Universidad de las Ciencias Informáticas

Facultad 2

Título: Línea base para la recuperación semántica de imágenes

Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas

Autor: Yilian Aguila Acuña

Tutores:

Dr.C Yanio Hernandez Heredia

Msc. Héctor Raúl González Diez.

Junio 2017

"(...) sin educación no hay Revolución posible, sin educación no hay socialismo posible, sin educación no hay ese hombre nuevo de que hablaba el Che (...)"

Declaración de Autoría

Declaramos ser autores de este trabajo de diploma y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los ____ días del mes _____ del año 2015.

_____

Yilian Aguila Acuña

_____ _____

Firma del tutor

Dr.C Yanio Hernandez Heredia

Firma del tutor

Msc. Héctor Raúl González Diez

Datos de contacto

Dr.C Yanio Hernandez Heredia: Ingeniero en Ciencias Informáticas, Dr en Ciencias Técnicas, Profesor Auxiliar.

Msc. Héctor Raúl González Diez: Lic. en Física Nuclear, Máster en informática aplicada, Profesor Asistente, Investigador en la línea de inteligencia artificial.

Agradecimientos

En primer lugar, agradecerle a mi país por permitirme entrar a la universidad y cumplir mis sueños de ser ingeniera y sentirme orgullosa de ser parte de la nueva generación de salvaguarda de la revolución.

A la universidad por enseñarme a superarme en la vida. Al colectivo de profesores que me brindaron sus conocimientos para poder alcanzar esta meta.

A mis dos tutores por brindarme su apoyo para alcanzar este título tan anhelado (Héctor gracias por soportarme todos los días a las 8:00 de la mañana en tu oficina, yo sé que no fue fácil, jaja; y Yanio gracias por atenderme y guiarme cuando lo necesité).

A mis padres por enseñarme, guiarme y sacrificarse para que hoy pueda decir que alcance una de mis metas. Gracias a mi papá por aconsejarme en cada momento de mi vida. A mi mamá por cuidarme y darme todo su amor. Espero que se sientan orgullosos.

A mis abuelas Cuca y Rosa: A pesar de que mi abuela Cuca no está conmigo fue y sigue siendo la luz que me guía, es esa personita que estaba ahí para mí a todo momento y que más amor me dio. A mi abuela Rosa por darme consejos y darme todo su amor.

Amias tías Sari, Agustina, Gladis, Yanet por darme sus consejos para la vida, por ayudarme cuando lo necesité, espero que se sientan orgullosas.

A mis hermanos Leonardo y Orlandito: Leonardo, por darme todo su cariño, es mi hermanito chiquito y lindo. Orlandito por ayudarme y quererme.

A mi nene lindo, Alejandro por estar a mi lado, por cuidarme, por darme su amor, por ayudarme, por estar en mis momentos malos y buenos, te amo.

A mis amigas lindas por soportarme, por ayudarme y por escucharme, Rosaibis, Mayara, Elisabet, Yisel y Elaine. A mis amigos bellos Manuel (Pluto), Yoan, Roniel, Yosamy. A mis padres aquí en la universidad Osmel y Beatriz, gracias por estar a mi lado.

A todas las amistades que hice en mi tránsito por la universidad, a mis entrenadores Ariel y Larrude por enseñarme que puedo conseguir todas las metas que me trace en la vida, a mis compañeras del equipo de fútbol y a mis compañeros del judo.

En general a todos los que de una u otra forma me ayudaron a cumplir mis metas.

Dedicatoria

Dedico este trabajo y mis estudios en especial a mis padres que siempre han estado a mi lado y por ser los pilares de mi vida, a uno de los más grandes amores de mi vida y por ser esa personita que más me quiso, mi abuela Cuca. Además, se la dedico a mi nene Alejandro por estar a mi lado y a toda mi familia, que de una forma u otra me guiaron en el camino para llegar aquí.

Resumen

El presente trabajo tiene como objetivo desarrollar una línea base para la recuperación semántica de imágenes, que permite extraer descriptores de imágenes, agrupar las características semejantes para crear estructuras de clasificación y recuperación.

Para la descripción de las imágenes se utilizó el descriptor SURF, obteniendo los puntos clave que se representaron posteriormente mediante el Bag of Words, creando el vocabulario por el cual se realiza la clasificación, utilizando las funciones Fitctree y Predict que tiene integrada el MATLAB. Para la validación se realizó la prueba del falso positivo y el falso negativo, mediante el método de validación cruzada. Además, se utilizó el método de selección estratificada para la división de la base de imágenes empleada. Con las pruebas que se le realizaron a la línea base se demuestra su correcto funcionamiento, retornando la clasificación probable que tiene cada una de las imágenes dadas.

Palabras claves

Clasificación, descriptor, imágenes, recuperación semántica.

ÍNDICE

INTRODUCCIÓN

“Una sola fotografía puede contener múltiples imágenes.” (Dando Moriyama), cuando se habla de una imagen lo primero que viene a la mente es un objeto o rostro de una persona, pero no siempre se tiene conocimiento de cuál es el concepto real de imagen, que no es más que una representación pictórica de un objeto o fenómeno que contiene información descriptiva de este (Comeche J. A., 2013).

Las imágenes a menudo son utilizadas para diversas actividades o motivos, sirven para adornar, enseñar, mostrar, etc…, a su vez pueden ser procesadas para sacar información, que en muchas ocasiones no se encuentran visibles a los ojos humanos, este procesamiento de imagen consiste en la manipulación de los datos contenidos en imágenes para convertirlos en información útil. El procesamiento de imágenes mediante medios de procesamiento digital de la información, tales como FPGA (Field Programmable Gate Array) aplicaciones específicas sobre DSP (Digital Signals Provessor), etc; constituye lo que se denomina Procesamiento Digital de Imágenes (PDI). (Gonzalez & Woods, 2002)

Con el desarrollo de nuevos sistemas y la creación de interfaces gráficas, el uso de las imágenes digitales se ha convertido en un elemento fundamental en la manipulación de información. El procesamiento de imágenes mejora la información pictórica para su posterior interpretación humana y la mejora en el procesado de las imágenes para su posterior almacenamiento, transmisión y representación de las mismas en máquinas automáticas de percepción. Por otra, parte estas imágenes proporcionan una gran capacidad de almacenamiento de información en espacio físico reducido. Por lo que se crean los sistemas de recuperación de imágenes para la representación y búsqueda de imágenes en forma digital.

La recuperación semántica de imágenes ha sido un problema estudiado en los últimos años dentro de la disciplina de reconocimiento de patrones. La misma cuenta con tres etapas esenciales, descripción de imágenes a través de características invariantes a las transformaciones a fines del espacio, representación de las imágenes en una estructura de clasificaciones a partir de las características comunes y recuperación haciendo uso de algoritmos de aprendizaje automáticos. La implementación de una línea base que sea tecnológicamente reutilizable, para la recuperación de imágenes, que tenga en cuenta estos tres elementos antes expuestos, es la base para el desarrollo de una nueva aplicación en el campo del procesamiento de imágenes y vídeos.

Una de las mayores ventajas de la recuperación semántica de imágenes es la capacidad de búsqueda de una imagen sin necesidad de proporcionar al sistema una descripción de la misma, pues se basa en características visuales que se extraen a partir de ciertos parámetros obtenidos directamente de la imagen, y otros tras haberla sometido a distintos tratamientos como trasformaciones, filtrado o máscaras. Estas características se comparan con las imágenes almacenadas previamente en la base de datos y posteriormente se puede aplicar un decisor, que podrá basarse en la media de distintos tipos de distancias, dependiendo de la finalidad del resultado y del tipo de imágenes que se estén tratando (Sarriá, 2010).

Otra ventaja de la recuperación semántica de imágenes, además de ser vital para el cumplimiento de tareas como la detección y recuperación de objetos y regiones de interés, es que puede reducir la brecha semántica a partir de que las mismas pueden ser utilizadas como una capa intermedia entre el usuario final y la computadora propiciando un mejor entendimiento entre ambos (Larin-Fonseca, Garea-Llano , & Chacón-Cabrera , 2012).

