

# Capítulo 1. Introducción

El WWW es la mayor fuente de imágenes que día a día se va incrementando. Según una encuesta realizada por el Centro de Bibliotecas de Cómputo en Línea (OCLC) en Enero de 2005, 11.5 billones de páginas web y del orden de mil millones de imágenes en línea, cuyo número aumenta anualmente de un 50% a un 70% de manera exponencial. Para que los usuarios manejen esta información de forma eficaz, se necesitan herramientas para buscar, recuperar y clasificarla.

La búsqueda, recuperación, visualización y procesamiento de imágenes se lleva a cabo en una gran variedad de dominios, como medicina, patrimonio cultural, medio ambiente, ingeniería, etc. Tradicionalmente las búsquedas de imágenes son a través de descripciones textuales, especificando claves. Las imágenes son recuperadas basándose en la asignación manual de claves; este método de asignación es lento, caro y poco práctico para una gran cantidad de imágenes. Motores de búsqueda como Google, Yahoo, etc. basan sus capacidades de búsqueda principalmente en características de texto asociadas a las imágenes. El uso de claves puede ser una forma efectiva para buscar imágenes si cada imagen es descrita exhaustivamente y el buscador de imágenes tiene una clara idea de cual descriptor usar. Además, la asignación de claves puede ser equívoca e incompleta, no pueden especificar por sí mismas (codificar) toda la información que se presenta en una imagen.

## 1.1 Definición del problema

Cuando un usuario se encuentra ante un gran repositorio, cuyo contenido en parte es desconocido, es difícil de expresar o no se tiene una clara idea de cómo describir en palabras una consulta en algún buscador, es por esto que se ve la necesidad de desarrollar métodos eficientes para la recuperación de información visual. Los resultados de recuperación de imágenes deben proveer las imágenes que un humano seleccionaría, al buscar en el repositorio donde se almacenan. Esta situación presenta un problema, ya que la idea que tiene un humano sobre la semántica de la imagen, es difícil de codificar en un algoritmo de manera automática. Los humanos tendemos a describir las imágenes basándonos en los objetos que están representados en ellas, este tipo de descripción es más probable que produzca resultados de correspondencia entre la percepción humana y el contenido de la imagen.

Actualmente existen métodos utilizados en sistemas de recuperación de imágenes, que se basan en la recuperación de características de bajo nivel (como color, forma, textura, espacio, etc.) [15, 16, 18]. El problema de estos métodos, es que en algunos casos no proveen una indexación correcta de las imágenes, ya que no reflejan la semántica que presenta una imagen, además

necesitan de un gran número de iteraciones para obtener resultados coherentes con las consultas especificadas. Por otra parte, existen trabajos donde integran anotaciones semánticas a las imágenes, los métodos que utilizan consisten en asignar estas anotaciones basadas en opiniones humanas, las cuales pueden representar ideas ambiguas o significados diferentes [7]. También existen métodos híbridos que integran características de bajo nivel y semántica [26], los problemas que se presentan son en el muestreo espacial al formar los vectores de características y en la cantidad requerida memoria del sistema. El problema que se pretende resolver es proponer un modelo que recupere la semántica de una imagen basándose en sus características de bajo nivel, de una forma fácil y que trate de mejorar los problemas que presentan los modelos híbridos existentes.

## 1.2 Objetivos

El objetivo general de esta investigación es el análisis de métodos de bajo nivel y el modelado de sistemas para la recuperación de imágenes en colecciones digitales, integrando técnicas de CBIR (Content-Based Image Retrieval). El propósito es mejorar el proceso de recuperación proponiendo un modelo basado en las ventajas que presentan algunos métodos de recuperación existentes. Aplicándose particularmente a la exploración de métodos para la recuperación de imágenes, por medio de la extracción de sus características de bajo nivel, las cuales se integran con características de alto nivel, es decir, métodos de interpretación de contenido de imágenes. Para cumplir el objetivo general, se lograron los siguientes objetivos específicos:

