenes es muy grande sólo si las imágenes no son similares.
2. **Eficiencia:** se pueden calcular rápidamente.
3. **Economía:** sus dimensiones son pequeñas para que no afecte la eficiencia del proceso de recuperación.
4. **Escalabilidad:** la ejecución de el sistema no es influenciado por el tamaño de la base de datos.
5. **Robustez:** cambios realizados en las condiciones de similitud de las imágenes de la base de datos no afecta la recuperación.

Muchos métodos de recuperación de imágenes se han propuesto, la mayoría basándose en variaciones en el uso de histogramas de color. El histograma representa la frecuencia relativa de

ocurrencia de varios niveles de colores en la imagen. Los colores del histograma de cada imagen se calculan y se almacenan en una base de datos.

Cuando los usuarios ejecutan una consulta, especificando la proporción de cada color o una imagen ejemplo, se calcula el histograma de colores. Los algoritmos de correspondencia, recuperan aquellas imágenes de las cuales los histogramas de colores tienen correspondencia con el histograma de la consulta. Este método de comparación de color, es eficiente si las imágenes u objetos son del mismo tipo (por ejemplo en forma) y solo tienen diferencia de color, recuperando una imagen con características similares a la imagen consulta [23].

2.3.2 Descripción de textura

La textura es un componente clave para la percepción visual del humano. Todos pueden reconocer una textura, pero es difícil describirla. A diferencia del color, la textura se distingue en una región y no en un punto. Las texturas pueden ser definidas como patrones homogéneos visuales, presentados en materiales, como madera, piedras, telas, etc. Las texturas no se pueden percibir fácilmente como objetos aislados. La textura tiene cualidades como periodicidad y escala, también se puede distinguir en términos de dirección, contraste y aspereza. En el análisis de similitud entre texturas en sistemas de recuperación de información visual, éstas son muy utilizadas para distinguir entre áreas con color similar. Una técnica es el cálculo del grado de contraste, aspereza, dirección y regularidad, también se puede usar la periodicidad, dirección y medidas aleatorias. Otros métodos son el uso de los filtros de Gabor y fractales, matrices de ocurrencia que almacenan información estadística, características psicológicas y Wavelets de Gabor [24].

Las texturas dependen de la resolución de la imagen, pareciendo un patrón repetitivo partiendo de cierta distancia. Éste patrón no se puede distinguir cuando la imagen se amplía, aunque pueden aparecer otro tipo de patrones.

Existen 2 formas de procesamiento para el análisis de texturas, una es el análisis estadístico, en donde se analizan valores estadísticos de niveles de grises, así como el número de picos y otras características espaciales; otra forma es el análisis de frecuencias, debido a que la textura es una repetición espacial de un patrón, se pueden obtener valores grandes o picos, utilizando la transformada de Fourier [25].

La textura es una característica discriminativa, la cual puede describirse, en base a su espacio, frecuencia o propiedades perceptuales. La periodicidad, aspereza, dirección y grado de complejidad, como se mencionó anteriormente, son algunos de los atributos que se pueden recuperar de una textura. Las características espaciales que se basan en estos atributos, son

interesantes para recuperar imágenes por similitud de textura. Al igual que los colores, la textura es conectada con efectos psicológicos.

Una descripción eficiente de la textura, se puede basar en propiedades estadísticas y estructurales de patrones de brillo. Existen técnicas para extraer descriptores de textura significativa de las imágenes, estas se basan en diferentes modelos que se clasifican de la siguiente manera [25]:

- Basados en espacio
 - Dimensión fractal.
 - Función de auto-correlación.
 - Matriz de co-ocurrencia.
 - Modelos auto-regresivos.
 - Modelos estocásticos.
- Basados en frecuencia
 - Espectro (Spectrum Power).
 - Transformadas Wavelets.
- Firmas de textura
 - Contraste, asperosidad y direccionamiento.
 - Energía textual.
 - Repetición, direccionamiento, complejidad.