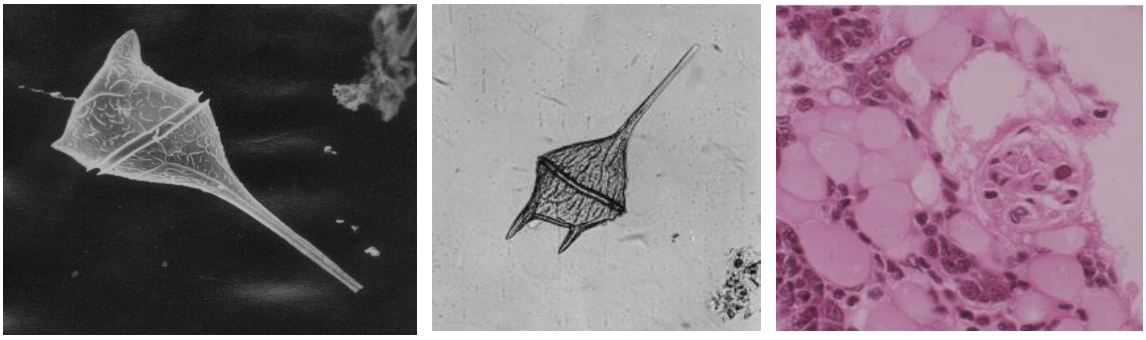


Ilustración 1 Microorganismos

Además, ha sido utilizada en diversas aplicaciones tanto industriales como en el quehacer científico. Esto ha permitido la identificación de algunas especies de fitoplancton, zooplancton, bacterias (tuberculosis y Vibrio cholerae), cuerpos de inclusión de virus en imágenes de tejido de camarón (IHHN y WSSV), cromosomas del abulón rojo, azul y amarillo y parásitos de peces. Otra aplicación es el estudio de superficies marinas reales a partir de imágenes remotas (Álvarez J. , 2017).

Otra de las ramas que emplea el tratamiento de imágenes es la medicina, para caracterizar la anatomía y las funciones del cuerpo humano, empleándola en aplicaciones para el estudio cerebral y cualquier región anatómica de la que se pueda obtener una imagen (R., 2004).

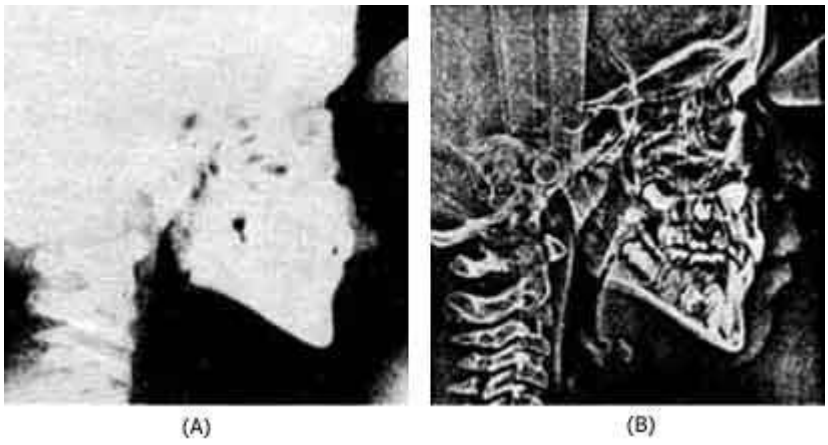


Ilustración 2 Imagen médica

Como muchos países, Cuba cuenta con un gran grupo de instituciones inmersas en el desarrollo de Software que trabajan en el campo del procesamiento de imágenes. Dentro de estas instituciones se encuentra la Universidad de Ciencias Informáticas (UCI), la cual centra sus esfuerzos en la formación de nuevos profesionales y a la vez desarrollar nuevos sistemas que sirvan de apoyo e informatización de las actividades empresariales.

La UCI cuenta con diversos proyectos y centros enfocados al desarrollo de software para diferentes ramas, entre ellos se encuentra el Centro de Informática Médica (CESIM), que es un centro dedicado al desarrollo de productos, sistemas, servicios y soluciones de alta calidad y competitividad para la optimización del trabajo y mejoramiento de la calidad de la atención médica, una de sus áreas se dedica al procesamiento de imágenes digitales, pero solo realiza el procesamiento de imágenes médicas.

Por otra parte, se encuentra El grupo de Procesamiento Digital de Señales y Geoinformación tiene como objetivo crear un espacio orientado a las investigaciones teóricas y aplicadas, que garanticen las inserciones de valor agregado en los sistemas desarrollados principalmente por el centro de desarrollo GEySED. Además de fomentar la integración de las líneas del centro e intencional el trabajo científico y la visibilidad del mismo, este grupo, entre sus temas de asociados se encuentran las técnicas de procesamiento de imágenes que incluye técnicas para el mejoramiento de la calidad de imágenes, así como técnicas de análisis, descripción y clasificación de imágenes. Este grupo, a pesar de realizar un procesamiento de imágenes más profundo, no trabaja la recuperación semántica de imágenes.

Por lo que una de las necesidades existentes hoy en la UCI es que no se cuenta con una línea base que realice una integración de los algoritmos básicos para la recuperación semántica de imágenes capaz de describir imágenes a través de características invariantes a las transformaciones a fines del espacio. Representar las imágenes en una estructura de clasificaciones a partir de las características comunes y recuperar haciendo uso de algoritmos de aprendizaje automáticos. Por lo que se hace difícil crear la base para el desarrollo de nuevas aplicaciones en el campo del procesamiento de imágenes y vídeos que aprovechen las ventajas que brinda el procesamiento de imágenes basado en la recuperación semántica de imágenes.

Teniendo en cuenta la situación problemática ante expuesta se tiene como **problema a resolve**r: ¿Cómo extraer descriptores de imágenes y agrupar las características semejantes para crear estructuras de clasificación y recuperación semántica de imágenes automática?

Se tiene como **objeto de estudio** la recuperación semántica de imágenes.

El **objetivo general** del trabajo es implementar una línea base que permita extraer descriptores de imágenes, agrupar las características semejantes para crear estructuras de clasificación y recuperación.
El objetivo general estará enmarcado en el **campo de acción:** la clasificación automática de imágenes.

Como **objetivos específicos** se plantea:

1. Caracterizar el marco teórico-conceptual de los Algoritmos de extracción de descriptores de imágenes y de las técnicas de aprendizaje automático.
2. Implementar un componente en MATLAB, para la Recuperación semántica de imágenes, utilizando los algoritmos seleccionados del estado del arte
3. Validar la solución implementada mediante el uso de base de datos internacionales disponibles para este tipo de soluciones.

Las **tareas** que darán cumplimiento a los objetivos planteados son:

1. Selección de los descriptores de representación imágenes de mejores resultados a nivel internacional.
2. Selección de los modelos de representación y agrupamiento de imágenes.
3. Selección de los modelos de clasificación y recuperación de imágenes.
4. Definir los requisitos y etapas del componente propuesto.

5. Seleccionar las bases de datos para evaluar la propuesta.
6. Seleccionar las métricas para evaluar la propuesta.
7. Validar el correcto funcionamiento del componente.
8. Diseñar y ejecutar un paquete de experimentos para evaluar la propuesta.
9. Obtener el documento final de tesis.

Métodos científicos

Métodos Teóricos

Método Análisis-histórico-lógico: Permite estudiar las diferentes etapas de la recuperación de imágenes para llegar al objetivo a través de la investigación de las leyes generales del funcionamiento y desarrollo del fenómeno y el estudio de su esencia.

Método analítico-Sintáctico: A través del análisis y estudio de los algoritmos utilizados para una correcta recuperación semántica de imágenes se pretende llegar a una solución óptima para el problema planteado.

Métodos empíricos

Método de Observación: Aprecia y comprende cómo se desarrolla el proceso de recuperación semántica de imágenes para poder definir y distinguir lo que se quiere hacer.

Método experimental: Implica la observación, manipulación, registro de las variables que se utilizaran el prototipo a implementar y así comprobar la idea a defender y validar su resultado.

Resultados esperados: el desarrollo del trabajo permitirá la integración de tres métodos para crear una línea base para el desarrollo de nuevas aplicaciones que empleen la recuperación semántica de imágenes.

Descripción de los capítulos.

Capítulo 1: Algoritmos para la recuperación semántica de imágenes, aborda los conceptos básicos del tema a desarrollar, así como el análisis de los posibles métodos a utilizar en la solución.

Capítulo 2: Propuesta de la solución, se expone la propuesta de solución, los algoritmos utilizados, se describe la herramienta y la metodología utilizada para el desarrollo del trabajo.

Capítulo 3: PRUEBAS, contiene la descripción de las pruebas realizadas a la solución, así como los resultados obtenidos, además de la verificación del correcto funcionamiento.

CAPÍTULO 1: ALGORITMOS PARA LA RECUPERACIÓN SEMÁNTICA DE IMÁGENES.

1.1 Recuperación semántica de imágenes.

La recuperación de información es el área del conocimiento que estudia los sistemas automatizados, informando al usuario de la existencia de documentos relacionados con la consulta realizada, es decir, trata esencialmente documentos en forma digital, de manera que se puedan ser representados y recuperados mediante algoritmos automáticos. Dentro de esta área se encuentra la rama dedicada al estudio e investigación de la documentación visual, surgida en la década de 1990, denominada Recuperación de imágenes.