- Análisis y selección de sistemas de recuperación de imágenes, como a-Lisp [13], AMORE [4], QBIC [3], SQUID [5], entre otros.
- Análisis de los métodos de recuperación de imágenes como los métodos probabilísticos, los métodos basados en color, los métodos jerárquicos con texto, Lógica difusa y los métodos híbridos.
- Análisis de algunos métodos tanto estructurales como globales para extracción de características de forma. Entre los estructurales se mencionan Código de cadena, Descomposición de polígono, Representación B-splines, Curva suavizada, Espacio de escala y Análisis sintáctico. De los métodos globales se pueden mencionar Excentricidad, Firmas de forma, Momentos, Evolución de la curva, Correspondencia por elasticidad, Representación de espacio tangencial y Transformaciones espectrales. (Se dedica el Capítulo 3 a la descripción y análisis de estos métodos).
- Integración para la recuperación de características de bajo nivel de los métodos evolución de la curva, representación de espacio tangencial, representación de color basada en histogramas,

elegidos debido a su eficiencia y fácil implementación para un sistema de recuperación visual (VIR) .

- Especificación del método de descripción, clasificación y asignación semántica de los objetos en las imágenes, basado en una ontología.
- Diseño de un modelo híbrido integrando características de bajo nivel, como color y forma, y de alto nivel, es decir la semántica de la imagen.
- Creación de una estructura ontológica que especifica y clasifica la semántica de las imágenes.
- Implementación, pruebas y corrección del modelo híbrido propuesto.
- Preparación del artículo sobre el modelo de indexación propuesto.

### **1.3 Actualidad en sistemas de recuperación de imágenes.**

En la actualidad existen una gran cantidad de máquinas de búsqueda de imágenes, comerciales y académicos. Por ejemplo, los sistemas Virage [1], VisualSEEK [2] y Query by Image Content (QBIC) [3] recuperan gráficos y video de colecciones en línea usando características visuales de bajo nivel de la imagen como color, textura y forma.

El sistema de recuperación Advanced Multimedia Oriented Retrieval Engine (AMORE) [4] y el sistema SQUID[5] proporcionan la recuperación de imágenes del Web utilizando consultas formadas por claves que especifican imágenes similares, bosquejos y predicados de SQL.

Para el usuario es difícil determinar cuales son las características de bajo nivel para establecer una consulta (recuperación de una imagen). En la actualidad los sistemas CBIR resuelven ese problema, proporcionando un ejemplo de imagen (mecanismo de retroalimentación – Query By Example, QBE), en lugar de especificar características de bajo nivel, para obtener imágenes similares, pero es necesario que el usuario tenga un buen ejemplo de lo que desea obtener como resultado previo al especificar dicha consulta. Nuevos sistemas de recuperación de información reemplazan las características de bajo nivel con características de alto nivel, conocidas como características semánticas, que son más entendibles para el usuario que las características de bajo nivel [6].

Los sistemas de recuperación de imágenes basados en contenido contribuyeron a la teoría de la recuperación de imágenes, de forma importante. Los CBIR actuales se basan en características de bajo nivel, como texturas, color, forma, localización, etc., las cuales no proporcionan una indexación de imágenes rápida y efectiva, ya que se requiere un gran número de iteraciones para obtener resultados que satisfagan las consultas de los usuarios. Estas características dan la ilusión al usuario que el sistema de recuperación puede entender el significado de la consulta, pero no

necesariamente reflejan el significado de una imagen (semántica de la imagen), como lo haría algún texto o anotación que llevaran las imágenes como menciona [7] en [8]. En el Capítulo 2 se muestra un análisis de sistemas comerciales existentes

### **1.3.1 Semántica en recuperación de imágenes**

En la actualidad no hay vocabularios o métodos aceptados para el análisis de imágenes de varios dominios. La semántica de una imagen define el significado que tiene ésta. Las formas de los objetos contenidos en una imagen son una característica importante del contenido semántico de la imagen y puede dar como resultado un índice en conjunto con alguna métrica de categorización estableciendo una anotación inteligente para su representación y su recuperación. El análisis y la recuperación de la semántica del contenido de una imagen es un problema difícil y ha sido una meta de la visión por computadora desde hace muchos años. [83]

La semántica de una imagen describe sus atributos tales como tipos de objetos, objetos individuales y eventos, acciones, emociones o incluso algún significado religioso. Puede tener varios niveles. Como se muestra a continuación:

- Semántica de tipo (fotografía de paisaje, imagen de la galería de imágenes, etc.)
- Composición de objeto (una bicicleta, un carro estacionado en la playa, una escena de la puesta de sol).
- Semántica abstracta (gente peleando, persona feliz, etc.).
- Semántica detallada (descripción detallada de una imagen). [6, 83].