2.3.3 Descripción de forma.

El definir la forma de los objetos de una imagen, representa la manera más certera para reconocer estos objetos y recuperar imágenes que contienen estos objetos. La correspondencia de objetos a través de la descripción de su forma, ha sido una de las tareas más difíciles para los sistemas de recuperación de imágenes basados en contenido, esto es debido a que la descripción de la forma de un objeto requiere muchos parámetros para ser representada. La forma de un objeto es su contorno, un patrón geométrico, que consiste en una serie de puntos, curvas, superficies, sólidos, etc.

La característica de forma se representa a través de un vector y se almacenan para después ser indexado. La entrada también puede ser un bosquejo que realiza el usuario.

Existen técnicas que se pueden utilizar para realizar los vectores de características y poder hacer el proceso de correspondencia de formas. En el proceso de correspondencia, se utilizan varios métodos como la deformación elástica de plantillas [26], histogramas de dirección de contornos, etc.

Hay tres categorías de técnicas para representar la forma de un objeto:

1. Método de vector de características, que consiste en representar la forma en un vector numérico y la similitud entre 2 formas, se calcula mediante la medición de distancia como la Euclidiana (Ecuación 2.1) o la Hausdorff (Ecuaciones 2.2 y 2.3).

$$d_E(F_1, F_2) = \sqrt{\sum_{i=0}^{Nc} (f_1^i - f_2^i)^2} \quad (2.1)$$

$$H(F_1, F_2) = \max(h(F_1, F_2), h(F_2, F_1))$$

donde $h(F_1, F_2) = \max_{f_1 \in F_1} \min_{f_2 \in F_2} \|f_1 - f_2\|$ (2.2) y (2.3)

donde f_1 y f_2 son los valores del vector de características en la posición i . Nc es el total de posiciones del vector de características. F_1 y F_2 son los dos vectores de características de dos imágenes diferentes.

2. Método relacional, donde la forma se divide en un conjunto de componentes que sobresalen. El descriptor final se compone de los descriptores de cada componente sobresaliente y la relación entre ellos.
3. Método de transformación, se describe una forma basándose en el esfuerzo para transformarla en otra forma. La distancia de transformación se utiliza para calcular la similitud entre dos formas. Se consideran factores como tolerancia a la oclusión y deformación, robustez al ruido, factibilidad de indexación, etc.

Existen distinciones entre métodos de representación de forma y métodos de descripción de formas. Los métodos de representación de formas son representaciones no numéricas de formas, donde la característica importante de cada una se conserva, como por ejemplo, en algún grafo. Los métodos de descripción de formas son representaciones numéricas de formas, los resultados de este tipo de métodos son los vectores de característica de forma. La rotación, traslación y escala de la forma de un objeto no cambia, los vectores de características deben ser invariantes a estas transformaciones.

Existen diferentes métodos de análisis de forma, de contorno o externos y globales o internos. Los de contorno se basan en los puntos de los bordes de la forma, analizan los contornos o utilizan la transformada de Fourier del contorno. Los métodos internos utilizan las siguientes técnicas: eje medio, métodos basados en momentos y descomposición de formas [27].

Cuando se especifica una imagen de consulta, se extraen sus características de color de la base de datos, aplicando el mismo procedimiento que a las imágenes de entrada. La recuperación de imágenes entonces, es realizada por el procedimiento de correspondencia, el cual compara las

características de la imagen consulta con las características de las imágenes almacenadas. El mecanismo de correspondencia, hace la recuperación de acuerdo a la métrica o medida de similitud seleccionada. Las imágenes almacenadas son ordenadas de acuerdo a su similitud con la imagen consulta.