La expresión Recuperación de imágenes fue utilizada por primera vez por Toshikazu Kato en 1992, denominándola Recuperación de Imagen Basada en Contenido ((Content-Based Image Retrieval o CBIR), considerada como el proceso automático de representación y búsqueda de imágenes en forma digital. (Álvarez S. P., 2007)

1.2 Etapas de la recuperación semántica de imágenes.

La recuperación de imagen cuenta de tres etapas fundamentales:

- La primera etapa de los Sistemas de Recuperación de imagen se basa en las representaciones textuales de las características de las imágenes.
- La segunda etapa se basa en los rasgos visuales de las imágenes.
- La tercera etapa emplea simultáneamente el código visual y el código textual para representar y recuperar imágenes.

La primera etapa está comprendida desde los inicios de la incorporación de imágenes como unidad de descripción documental a las colecciones digitales hasta la década de 1990. En esta etapa se aplica las mismas técnicas que se empleaban con documentos textuales, como resultado se importa todo texto que acompañe a la imagen para extraer automáticamente los términos de indización que representarán el contenido de la misma. En la actualidad se utiliza este enfoque donde la cantidad inherente de imágenes incorporadas a la colección de manera constante hace impensable cualquier otro procedimiento, como por ejemplo los motores de búsqueda generalistas Google y Bing. (Feng, Brussee, Blanken, & Veenstra, 2007)

Los principales problemas que presenta esta primera etapa son:

- La dependencia de analistas humanos: Este enfoque depende de descriptores manuales de la colección mediante la adaptación de normas de representación pre-existentes y con la ayuda de vocabularios controlados y otras herramientas documentales.
- La inconsistencia de la descripción entre analistas.
- El volumen de la documentación.
- Dificultad de descripción de las propiedades perceptuales o de bajo nivel (color, forma...) mediante el código lingüístico.

Para darle solución a estos problemas se desarrolla una segunda etapa a finales de la de cada de 1990, denominada Recuperación de imagen basada en contenido (Content-Based Image Retrieval o CBIR). El principal aporte de este enfoque es que ahora la imagen se describe mediante sus propiedades perceptuales, a través de las características psicológicas percibidas por el ojo humano, principalmente el color, la textura, la forma o las relaciones especiales.

La recuperación de imagen parte de un modelo estándar que permite representar cualquier color mediante tres números, que representan los colores primarios: rojo, verde y azul, donde cada pixel de la imagen posee un color y puede ser descrito mediante una triada de números. La textura indica una región de la imagen que presenta un patrón tronco del árbol, ramas, hojas… que no posee un único color, pero que se repite hasta ser percibido por el ojo humano como una propiedad de la imagen. (GONZALEZ & WOODS, 2008, pp. 706-712)

Para representar numéricamente las formas presentes en una imagen se utilizan dos números por cada pixel de la imagen correspondientes al nivel de variación de la intensidad en dichos pixeles con respecto a los ejes de abscisas y ordenadas. A estos dos números se le denominan valores de borde y tienen propiedades geométricas de señalar la dirección del máximo cambio de intensidad en dicho pixel y la magnitud del cambio en dicha dirección (GONZALEZ & WOODS, 2008). Una vez detectadas ciertas regiones en la imagen, se utilizan varios métodos para representar su posición relativa entre los que se destaca las cadenas 2D, donde una primera cadena barre las imágenes de arriba abajo indicando por orden las regiones que se van encontrando, y una segunda cadena barre la imagen de izquierda a derecha indicando, igualmente por orden, las regiones que se van topando. (Mosquera González, Carriera Nouche, & Hónzalez Penedo, 2011, pp. 571-572)

Los sistemas CBIR permiten realizar consultas en las que el usuario escoge el color o colores más destacados de la imagen que busca, además puede señalar las formas más sobresalientes de la imagen y pueden definir regiones completándolas con colores sólidos y en la localización espacial deseada. A pesar de resolver muchos de los problemas de la segunda etapa, los sistemas CBIR también presentan una serie de problemas entre los que destacan:

- Las limitaciones formales que abarcan los problemas al nivel de repetición de las propiedades o atributos visuales.
- Las diferencias numéricas entre dos colores en ocasiones no se corresponden con la diferencia percibida por el ser humano en relación a tales colores.
- Se pueden originar repeticiones numéricas muy distintas por formas percibida como únicas por los seres humanos.
- Las dificultades a la hora de efectuar las consultas por parte del usuario, en última instancia motivada por el vacío semántico traen consigo las limitaciones semánticas.
- La subjetividad humana inherente a la interpretación de imágenes.

Para darle solución a estos problemas surge recientemente una tercera etapa denominada Recuperación de Información Visual Basada en la Semántica (SBVIR). Este enfoque es caracterizado fundamentalmente por la confluencia de la descripción del contenido visual y la descripción lingüística en una imagen. Los sistemas de recuperación de imagen de tercera generación emplean la notación de imágenes mediante descriptores textuales para representar propiedades extrínsecas de carácter subjetivo o semántico. Esta etapa limita el código lingüístico a aquellos niveles en los que todavía no es posible manejarse con éxito mediante un código puramente visual, empleando el conocimiento humano en dos manifestaciones principales: las etiquetas asignadas por usuarios humanos y la retroalimentación por relevancia durante la recuperación. (Comeche J. A., 2013)

1.3 Principales módulos para la recuperación de imágenes.

En forma general la recuperación de imágenes o los sistemas que dan cumplimiento al objetivo de recuperación cuentan con tres etapas o módulos fundamentales:

- Módulo de Descripción: es el encargado de representar numéricamente las propiedades o cualidades de la imagen. Las propiedades se pueden clasificar en dos clases (Comeche J. A., 2013):
    - Propiedades intrínsecas de la imagen: conjunto de rasgos visuales que caracterizan toda imagen como el color, la textura, la forma y las relaciones especiales. Estas propiedades también son conocidas como propiedades de bajo nivel (Comeche J. A., 2013).
    - Propiedades extrínsecas de la imagen: son todos los elementos no propiamente visuales, estas propiedades se dividen en dos subdivisiones (MUller, Clough, Deselaers, & Caputo, 2010):
        - Propiedades de nivel medio: engloba la determinación automática de límites, contornos, objetos y conceptos extraídos de la imagen en su integridad.
        - Propiedades de nivel alto: engloba los elementos objetivos que se incluyen en los metadatos o a propiedades de carácter subjetivo extraídas a raíz de la contemplación de la imagen y que suelen incorporarse en el apartado "Descripción" o "notas" en los metadatos.

- Módulo de Consultas: Le da la opción al usuario de introducir la consulta o expresar su necesidad informativa. La consulta se puede introducir mediante varios métodos que juegan un papel importante en la disminución de la brecha semántica en sistema de recuperación de imágenes, hay cuatro tipos de consultas: (B., 2010)

    - Consulta con imagen de ejemplo (Query by image example): el usuario provee una imagen o selecciona las palabras claves o conceptos que describen su contenido o el de una región de interés (ROI sus siglas en ingles), que para su criterio es representativa de lo que busca. Este tipo de consulta tiene dos desventajas fundamentales, una es que, en la mayoría de los casos, es difícil para los usuarios proveer una imagen apropiada para inicial la búsqueda. La otra desventaja es la dificultad de capturar la intención del usuario, es decir, que el sistema pueda identificar las características que lo llevaron a escogerla como ejemplo.

    - Consulta por características de la imagen (Query by image feaure): el usuario provee las características visuales representativas de la imagen que busca. Se puede realizar mediante bocetos en blanco y negro, objetos complejos de imágenes en color o mediante bocetos en tres dimensiones, especificando la consulta con una foto del ambiente creado, capturada desde algún punto de vista específico, de manera que el boceo incluya la relación espacial de los objetos.

    - Consulta con texto (Query by text): el usuario especifica su búsqueda mediante palabras claves o lenguaje natural. Este tipo de consulta también puede ser restringida con sintaxis o vocabularios definidos en una antología, permitiéndole al usuario recorrer la antología para seleccionar los términos que va a incluir en la consulta o digitalizar la consulta y verificar que los términos estañen la ontología.

    - Consultas híbridas: frecuentemente se usan en sistemas como Paragrab, que utiliza técnicas de procesamiento de lenguaje natural y refinamiento, además combina metadatos de consultas, uno basado en texto y reconocimiento del habla y el otro basado en características visuales. SPIRS (Spine Pathology and Image Retrieval in Meical Applications) combina los tres tipos de consulta y el proyecto IRMA (Image Retrieval in Medical Application) implementa consultas híbridas con bosquejos e imágenes de ejemplo.

- Módulo de búsqueda: se encarga de la extracción automática de las imágenes más relevantes existentes en la colección en relación a cada consulta y su ordenación por orden de relevancia. Este proceso se realiza mediante un algoritmo de similaridad. (Comeche J. A., 2013)

1.4 Recuperación de imagen basada en contenido (Content-Based Image Retrieval o CBIR)

La solución del problema estará basada en la segunda etapa de la Recuperación de imágenes definida anteriormente como los sistemas CBIR, que cuenta con una arquitectura formada por cuatro partes fundamentales, descripción, clasificación y representación.