En la actualidad para que un usuario encuentre en la Web la información que necesita, se hace uso de la nueva tecnología Semantic Web, que fue propuesta por Tim Berners-Lee, que desarrolla y utiliza lenguajes que son entendibles por la computadora, en lugar de por los humanos, en expresiones de contenido semántico en páginas Web [9]. Esta tecnología comprende varias herramientas y estándares para crear un medio universal de intercambio de información, asignando a los documentos (páginas Web) significado (semántica) entendible por la computadora. Para esta investigación de tesis, se tomarán los enfoques que abarcan las técnicas de componentes del Semantic Web que se refieren a objetos, sus relaciones y el vocabulario para describir propiedades y clases, con una jerarquía de semántica para dichas propiedades y clases. Para compartir este conocimiento (semántica) es clave utilizar una terminología como las ontologías que facilitan tanto la búsqueda y clasificación de recursos (imágenes), para lo cual se requiere identificar la información semántica de estos recursos [9].

### **1.3.2 Anotaciones ontológicas en recuperación de imágenes**

Las ontologías proveen significados para explicar conceptos y las relaciones entre ellos, describiendo una especificación formal de cierto dominio. De una manera formal una ontología es una especificación explícita de una conceptualización, (pieza de conocimiento). Para hacer esto, las ontologías hacen uso de la jerarquía de clases, dando propiedades adicionales es decir, con el uso de subclases. Dependiendo de su nivel de generalidad, cada tipo diferente de ontología satisface diferentes roles en el proceso de construcción de sistemas basados en conocimiento. Existen diferentes sistemas de recuperación de imágenes, que utilizan clasificación semántica de alto nivel, como el sistema Semantics-sensitive Integrates Mathching (SIMPLIcity) [10].

### **1.4 Motivación**

El principal interés de esta investigación es la exploración de métodos de recuperación semántica, incorporando mecanismos alternativos a la búsqueda para obtener mejores resultados. Como es la aplicación de la búsqueda semántica de objetos de una imagen, que permite encontrar imágenes de manera rápida y precisa, agregando previamente significado semántico a los objetos dentro de una imagen y apoyándose en las ventajas que ofrece la estructura de las ontologías, para describir e interpretar conceptos del dominio.

Los sistemas de hoy en día presentan algunos problemas de eficiencia, como los enfocados sólo a imágenes que muestren escenas geográficamente bien definidas; algunos otros se basan en color y no pueden relacionarse con el contenido de una imagen [30]; otros trabajos se basan en anotaciones, resultando muy costosos y tardados. Además otras investigaciones aplican codificación, pero debido al incremento de la cantidad de archivos imagen, se ven limitados por su vocabulario predefinido, siendo incompleto, burdo y ambiguo [23].

Las búsquedas del usuario (queries) se traducen en características de la imagen, las cuales se utilizarán como índices para el análisis de búsqueda y similitud con otras imágenes. La decisión de la similitud entre la imagen recuperada y la solicitud del usuario (query) se calcula a partir de la convergencia de las formas y la comparación de segmentos de imágenes con las respectivas anotaciones ontológicas de objetos de la imagen, obteniendo así, un mecanismo de entendimiento de semántica de imágenes [35].

### **1.5 Alcances y Contribuciones.**

A continuación se enumeran las restricciones y alcances que se definieron en el proceso de esta investigación:

- Análisis de los métodos probabilísticos, los métodos basados en color, los métodos jerárquicos con texto, Lógica difusa y los métodos híbridos para la recuperación de imágenes.
- Análisis de los métodos globales y estructurales para la representación de formas anteriormente definidos en los objetivos de esta tesis (Sección 1.2).
- El nivel de semántica con el que se trabajará en las imágenes es el nivel de descripción de objetos.
- Las colecciones de imágenes están restringidas en dominios específicos.
- El sistema VIR propuesto no trabaja en línea.
- El producto final es un prototipo y no un software comercial. El resultado final de este trabajo de investigación es un modelo híbrido de indexación de formas, es decir, que integre tanto características de bajo nivel (el color y forma), como características de alto nivel (significado de la imagen a nivel de objeto).
- A nivel conceptual, obtención de un método híbrido (basado en características de la imagen de bajo nivel y en descripciones semánticas) de indexación y recuperación de imagen que esté basada en contenido.
- A nivel práctico:
  - La realización de un análisis del desempeño de algunos métodos de sistemas VIR.
  - La creación de una estructura ontológica para almacenar la semántica y clasificar a las imágenes mediante la relación de los vectores de características con anotaciones ontológicas de las imágenes previamente definidas.
  - Implementación (software) del modelo propuesto.
  - Definición de la metodología sobre el uso de las técnicas de recuperación de información visual basado en contenido.