2.4 Métodos para la recuperación de información basados en contenido.

De la bibliografía que se revisó, se encontraron algunos sistemas de recuperación de imágenes que trabajan en varios ámbitos. Para conocer las características con las que cuentan los sistemas actuales y como funcionan, se realizó un pequeño análisis de algunos de ellos. Las formas de objetos contenidos en las imágenes son características importantes del contenido semántico de la imagen. La forma de los objetos puede ser usada para definir un índice y así obtener su representación y medir la similitud. Se pueden recuperar imágenes de una base de datos utilizando técnicas de coincidencia de patrones.

Algunos trabajos codifican la configuración global de una imagen, usando relaciones fotométricas y espaciales dentro y entre regiones de la imagen. Este método presenta problemas ya que ésta codificación solo es efectiva para categorías de escenas que están geoméricamente bien definidas, no se puede generalizar a imágenes donde las partes y objetos se localizan aleatoriamente [28].

Los sistemas de búsqueda de imágenes existentes abarcan la búsqueda a través de texto, búsqueda por similitud de características de imágenes, búsqueda por características de segmentos y búsqueda por medio de esqueletos de imágenes. Algunos sistemas combinan texto e imágenes como datos.

2.4.1 Métodos probabilísticos.

Trabajos como [29] proponen hacer búsqueda de imágenes calculando la probabilidad que tiene cada imagen de relacionarse con los campos de la consulta. Se establece un modelo jerárquico, donde los datos son modelados generando una jerarquía de nodos y cada hoja de la jerarquía corresponde a un cluster. Cada nodo en el árbol tiene una probabilidad de generar alguna palabra y también tiene la probabilidad de generar un segmento de la imagen con las características obtenidas. Utilizan vectores de características que toman las características extraídas como el tamaño, posición, color, textura y forma, usando la distribución Gaussiana para segmentos de imagen, los almacenan en un modelo jerárquico porque soporta mejor la recuperación de grandes colecciones de imágenes. Una ventaja que presenta esta estructura utilizada es que, debido a que la información de

la imagen se comparte entre los niveles más altos de los nodos, su representación es más compacta que una estructura no jerárquica, proporcionando una ejecución adecuada. Al realizar la prueba de este modelo, utilizaron el algoritmo de Máxima Expectativa, donde se introducen variables ocultas, que inicializan el clúster y los niveles. Este método no presentó buenos resultados al utilizar sólo características de imágenes, ya que se pierde parte de la semántica. La “clusterización” es más apropiada utilizando tanto características de segmentos de imágenes como el texto, ya que las personas relacionamos las imágenes usando tanto semántica como contenido visual.

El modelo identifica que palabras y segmentos de imágenes pueden ser relacionados dando una determinada entrada.

2.4.2 Métodos basados en color.

En trabajos como el de Madirakshi [30], se enfocan en el uso de conocimiento de dominio, para la recuperación de imágenes para aplicaciones especializadas en dominios restringidos. En este trabajo se enfocan principalmente en recuperar imágenes de flores introduciendo una imagen ejemplo o el nombre de algún color. Definen un sistema de clasificación de color en lenguaje natural, utilizado para interpretar el conocimiento de dominio que se basa en colores, a reglas para la segmentación automática de la región de interés a partir del fondo. El objeto es segmentado del fondo de la imagen, antes de que la forma sea computada para extraer la región que presenta una alta probabilidad de ser parte de la flor. El problema que presenta la recuperación a través de especificar el color, es que el color como característica podría no relacionarse con el contenido de la imagen, por ejemplo si se busca un tipo determinado de flor.

Otros trabajos como [31], integran el acceso por contenido a base de datos, realizando búsquedas por descripción textual, por formas de la imagen, por similitud de patrones y por color. Especificando un método que permita acceder simultáneamente por contenido textual a través del etiquetado de metadatos por el autor, por tolerancia y porcentaje de color, por similitud de correlaciones espaciales de los colores en las imágenes y por formas geométricas. Dichos metadatos los codifican en RDF (Resource Description Framework) y los incorporan al propio fichero imagen. Aunque su trabajo intenta unificar métodos de clasificación automática y clasificación en base a anotaciones, su proceso de recuperación según sus resultados, fue costoso y subjetivo. Tomando en cuenta que el enfoque de su dominio de aplicación, se reducía a imágenes de un catálogo real de telas de diferentes colores y patrones.