Primeramente, se realiza la Descripción, proceso mediante el cual se procesan todas las características de cada una de las imágenes para cuantificar todas las que sean capaces de diferenciarla de otras imágenes. Posteriormente se realiza la representación para transformar la información visual en un espacio vectorial para su posterior clasificación. Por último, se realiza la Clasificación para realizar la comparación entre dos histogramas, mediante el cálculo de su diferencia, ya que cada imagen tiene asociada una serie de histogramas que contienen su información y para realizar la comparación entre dos histogramas se necesita calcular su diferencia mediante métricas

1.4.1 Descriptores de imágenes.

La información de las imágenes digitales se encuentra codificada mediante la magnitud de cada uno de los pixeles, que representan el nexo de la unión entre el contenido abstracto de sus valores y las características propias de una imagen. Debido a la necesidad de extraer la descripción de la información de las imágenes, se crean los descriptores de imágenes, teniendo en cuenta las propiedades siguientes (García Ó. B., 2011):

- *Simplicidad:* con el objetivo de permitir una fácil interpretación del contenido de las características de la imagen, el descriptor de imágenes debe representar dichas características de manera clara y sencilla.
- *Repetibilidad*: cuando se genera un descriptor a partir de una imagen, este debe ser independiente del momento en que se genera.

- *Diferenciabilidad*: Dada una imagen, el descriptor generado debe poseer alto grado de discriminación respecto de otras imágenes y al mismo tiempo contener información que permita establecer una relación entre imágenes similares.
- *Invarianza*: los descriptores que representen dos imágenes con deformaciones en su representación, deben aportar la robustez necesaria para poder relacionarlas aún bajo diferentes transformaciones.
- *Eficiencia:* los recursos utilizados para generar el descriptor deben ser aceptables para que sean utilizados en aplicaciones con restricciones críticas de espacio y/o tiempo.

Clasificación de los descriptores de imagen.

Los descriptores de imagen se clasifican según el nivel de abstracción de la representación. En el nivel bajo se encuentran los descriptores visuales que describen características elementales como color, forma o textura y en el nivel superior los descriptores son más específicos pues aportan información sobre los objetos de la imagen. Los descriptores se clasifican en dos grande grupos (García Ó. B., 2011):

- Descriptores de información general: son los descriptores de bajo nivel, que describen a la imagen según el color, forma, textura, regiones y movimientos.
- Descriptores de información de dominio específico: proporcionan información acerca de los objetos y eventos que constituyen la imagen. A estos descriptores también se le conocen como descriptores semánticos y lo que hacen es utilizar los descriptores de bajo nivel para cubrir el “gap” existente entre las características visuales disponibles y las diferentes categorías semánticas (García Ó. B., 2011).

Los descriptores de información global se clasifican según la región de la imagen sobre la cual realizan las operaciones para general los resultados que componen el descriptor:

- Descriptores globales: El contenido de la imagen es resumido en un único vector o matriz. Encapsula una gran cantidad de información de la imagen en una pequeña cantidad de datos para describirla. Este tipo de descriptor ha sido utilizado para diferentes tareas, a pesar de su simplicidad, debido a su bajo costo computacional y a unas prestaciones relativamente buenas (García Ó. B., 2011).
- Descriptores locales: Actúa sobre regiones de interés previamente calculadas o identificadas, construyendo un vector o matriz de características de esa región, donde se recoge información tanto del punto de interés como de la región adyacente al mismo o vecindario (García Ó. B., 2011).

Para la caracterización del contenido de imágenes de forma automatizada se han desarrollado múltiples descriptores visuales, muchos pertenecen a estándares como el MPEG-7 desarrollado por MPEG (Motion Picture Expert Group), que reúne una colección de descriptores visuales aplicables para su implementación en tareas de recuperación de contenido multimedia, comparación y clasificación de imágenes o realización de resúmenes de vídeo. Otros descriptores utilizados para diferentes tareas de tratamiento de imágenes, no pertenecen a ningún estándar, pero tienen contribuyen al desarrollo de nuevas técnicas y nuevos descriptores (García Ó. B., 2011).

Los descriptores globales, en comparación con los descriptores locales, tienen una gran desventaja, pues no son capaces de manejar las transformaciones a fines del espacio. Por lo que la investigación se centró en los descriptores locales.

Descriptores locales.

- **Scale Invariant Facture Transform (SIFT).**

El algoritmo Scale Facture Transform (SIFT) fue creado por Lowe, es un algoritmo capaz de detectar puntos característicos estables en una imagen, donde los puntos son invariantes frente a diferentes transformaciones como traslación, escala, rotación, iluminación y transformaciones afines. Este algoritmo realiza la correspondencia entre puntos basada en los vectores de características de cada punto que componen el descriptor de la imagen. Según la implementación de Lowe, el algoritmo SIFT cuenta principalmente con cuatro etapas (Mikolajczyk & Schmid. , 2005):

1. Detección de Extremos en el Espacio Escala: Se realiza una búsqueda sobre las diferentes escalas y dimensiones de la imagen para la identificación de posibles puntos de interés que son invariantes a los cambios de orientación y escalado.Este proceso se realiza mediante las funciones DoG (Difference-of-Gaussian).

2. Localización de los Puntos Clave: se aplica una medida de estabilidad sobre los puntos de interés para destacar aquellos que no sean adecuados y filtrar los puntos claves, también llamados puntos de interés.

3. Asignación de la Orientación: A cada punto de interés extraídos en la imagen de asignan una o más orientaciones, basándose en las direcciones locales presentes en la imagen gradiente. Posteriormente, todas las operaciones se realizan sobre los datos transformados según la orientación, escala y localización dentro de la imagen asignada en esta etapa, proporcionando así la invarianza respecto de estas transformaciones.

4. Descriptor del Punto de Interés: Hace referencia a la representación de los puntos clave como una medida de los gradientes locales de la imagen en las proximidades de dichos puntos claves y respecto de una determinada escala. Cada punto de interés corresponde a un vector de características compuesto por 128 elementos, que le confiere una invarianza parcial a deformaciones de forma, así como a cambios de iluminación.

La comparación entre objetos pertenecientes a dos imágenes diferentes se lleva a cabo mediante la comparación de los puntos de interés, de ahí la importancia de la estabilidad de dichos puntos de interés. Brown y Lowe proponen una función 3D para eliminar aquellos puntos que se encuentren en bordes o que presenten bajo contraste, ya que son más susceptibles al ruido, y así asegurar dicha estabilidad (Mikolajczyk & Schmid. , 2005).

Detección de Extremos en el Espacio Escala:

El descriptor SIFT es construido a partir del espacio Gaussiano de la imagen original, en el cual se pueden detectar de manera efectiva las posiciones de los puntos claves, invariantes a cambios de escala de la imagen. En esta primera etapa se tiene como objetivo obtener los puntos candidatos de la imagen que pueden ser identificados de forma repetida bajo diferentes vistas del mismo objeto (Mikolajczyk & Schmid. , 2005).

El espacio-escala Gaussiano de una imagen L(x, y, σ) es definido como la convolución de funciones 2D Gaussianas G(x, y, σ) de diferentes valores *o* con la imagen original I(x, y):

$$L(x, y, \sigma) = \quad G(x, y, \sigma) \quad * \quad I(x, y)$$

Siendo (x, y) las coordenadas especiales y *o* el factor de escala.

El algoritmo utiliza la función DoG (Diferencia Gaussiana) que se forma a partir de la derivada escalar de la Gaussiana escalada especialmente. Esta función DoG D(x, y, σ) se obtiene mediante la sustitución de escalas posteriores en cada octava (Mikolajczyk & Schmid. , 2005):

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

Donde k es una constante multiplicativa del factor de escala. La función DoG es utilizada por su eficiencia en cuanto a costo computacional, las imágenes suavizadas L(x, y, σ) son calculadas para la descripción de características en el espacio-escala, y por lo tanto, puede obtenerse como una simple resta. Además, Mikolajczyk asegura que los máximos y mínimos del Laplaciano de la Gaussiana respecto de una escala normalizada produce las características de imágenes más estables en comparación con otras funciones como el Gradiente, el Hessiano o el Harris Corner Detector, pudiéndose aproximar el Laplaciano de la Gausseana de escala normalizada mediante la función DoG (Mikolajczyk & Schmid. , 2005).

Una octava es el conjunto de imágenes Gausseanas suavizadas junto con las imágenes DoG y el conjunto de las octavas es construido mediante el muestreo sucesivo de la imagen original por un factor 2. Cada octava es dividida en números enteros de sub-niveles o escalas s. Luego de analizar una octava completa, se obtiene la primera imagen de la octava siguiente mediante el muestreo de la primera de las imágenes de la octava predecesora con un valor de θ del doble respecto a la actual (Mikolajczyk & Schmid. , 2005)l.