## 1.6 Estructura de la tesis

A continuación se define la estructura de este documento de tesis

- En el *Capítulo 1* se presenta una breve introducción de las búsquedas de imágenes que se hacen en la actualidad, se define el problema al que se enfoca esta tesis y se muestran los alcances y contribuciones del trabajo de investigación.
- En el *Capítulo 2* se definen conceptos de representación de imagen y se muestra un pequeño análisis de sistemas comerciales existentes, se explican los métodos en que se basa la

representación de imágenes y una breve descripción de algunos trabajos existentes que hacen uso de estos métodos.

- En el *Capítulo 3* se presenta una clasificación y el análisis de algunas técnicas de representación de formas, presentando su funcionamiento y algunas características de cada método.
- El *Capítulo 4* muestra el concepto de ontología, las propiedades, características y clasificaciones. También se presentan los pasos que se siguieron para su diseño, la definición del dominio y se presenta la estructura de la ontología que se diseñó y se utilizó en este trabajo.
- El *Capítulo 5* muestra el diseño de la implementación del modelo de indexación de imágenes basado en una ontología, se explican las técnicas que se eligieron y los procesos que se realizan para llevar a cabo la recuperación de las imágenes.
- El *Capítulo 6* muestra las pruebas, evaluaciones tanto del sistema como del modelo propuesto y los resultados obtenidos del sistema, se muestran ejemplos de algunas consultas. También se muestran gráficas de los resultados que obtuvo el sistema, comparando los resultados obtenidos al utilizar términos ontológicos y los obtenidos que se obtuvieron al no utilizar términos ontológicos.
- Por último el *Capítulo 7* muestra las conclusiones de ésta tesis, referente al sistema de recuperación implementado y al modelo de indexación propuesto; se presentan algunas posibles líneas como trabajo a futuro de éste proyecto.

## **1.7 Metodología.**

En la investigación primero se analizaron los métodos existentes que se utilizan en la extracción las características de una imagen y los métodos para interpretar el significado, después se seleccionaron los prototipos de los métodos más eficientes para estudiarlos, se realizaron pruebas de modelado en una aplicación para analizar ventajas y desventajas. Las pruebas realizadas nos llevaron al diseño de un modelo que se basa en las ventajas que presentan los métodos estudiados.

En base a esto, se implementó una aplicación que implementa este modelo propuesto, para después hacer las pruebas necesarias y la corrección del modelo propuesto y así obtener un modelo óptimo y eficaz. La contribución final de este trabajo es un método híbrido de recuperación de imágenes basándose en la integración de técnicas de extracción de características de bajo nivel y clasificación de la semántica.

## **1.8 Conclusiones**

El propósito de esta tesis es analizar mecanismos alternativos a la búsqueda de imágenes, a través del uso de ontologías que describen semánticamente el significado de una imagen. En este capítulo se planteó el problema a resolver, así como los objetivos de esta tesis. Se analizaron algunos métodos utilizados para la recuperación de imágenes. También se analizaron las características de algunos sistemas actuales de recuperación de imágenes y las técnicas que utilizan en la extracción de características así como en la comparación entre imágenes para su recuperación.

Los sistemas actuales recuperan imágenes basándose en características de bajo nivel, cada sistema tiene características particulares en su forma de realizar los procesos de comparación. Cada uno presenta una ventaja para un cierto tipo de imagen y un algún tipo especial de recuperación de información. Debido a que las ontologías pueden definir el significado de algún concepto, se diseñará una estructura ontológica para clasificar las características de alto nivel que interpreten el significado de una imagen, es decir, la semántica. Se decidió elegir el nivel de descripción de objetos como nivel de semántica, ya que los objetos se definen de manera menos ambigua.

Los permisos para poder publicar en este trabajo las interfaces de los sistemas comerciales revisados en éste capítulo, se encuentran en el Apéndice C.