2.4.3 Métodos jerárquicos con texto.

Existen muchas aplicaciones de utilidad en la recuperación de imágenes biomédicas, que ayudan a diagnosticar, intervenir y dar seguimiento a un paciente. Tradicionalmente indexan su base de datos utilizando texto, basado en palabras claves y categorías jerárquicas. En este ámbito existen actualmente códigos de etiquetado para la clasificación de enfermedades, pero resulta poco práctico usar estos sistemas de codificación, porque la cantidad de los archivos imagen se incrementa rápidamente. Otra desventaja que presentan estos métodos es que los sistemas están limitados por su vocabulario predefinido que es, subjetivo, incompleto, burdo y ambiguo como se menciona en [32] y [33]. Su trabajo pretende encontrar el mejor y más pequeño subconjunto de características de una imagen, que capture su semántica y que a su vez la semántica se clasifique en categorías. Por lo que se requiere un conjunto explícito de etiquetas de imagen y suficiente espacio de memoria de la computadora, tanto para almacenar grandes matrices de atributos de características como para soportar la búsqueda extensa sobre esta matriz.

2.4.4 Lógica difusa.

Otro enfoque utilizado para la recuperación de imágenes consiste en el uso de la Lógica Difusa. En [34] se aplica la lógica difusa, utilizando variables de lenguaje difuso para describir el grado de similitud de las características de la imagen, hacer uso de la inferencia difusa para asignar pesos entre las características de la imagen y expresar la subjetividad de la percepción humana a través de reglas difusas. Este trabajo propone una mejora en el histograma tradicional llamado Histograma de Área Promedio (Average Area Histogram - AAH) para representar las características de color, basado en las características de área de las regiones formadas por píxeles de cada color. El sistema de recuperación basado en lógica difusa, consta de 4 partes:

1. **Extracción de características:** El color se representa por el histograma HSV. Utilizan momentos invariantes para extraer la forma.
2. **Fuzzier:** Basándose en la distancia de color y de forma de la imagen solicitada y cada imagen de la base de datos, se infiere la similitud de las imágenes partiendo de variables como "muy similar", "similar" y "no similar" que describen la diferencia e características.
3. **Inferencia difusa:** Se crea una base de regla difusa, de acuerdo al conocimiento general de un objeto y los requerimientos del usuario.
4. **Defuzzier:** Se obtiene la similitud de 2 imágenes descrita por 3 variables difusas: "muy similar", "similar" y "no similar".

La ventaja que presenta dicha investigación es que para imágenes con una gran cantidad de apariciones, el método es robusto en la clasificación de imágenes.

2.4.5 Métodos híbridos.

Finalmente también existen trabajos que presentan métodos híbridos para recuperación de información visual, que combinan análisis de objetos dentro de las imágenes con su indexación automática de descripción textual [26]. Aplicando técnicas semánticas para la descripción de información visual en sistemas que usan características de bajo nivel, las consultas de los usuarios son tomadas según el cálculo de la convergencia de formas con su correspondiente anotación ontológica del objeto en la imagen. El vector de características contiene tanto características de bajo nivel como descripción semántica. Los problemas que presenta este método es el muestreo espacial durante la generación del vector de características, así como la capacidad de memoria requerida. Aún así, demuestra ser una alternativa que facilita la recuperación de imágenes además de mostrar un eficiente funcionamiento, con mayor rapidez, debido al menor número de iteraciones necesarias en el proceso de búsqueda al utilizar ontologías [26].