En diferentes niveles de escala, el espacio-escala L(x,y, σ), representa la misma información a diferentes niveles de la escala, permitiendo una reducción de la redundancia, por lo que se producen s+3 imágenes por cada una de las octavas y por lo tanto s+2 DoG imágenes por cada búsqueda de extremos. De acuerdo con los resultados de Lowe, es el valor de s=3 el que mejor resultados consigue.

Localización de puntos claves estables.

En esta segunda etapa se realiza un estudio de la estabilidad de cada uno de los puntos claves candidatos que han sido calculados. Para extraer los puntos que no están firmemente situado o que están sobre bordes o aquellos con bajo constaste, Lowe utiliza los criterios siguientes, ya que estos puntos son bastante vulnerables al ruido y por lo tanto no podrán ser detectados bajo pequeños cambios de iluminación o variación del punto de vista de la imagen (Mikolajczyk & Schmid. , 2005).

- Los puntos con bajo contraste son excluidos de la siguiente etapa mediante un proceso de umbralización por el cual se realiza una comparación y todos los puntos, cuyo valor sea menor que dicho umbral D, son expulsados.
- Los puntos situados sobre bordes de manera difusa son eliminados mediante la propiedad de la función DoG atendiendo a la gran curvatura que presenta en la

dirección paralela al borde y la pequeña curvatura que se observa en la dirección perpendicular. Mediante la matriz Hessiano sobre la localización y escala del punto medio:

$$H = \begin{bmatrix} \frac{\partial^2 D}{\partial x^2} & \frac{\partial^2 D}{\partial x \partial y} \\ \frac{\partial^2 D}{\partial x \partial y} & \frac{\partial^2 D}{\partial y^2} \end{bmatrix}$$

Donde D es la imagen DoG(x, y,θ) respecto de la escala s. Las derivadas se calculan mediante la resta del valor de los puntos vecinos. Mediante la desigualdad siguiente se localizan los puntos en los bordes por lo que aquellos que no satisfagan dicha desigualdad serán descartados debido a su inestabilidad.

$$\frac{\left(\frac{\partial^2 D}{\partial x^2} + \frac{\partial^2 D}{\partial y^2}\right)^2}{\left(\frac{\partial^2 D}{\partial x^2} * \frac{\partial^2 D}{\partial y^2}\right) - \left(\frac{\partial^2 D}{\partial x \partial y}\right)^2} \quad \frac{(\boldsymbol{r}+\mathbf{1})^2}{\boldsymbol{r}}$$

Tras descartar los puntos inestables, al resto de los puntos claves se les asignan una orientación

Asignación de la orientación.

La asignación, de una orientación basada en las propiedades locales de la imagen y representando el descriptor respecto de esta orientación, a cada uno de los puntos, se realiza para garantizar la invarianza respecto a la rotación. Para cada uno de los puntos de interés se calcula la magnitud del gradiente, m, y su orientación θ, mediante la siguiente ecuación:

$$m(x,y) = \sqrt{L(x+1,y) - L(x-1,y)^2 + L(x,y+1) - L(x,y-1)^2}$$

$$\theta(x,y) = arctan \frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}$$

Donde L representa la imagen gaussiana suavizada cuya escala resulta más próxima a la escala del punto de interés actual (Mikolajczyk & Schmid. , 2005).

Posteriormente se crea un histograma con 36 bins, cada uno de ellos con una longitud de 10◦ para cubrir el rango de los 360◦ posibles. La dirección dominante del gradiente se corresponde con el bins cuyo valor es el más alto y por lo tanto es el escogido como orientación dominante, aunque se puede dar el caso donde haya más de una dirección dominante. Por lo que, si hay un bins con un valor por encima del 80% del valor, también es considerado como dirección dominante. Los puntos con más de una dirección

dominante supondrán una mayor estabilidad al mismo. Para precisar se utiliza una parábola para, mediante la interpolación de los tres valores más altos del histograma, obtener el valor del pico.

Las principales orientaciones del histograma se asignan al punto de interés para que así el descriptor pueda ser representado respecto de estos.

Descriptor del punto de interés.

En la última etapa se crea un vector de características para cada uno de los puntos de interés que contiene una estadística local de las orientaciones del gradiente de la escala de espacio gaussiano. Alrededor del punto de interés, se realiza un muestreo de las orientaciones y magnitudes del gradiente de la imagen sobre regiones de 16X16. Cada una de las muestras son ponderadas, tanto por la magnitud de su gradiente, como por una función 3D gaussiana evitando cambios bruscos en el descriptor ante pequeños cambios en la posición de la ventana y al mismo tiempo asignado menor énfasis a los puntos más alejados del punto de interés (Mikolajczyk & Schmid. , 2005).

El contenido de los histogramas de orientaciones formado de las muestras de cada región de 16X16, son resumidos en sub-regiones de 4X4. Donde cada uno de los histogramas se compone de 8 bins, que almacenan las orientaciones posibles proporcionales a 45◦, donde la magnitud de cada flecha representa al valor acumulado para cada bin. Para cada uno de los puntos de interés, se obtienen 16 histogramas respecto de las orientaciones de los puntos de cada región (Mikolajczyk & Schmid. , 2005).

El descriptor de cada punto de interés, finalmente, está formado por un vector que contiene los valores de las 8 orientaciones de los 4X4 histogramas componiendo un vector de 4X4X8=128 elementos.

Correspondencia entre puntos clave (matching)

El termino matching se refiere al cálculo de un valor que represente el grado de similitud entre las dos imágenes. El cálculo de este valor se realiza mediante la aplicación de una métrica o fórmula de la distancia entre ambas imágenes, este valor es conocido también como score. Además, es necesario establecer las correspondencias entre los puntos de interés, esta correspondencia se lleva a cabo mediante el cálculo de la distancia euclidea entre los vectores de características pertenecientes a diferentes puntos de interés, generando, a su vez, otro valor que será utilizado para determinar cuál de los

puntos de la imagen comparada corresponde con su homólogo, en el caso de existir, de la primera de las imágenes (Mikolajczyk & Schmid. , 2005).

Ilustración 3Diagrama de bloque del descriptor SIFT.

El diagrama de bloque representa y hace un resumen de las etapas de funcionamiento del proceso de comparación entre dos imágines y el cálculo del score o punto entre las mismas.

- **Speeded Up Robuts Feature (SURF).**

Speeded Up Robuts Feature (SURF) fue desarrollado por Herbert Bay como un detector de puntos de interés y descriptor robustos. El SURF a pesar de ser similar al SIFT en cuanto a la filosofía, presenta una gran diferencia y principalmente dos mejoras expuestas a continuación:

- La velocidad de cálculo es considerablemente superior sin ocasionar perdida del rendimiento.
- Tiene una mayor robustez ante posibles transformaciones de la imagen.

El SURF, para conseguir estas mejoras, reduce la dimensionalidad y complejidad del cálculo de los vectores de características de los puntos de interés obtenidos, mientras continúan siendo suficiente características e igualmente repetitivos (Mikolajczyk & Schmid. , 2005).

Con respecto al descriptor SIFT, las principales diferencias del SURF son:

- La longitud o nacionalización de los vectores de los puntos de interés se reducen a la mitad de la longitud del descriptor SIFT, con una dimensión de 64
- El SURF siempre utiliza la imagen original.
- Para el cálculo de la posición y la escala de los puntos de interés utiliza el determinante de la matriz Hessiana.

HISTOGRAMAS DE GRADIENTES ORIENTADOS (HOG). (Cruz, 2013)

El descriptor HOG se utiliza para la búsqueda de objetos en una imagen ya que es capaz de detectar la presencia de patrones en una escena, describiendo por medio de la distribución de los gradientes la forma de un objeto en una imagen.

Este descriptor calcula los gradientes dividiendo la imagen en una serie de bloques distribuidos a lo largo y ancho de la misma y con cierto solapamiento entre ellos. El avance de los bloques se realiza mediante la eliminación de la columna de las celdas de la izquierda y añadiendo la columna de la derecha para el desplazamiento horizontal, en caso del vertical, se elimina la fila de las celdas de arriba, añadiéndola al final de la celda de abajo. Cada bloque es dividido en sub-regiones o celdas para calcular en cada uno de ellos el histograma de los gradientes orientados de tal forma que se logra mejorar el resultado. (García R. H., 2014)

Para aplicar HOG a un área determinada, primeramente, se calculan las derivadas especiales de dicha imagen a lo largo de los ejes x e y (Ix y Iy). El siguiente paso es dividir la imagen en celdas para calcular los histogramas de cada una de estas celdas. Para cada pixel de la celda se tiene un cierto peso en el histograma de orientación, basado en el valor calculado de la magnitud de su gradiente. A través del histograma de cada una de las celdas se hace la representación de la imagen. Los histogramas quedan ordenados en un vector según su peso, conformándose el vector de características de la imagen. Una imagen omnidireccional contiene los mismos píxeles en una fila, aunque la imagen esté rotada, lo que significa que se va a obtener un vector de características invariante ante una rotación.