2.5 Extracción y asignación de semántica de imágenes.

El término semántica, se refiere a los aspectos del significado o interpretación de un determinado código simbólico, lenguaje o representación formal. Cualquier medio de expresión (código, lenguas, etc), admite una correspondencia entre expresiones, de símbolos o palabras y situaciones o conjuntos, de cosas encontrables o inferibles en el mundo físico y cosas abstractas, que pueden ser descritas por dicho medio de expresión.

Algunos trabajos existentes se enfocan en la extracción y asignación de semántica en las imágenes. A continuación se explican brevemente.

En [35], se diseñó una herramienta para la asignación y búsqueda semántica en una colección de imágenes de arte, utilizando para ello, varias ontologías representadas en RDF *Schema*. Almacenaron 3 grupos de descriptores, los relacionados a la producción que son título de la obra, creador, fecha, periodo, técnica y cultura; los de tipo físico como materiales, medidas y tipo y los administrativos como localización, colección a la que pertenecen, fuente y derechos. Dicha herramienta además provee un espacio donde introducir información a almacenar, la cual no concuerda en ningún espacio de la estructura o no está presente en las ontologías. Las ontologías contienen términos que las relacionan entre sí. Cuando un usuario busca un término, se puede invocar a un mecanismo de sugerencias, donde se despliegan las opciones relacionadas con la entrada. Este trabajo presenta limitaciones en la expresión en cuanto a la necesidad de expresar las propiedades de cardinalidad y equivalencia entre recursos (clases, instancias, propiedades).

El sistema SemView, como especifica [36], soporta consultas visuales y de semántica. Resume el contenido visual y semántico de cada base de datos en una meta base de datos, la cual es una colección de metadatos estadísticos visuales y semánticos que se generan automáticamente. La consulta es guiada a las bases de datos respectivas y las imágenes más relacionadas se despliegan al usuario. Si la consulta es semántica, ésta se compara con los metadatos semánticos, para después generar una lista de las ubicaciones de las imágenes relacionadas a la consulta. En la extracción de la semántica de la imagen, siguieron el método de [37] que consiste en 3 niveles: nivel de píxel, que se utiliza una clasificación de color-textura para formar un catálogo semántico, basado en las categorías semánticas de imágenes de prueba; a nivel de región, se utiliza el catálogo para segmentar las imágenes en regiones; a nivel de imagen, se definen el contenido y contexto de las regiones de la imagen y se representan para apoyar a la extracción de semántica de las imágenes.

Para establecer la semántica, definen las categorías principales de semántica, en base a las restricciones de imágenes. Entrenando imágenes con color-textura que pertenecen a esa categoría y clasificando cada píxel en una celda del espacio de color-textura. Para cada celda se cuenta el número de píxeles que contiene, almacenándose en una tabla estadística entre celdas y categorías semánticas. Estos datos estadísticos se calculan por la similitud de las distribuciones, entre el contenido semántico de las imágenes de la base de datos y los metadatos semánticos.

En investigaciones como en [38], trabajan con generación de anotaciones automáticas a la imagen, para clasificarla basándose en clasificadores binarios. También se trabaja con agrupaciones confiables, que organizan las bases de datos de acuerdo a conceptos, extraen características de bajo nivel como el color y de alto nivel como autor. Las características de bajo nivel se extraen utilizando diferentes descriptores: un descriptor para la estructura de color, que expresa características de color locales. Crean un histograma de color en un bloque de 8x8, un descriptor de color escalable: que mide la distribución del color en toda la imagen, produciendo un histograma de color creado por la transformada Haar, un descriptor de capa de color, que captura la distribución espacial de color de una imagen, un descriptor de histograma de bordes, que captura la distribución espacial de los bordes, dividiendo la imagen en 16 bloques y calculando 5 direcciones de bordes para cada bloque y un descriptor de textura homogénea, que describe la dirección, regularidad y patrón en la imagen. Para la clasificación de semántica de la imagen, se utilizan los descriptores de bajo nivel basándose en estadísticas. Un método que utilizaron es el clasificador del K-vecino más cercano, que define límites no lineales dando la misma etiqueta, a una consulta del principal k vecino más cercano en el espacio de características.

Este tipo de clasificaciones se orientó a distinguir imágenes en el interior/exterior y entre ciudades/paisajes, por lo que no puede manejar un mayor enfoque de definición de objetos.

Otras investigaciones como [39] generan las anotaciones de las imágenes, dividiendo cada imagen en regiones. Cada región es descrita por un vector de características, que es valorado continuamente dado un conjunto de imágenes de entrenamiento, con anotaciones, calculando un modelo probabilístico de características de la imagen y palabras, lo cual permitirá predecir la probabilidad de generar una palabra, dadas las regiones de las imágenes. Esto se puede utilizar para hacer las anotaciones automáticamente a las imágenes y recuperar las imágenes de una consulta de texto. La asociación de diferentes regiones especifica el contexto, las asociaciones de palabras con la región de la imagen determinan el significado. El modelo que proponen, calcula la probabilidad de características de imagen de diferentes regiones, utilizan un conjunto ya entrenado, asocian esta probabilidad para anotar y recuperar las imágenes. Es un modelo estadístico para aprender automáticamente la semántica de las imágenes. Se predice la probabilidad de generar una palabra dadas las características obtenidas sobre las diferentes regiones de una imagen. La ventaja de este método es que se pueden predecir las probabilidades de forma natural, sin necesidad de definir etapas intermedias de agrupación.

2.6 Conclusiones.

En este capítulo primero se analizaron algunos sistemas comerciales de recuperación de información visual, se estudiaron sus características de recuperación, cada uno se basa en diferentes características como se observó, la mayoría de ellas se basan en color y forma, por esto se decidió utilizar estas características como características de bajo nivel, para hacer la comparación. Se tomarán las técnicas que utilizan los sistemas que se analizaron, como el cálculo de histograma para la comparación de color como el sistema CIRES [16] y el método de evolución de la curva utilizado en el sistema SQUID [5]. También se realizó el análisis de algunas técnicas para la representación del contenido de imágenes, como es la representación basada en color, representación basada en textura, la representación basada en forma y la representación semántica. También se analizaron diferentes trabajos donde se utilizan métodos basados en contenido para la recuperación de imágenes, como los métodos probabilísticos, el uso de lógica difusa, métodos híbridos, entre otros. En la asignación de semántica en las imágenes se describen brevemente algunas clasificaciones y niveles de semántica establecidos. En base a las ventajas de los métodos analizadas sobre descripción de color, textura y forma, se puede decir que la forma de un objeto, es una de las características más importantes de bajo nivel para la percepción humana ya que distinguimos a los objetos por su forma, este trabajo de tesis se enfocará a la búsqueda de similitud de imágenes. Se decidió representar la forma de un objeto, por medio de un vector de características, que como se analizó, se puede definir de una manera sencilla y con el cual se puede calcular fácilmente la

diferencia entre dos formas, aplicando la distancia Euclideana o la Hausdorff. Además se decidió utilizar la característica de color, ya que la mayoría de los sistemas se basan en variaciones de histogramas, mostrando ser eficiente en imágenes del mismo tipo. La indexación se basará principalmente en la forma, aplicando el color como filtro, para después combinar el conjunto resultante con su representación y clasificación de semántica que define a una imagen, mediante una estructura ontológica. La técnica de representación de forma que se utilizó se define en el siguiente capítulo. El objetivo es mejorar el desempeño de técnicas de recuperación de información como los métodos jerárquicos con texto que se pueden integrar en los métodos híbridos. Para esto se utilizará una estructura ontológica, que facilite el manejo del proceso de recuperación definiendo un modelo rápido, fácil y sencillo de implementar. Integrando técnicas de recuperación con las que se obtengan resultados satisfactorios en las búsquedas especificadas.