1.4.2 Representación

La representación de imágenes permite transformar la información visual en un espacio vectorial para su posterior clasificación. Las mayorías de las representaciones locales presentan una limitación común, dada porque no tienen en cuenta la disposición estructural a partir de las relaciones espacio-temporales entre los descriptores. La disponibilidad estructural de las características aporta una información adicional a la representación. Además, resulta una forma de hacer frente a variaciones de la imagen debido al ruido, oclusiones parciales y cambios de perspectivas. De esta manera es posible obtener una representación más general de la imagen a partir de las características locales. Son varios los modelos existentes para la representación de imágenes, por lo que la investigación se centrara en tres de estos modelos, el IF-IDF, el Bag of Word y el n gramas (García R. H., 2014).

- **Bag of Words (BoW)** (Gardcía, 2014)

Bag of Words (BoW) es un algoritmo que cuenta cuántas veces aparece una palabra en un documento. Estos recuentos de palabras nos permiten comparar documentos y evaluar sus similitudes para aplicaciones como búsqueda, clasificación de documentos y modelado de temas. BoW es un método para preparar el texto para la entrada en una red de aprendizaje profundo. En la visión por computadora, el modelo de bolsa de palabras (modelo de BoW) se puede aplicar a la clasificación de imagen, tratando las características de la imagen como palabras (Bag of words, 2017).

BoW enumera las palabras con sus recuentos de palabras por documento. En la tabla donde las palabras y los documentos se convierten efectivamente en vectores se almacenan, cada fila es una palabra, cada columna es un documento y cada celda es un recuento de palabras. Cada uno de los documentos del cuerpo está representado por columnas de igual longitud. Esos son vectores de recuento de palabras, una salida despojada de contexto.

Antes de que sean alimentados a la red neuronal, cada vector de las cuentas de palabras se normaliza de tal manera que todos los elementos del vector se suman a uno. De este modo, las frecuencias de cada palabra se convierten efectivamente para representar las probabilidades de la aparición de esas palabras en el documento. Las probabilidades que sobrepasen ciertos niveles activarán nodos en la red e influirán en la clasificación del documento.

- **TF-IDF.**

Tf-idf es una técnica de minería de texto utilizada para categorizar documentos y el peso de tf-idf es un peso que se utiliza con frecuencia en la recuperación de información y en la extracción de texto. Este peso es una medida estadística utilizada para evaluar la importancia de una palabra para un documento en una colección o corpus. La importancia aumenta proporcionalmente al número de veces que una palabra aparece en el documento, pero está compensada por la frecuencia de la palabra en el corpus. Las variaciones del esquema de ponderación tf-idf son utilizadas con frecuencia por los motores de búsqueda como una herramienta central para anotar y clasificar la relevancia de un documento dada una consulta del usuario (Tf-idf weighting, 2008).

Típicamente, el peso tf-idf está compuesto por dos términos: el primero calcula la Frecuencia de Término normalizada (TF), dividiendo el número de veces que una palabra aparece en un documento entre el número total de palabras en ese documento; El segundo término es la Frecuencia Inversa de Documento (IDF), computada como el

logaritmo del número de documentos en el cuerpo dividido por el número de documentos donde aparece el término específico (Wu, Luk, Wong, & K., 2008).

- TF: Término Frecuencia, que mide la frecuencia con que un término aparece en un documento. Como cada documento es diferente en longitud, es posible que un término aparezca muchas más veces en documentos largos que en documentos más cortos. Por lo tanto, el término frecuencia es a menudo dividido por la longitud del documento (es decir, el número total de términos en el documento) como una forma de normalización:
  - ✓ TF (t) = (Número de veces que el término t aparece en un documento) / (Número total de términos del documento).
- IDF: Inverse Document Frequency, que mide la importancia de un término. Mientras se calcula TF, todos los términos se consideran igualmente importantes. Sin embargo, se sabe que ciertos términos, como "es", "de" y "eso", pueden aparecer muchas veces, pero tienen poca importancia. Por lo tanto, tenemos que sopesar los términos frecuentes mientras aumentamos los raros, calculando lo siguiente:
  - ✓ IDF (t) = log_e (Número total de documentos / Número de documentos con el término t en ella).

En primer lugar, TF-IDF mide el número de veces que las palabras aparecen en un documento dado (es decir, frecuencia del término). Pero porque las palabras como "y" o "el" aparecen con frecuencia en todos los documentos, éstos son sistemáticamente descontados. Esa es la parte de la frecuencia del documento inverso. Cuantos más documentos aparezca una palabra, menos valiosa será esa palabra como señal. Que se destina a dejar sólo las frecuentes y distintivas palabras como marcadores. La relevancia de cada palabra TF-IDF es un formato de datos normalizado que también agrega hasta uno (Wu, Luk, Wong, & K., 2008).

Estas palabras marcadoras se alimentan a continuación a la red neural como rasgos con el fin de determinar el tema cubierto por el documento que los contiene. Aunque simple, TF-IDF es increíblemente potente, y contribuye a herramientas omnipresentes y útiles como la búsqueda de Google.

- **N-gramas.**

El n-gramas es una subsecuencia de n elementos consecutivos en una secuencia dada que utiliza las N-1 palabras anteriores para predecir la siguiente palabra. Permite

capturar las probabilidades de secuencias de palabras mediante técnicas estadísticas simples (Rodriguez, 2012).

Los n-gramas son utilizados en el procesamiento estadísticos de lenguaje natural de textos, así como en el análisis de imágenes. Utilizan como modelos de lenguaje (LM) unigamas, bigramas, trigramas,…, n-gramas. Mayormente estos modelos se encuentran en cuerpos de datos muy grandes como diarios y blog (Gravano, 2014).

Los n-gramas utilizan la regla de la cadena $P(A \wedge B) = P(A|B) \cdot P(B)$ para estimar la probabilidad de una oración. Para calcular estas probabilidades se debe tener un corpus suficientemente extenso y contar la cantidad de veces que aparece una palabra, la cantidad de veces que aparecen dos palabras juntas, tres palabras, etc. Pero como en general es difícil es complicado y no siempre se cuenta con un corpus suficientemente extenso, se utiliza el supuesto de Markov que plantea que la probabilidad de una palabra depende solamente de las n-1 palabras anteriores (Gravano, 2014).

$$P(W_n|W_1^{n-1}) \approx P(W_n|W_{n-N+1}^{n-1})$$

1.4.3 Clasificación.

La segunda etapa de los sistemas CBIR tiene como objetivo analizar los datos obtenidos en la etapa de descripción, para aprender a distinguir entre diferentes objetos a través de funciones de distancias, esta etapa se conoce como clasificación.

Con la clasificación lo que se busca es aprender funciones que transformen el espacio de entrada de los datos, de manera que los elementos de la misma clase se encuentren lo más cercano posible y los elementos de clases distintas se encuentren lo más lejos posible. Para el desarrollo de la línea base que se desea implementar se puede utilizar la función fitctree, que es una función MATLAB que devuelve un árbol de clasificación basado en las variables de entrada X, también conocidas como predictores, características o atributos; y las salidas Y, que son las respuestas o etiquetas. Además se puede utilizar la función MATLAB predict para predecir la salida del modelo identificado a partir de los datos de entrada y salida.

1.5 CONCLUSIONES PARCIALES

En este capítulo se realizó un estudio detallado de la Recuperación semántica de imágenes, definida como el estudio e investigación de la documentación visual. La

recuperación semántica de imágenes con tres etapas fundamentales: la primera etapa basada en las representaciones textuales de las características de las imágenes, la segunda basada en los rasgos visuales de las imágenes y, por último, la tercera etapa emplea simultáneamente el código visual y el código textual para representar y recuperar imágenes.

Además, se expuso las características de los tres módulos fundamentales que conforman los sistemas de recuperación: Módulo de Descripción, Módulo de Consulta y Módulo de Búsqueda. También se analizó la arquitectura CBIR, que permite realizar consultas en las que el usuario escoge el color o colores más destacados de la imagen que busca, además puede señalar las formas más sobresalientes de la imagen y puede definir regiones completándolas con colores sólidos y en la localización espacial deseada. Se defino también, las tres partes que la conforman: descripción, representación y clasificación.

Se analizaron varios de los métodos que se pueden utilizar para la descripción, específicamente el SIFT, el SURF y el HOG. Para la representación se analizaron los métodos IF-IDF, Bag of Word y n gramas. Por último se para la clasificación se determinó que se pueden utilizar las funciones MATLAB fitctree y predict.

CAPÍTULO 2: PROPUESTA DE LA SOLUCIÓN.

2.1 PROPUESTA DE LA LÍNEA BASE.

La línea base que se propone como solución se rige por la arquitectura de los sistemas CBIR, modelo de recuperación visual de imágenes basado en el uso de las características intrínsecas de los documentos que son extraídas y representadas automáticamente a través de estructuras de datos numéricos, que, para este caso, los documentos a procesar son imágenes. Esta propuesta cuenta con tres procesos fundamentales: descripción, representación y clasificación, por los cuales va a pasar una serie de imágenes pertenecientes a una base de datos de imágenes.

A través del primer proceso se extraerán las principales propiedades de las imágenes y sus puntos clave que son invariantes a las transformaciones a fines. Este proceso se realizará mediante el algoritmo SURF, porque reduce la dimensionalidad y complejidad del cálculo de los vectores de características de los puntos de interés obtenidos, mientras continúan siendo suficiente características e igualmente repetitivos. (Mikolajczyk & Schmid. , 2005)

Mediante el segundo proceso, se transformará la información visual en un espacio vectorial para su posterior clasificación. El algoritmo utilizado para este proceso es el Bag-of-Words (BoW), mediante el cual se tratan documentos para definir el número de concurrencias de cada palabra, siguiendo el mismo proceso, este algoritmo trata las imágenes como documentos y crea un histograma que cuenta el número de concurrencia de ciertas características de la imagen. Además de todos los modelos es el más simple, debido a que no codifica ninguna relación de los descriptores, además que, a partir de este modelo, es posible crear modelos que expresen las relaciones entre los descriptores y reduce el costo computacional de la representación.

Por último, a través del tercer proceso se analizarán los datos obtenidos en la descripción de las imágenes para distinguir entre diferentes objetos utilizando las funciones MATLAB fitctree y predict, de manera que se le pueda asignar a una imagen una clase específica y predecir la salida de dicha asignación.

El proceso que realiza la propuesta se muestra en la figura siguiente:

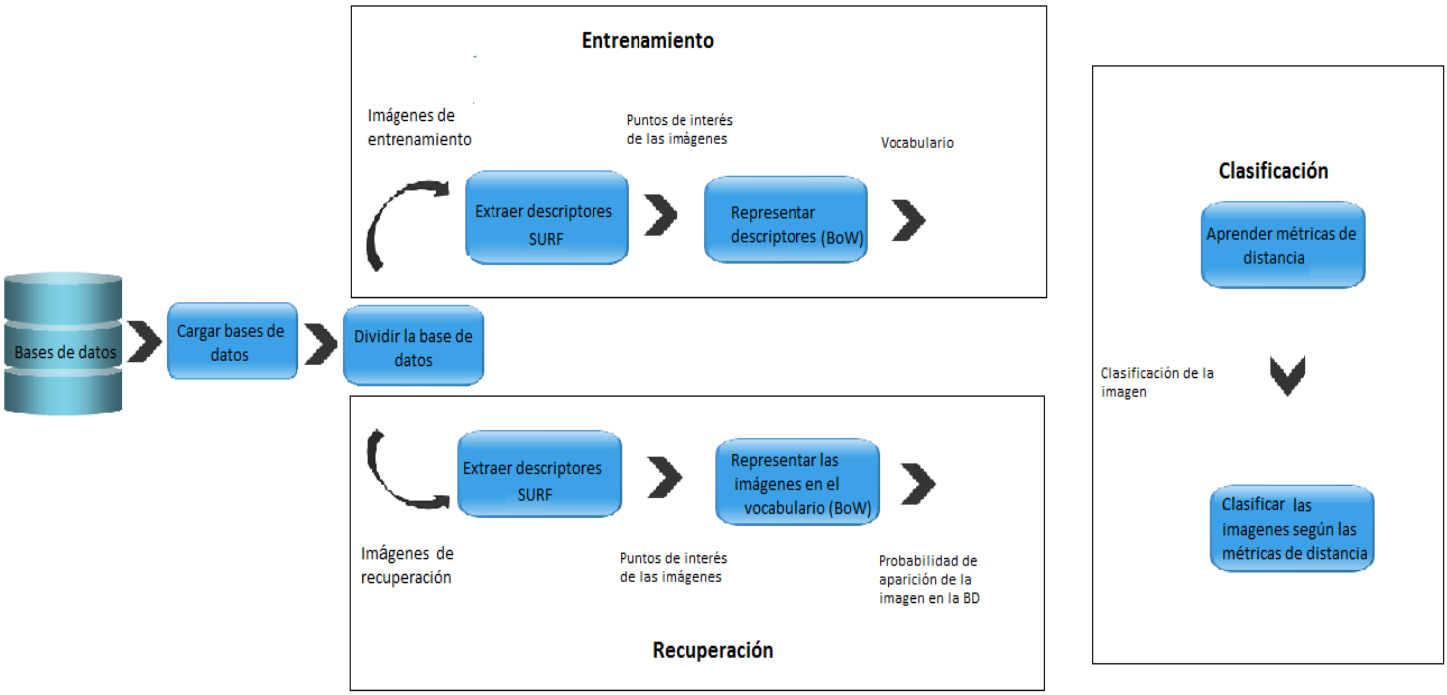


Ilustración 4 Proceso de recuperación de imágenes.

2.2 ETAPAS DE LA LÍNEA BASE.

El proceso que realiza la propuesta de la línea base definida anteriormente realiza los siguientes pasos:

1. Se carga la base de imágenes para dividirla en imágenes de entrenamiento e imágenes de recuperación.
2. Se procesan las imágenes de entrenamiento, extrayendo los descriptores para obtener los puntos de interés, que son representado posteriormente y se obtiene como resultado el vocabulario.
3. Se procesan las imágenes de recuperación, extrayendo los descriptores para posteriormente representar los puntos de interés obtenidos en el vocabulario que se creó en el paso anterior.
4. Por último se realiza la predicción y clasificación de las imágenes.

2.3 DESARROLLO DE LOS ALGORITMOS

2.3.1 Descriptor SURF.

Para la extracción de los descriptores se utilizará el descriptor SURF, como se ha indicado anteriormente. Este método cuenta, con cuatros pasos, los cuales son:

1. Detección de puntos de interés
2. Asignación de la orientación
3. Creación del descriptor
4. Matching entre puntos clave

Detección de puntos de interés

En la primera etapa el descriptor SURF utiliza el valor del determinante de la matriz Hessiana para la localización y la escala de los puntos por su rendimiento en cuanto a la velocidad del cálculo y a la precisión. Este descriptor, con respecto a otros, no utiliza diferentes medidas para el cálculo de la posición y la escala de los puntos de interés individualmente, sino que utiliza el valor del determinante de la matriz Herssianaen en ambos casos (Mikolajczyk & Schmid. , 2005).

Dado un punto p = (x, y) de la imagen I, la matriz Hessiana H (p, σ) del punto p perteneciente a la escala σ se define como:

$$H(p,\sigma) = \begin{bmatrix} Lxx(p.\sigma) & Lxy(p.\sigma) \\ Lxy(p.\sigma) & Lyy(p.\sigma) \end{bmatrix}$$

donde Lxx (p, σ) representa la convolución de la derivada parcial de segundo orden de la Gaussiana $\frac{\partial^2}{\partial x^2} \quad g(\sigma)$ con la imagen I en el punto p. De manera análoga ocurre con los términos L xy (p, σ), L yy (p, σ) de la matriz.

Debido a una serie de limitaciones de los filtros gaussianos, como la necesidad de ser discretizados, la falta de prevención total del indeseado efecto aliasing, etc.…, se implementó en el descriptor SURF los filtros tipo caja, que es una alternativa a los filtros gaussianos. Estos filtros realizan una aproximación de las derivadas parciales de segundo orden de las gaussianas para ser evaluadas de manera rápida usando imágenes integrales independientemente del tamaño de estas (Mikolajczyk & Schmid. , 2005).

Las imágenes integrales se calculan mediante la siguiente fórmula:

$$Ii\sum(x,y) = \sum_{i=1}^{i<x}\sum_{j=1}^{jzy} I \quad (i,j)$$

donde (x, y) representan la posición del punto en la imagen y Ii (x, y) representa la intensidad de la imagen en el punto.

Luego de calcular la imagen integral se calcula la suma de las intensidades de una región mediante la siguiente operación:

$$\sum I = \quad Ii_D + \quad Ii_{DA} + \quad Ii_B + \quad Ii_C$$

El tiempo necesario para calcular las operaciones de convolución es independiente del tamaño de la imagen.

El espacio escala se analiza mediante la elevación del tamaño del filtro, directamente en la imagen original.

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:

$$\boldsymbol{det}(\boldsymbol{H_{aprox}}) = \quad \boldsymbol{D_{xx}D_{yy}} - \quad (\boldsymbol{0.9D_{xy}})^{\boldsymbol{2}}$$

donde el valor de 0,9 está relacionado con la aproximación del filtro gaussiano.

Tras la convolución de la imagen original con un filtro de dimensiones 9X9 se obtiene una imagen que es considerada como la escala inicial o como la máxima resolución espacial. Las capas sucesivas se obtienen mediante la aplicación gradual de filtros de mayores dimensiones, evitando así los efectos de aliasing en la imagen.

El espacio escala del descriptor SURF está divido en octavas, compuestas por un número fijo de imágenes como resultado de la convolución de la misma imagen original con una serie de filtros cada vez más grandes. Dentro de una misma octava, el incremento o paso de los filtros de una misma octava es el doble respecto del paso de la octava anterior y al mismo tiempo el primero de los filtros de cada octava es el segundo de la octava predecesora, obteniendo una serie de octavas con sus respectivos filtros (Mikolajczyk & Schmid. , 2005).

Para calcular la localización de todos los puntos de interés en todas las escalas se eliminan los puntos que no cumplen la condición de máximo en un vecindario de 3X3X3. El máximo determinante de la matriz Hessiana es interpolado en la escala y posición de la imagen.

Asignación de la orientación

En esta etapa se le asigna una orientación a cada uno de los puntos de interés obtenidos en la etapa anterior. Primero se calcula la respuesta Haar en ambas direcciones x e y mediante las funciones. Para el cálculo del área de interés se toma un área circular entrada en el punto de interés y un radio 6s, siendo s la escala en la que el punto de interés ha sido detectado. La etapa de muestreo va a depender de la escala y se toma

como valor s y las funciones ondulares de Haar toma el valor 4s, donde a mayor valor de escala mayor es la dimensión de las funciones onduladas (Mikolajczyk & Schmid. , 2005).

Luego de realizar todas las operaciones antes expuestas, se procede al filtrado mediante las máscaras Haar utilizando imágenes integrales para obtener las respuestas en ambas direcciones. Estas respuestas son representadas mediante vectores en el espacio colocando las respuestas horizontales y verticales en el eje de abscisas y ordenadas respectivamente. Así se obtiene una orientación dominante por cada sector mediante la suma de todas las respuestas dentro de una ventana de orientación móvil. Finalmente, la orientación del punto de interés será aquella cuyo vector sea el más grande dentro de los 6 sectores en los que han sido dividida el área circulas alrededor del punto de interés.

Creación del descriptor

En la última etapa se concreta la creación del descriptor SURF. Como primer paso se construye una región alrededor del punto de interés y orientada en relación a la orientación calculada en la etapa anterior. A su vez, esta región es dividida en sub-regiones dentro de cada una de las cuales se calculan las respuestas Haar de puntos con una separación de muestreo de 5X5 en ambas direcciones.

En cada una de las sub-regiones se suman las respuestas de Haar en las direcciones horizontal y vertical respectivamente relativas a la orientación del punto de interés obteniendo un valor representativo por cada una de las sub-regiones, a la vez se suman los valores absolutos de las respuestas Haar en cada una de las sub-regiones para obtener información de la polaridad sobre los cambios de intensidad. Así cada una de las sub-regiones queda representada por un vector v de componentes y por lo tanto, englobando las 4 x 4 sub-regiones, resulta un descriptor SURF con una longitud de 64 valores para cada uno de los puntos de interés identificados (Mikolajczyk & Schmid. , 2005).

Matching entre puntos clave

El matching entre los puntos clave, al igual que en el descriptor SIFT, hace una correspondencia de los puntos claves identificados entre dos imágenes. En el caso del SURF, la estrategia para establecer las correspondencias entre los puntos claves de ambas imágenes es la de “el vecino más próximo”.

2.3.2 Bag-of-Word.

El modelo Bag of Words es utilizado en el procesamiento de imágenes, representando una imagen en función de la frecuencia de una serie de elementos visuales. Este método se utilizará para la clasificación de las imágenes.

Para la creación de un modelo BoW es necesario disponer de conjuntos de imágenes, etiquetados y normalmente grandes, llamados datasets. Los datasets están formados por cientos o miles de imágenes de las que normalmente se conoce su contenido. En cada imagen del datasets, es necesario seleccionar una serie de características que permiten representarla. Por lo que es necesario seleccionar las posiciones en las que se extraerán dichas características (Bag of words, 2017).

Se hace necesario dividir el datasets en dos conjuntos disjuntos, uno de entrenamiento y otro de prueba o tests. El conjunto de entrenamiento suele ser mayor al de test, para darle al sistema más robustez, aunque no es imprescindible. Ninguna imagen de prueba ha de estar en el conjunto de entrenamiento y viceversa. Los descriptores extraídos son usados para construir un diccionario o conjunto de palabras que representan los puntos clave. Para la construcción del diccionario se utiliza única y exclusivamente los descriptores pertenecientes a la imagen del conjunto de entrenamiento (Bag of words, 2017).

Una vez que el diccionario está construido, cada imagen del datasets se representa mediante histogramas que contiene con qué frecuencia aparecen en esa imagen cada una de las palabras visuales del diccionario. Para ello se mide la distancia de cada descriptor extraído de una imagen con todas las palabras visuales del diccionario, dicho descriptor quedara representado por palabras visuales cuya distancia sea menor. Repetido este proceso para todos los descriptores, cada imagen del datasets quedará representada por vectores cuyo tamaño será el mismo que el número de palabras visuales del diccionario. Cada elemento del vector indicara la frecuencia de esa palabra dentro de la imagen correspondiente (Bag of words, 2017).

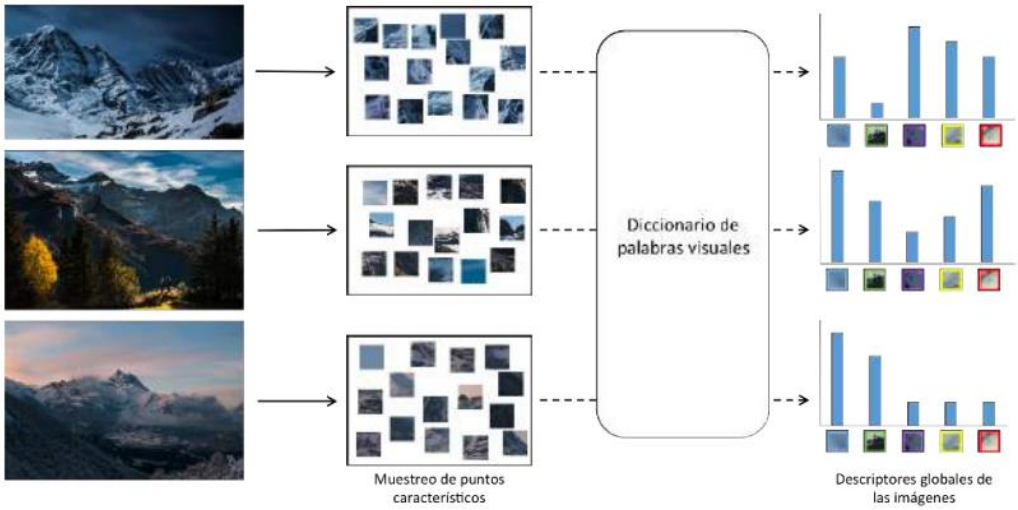


Ilustración 5 Diagramas de representación de imágenes 1.

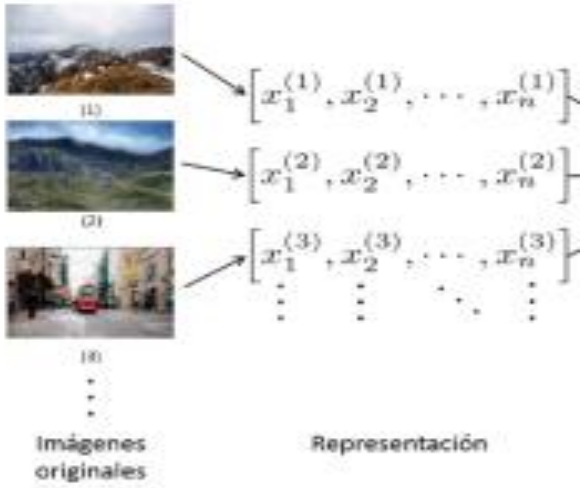


Ilustración 6 Diagramas de representación de imágenes 2.

2.3.3 Funciones MATLAB Fitctree y Predict.

Fitctree

La función MATLAB fitctree devuelve un árbol de clasificación basado en las variables de entrada X y salida Y, donde cada nodo de bifurcación se divide en base a los valores de una columna X, permitiendo la clasificación multiclase a partir de la matriz de datos y el vector de etiquetas.

Un árbol consiste esencialmente en un conjunto secuencial de condiciones y acciones que relacionan unos determinados factores con un resultado o decisión. Un árbol de decisión es una forma de representar el conocimiento obtenido en el proceso de

aprendizaje inductivo. Se crea una partición del espacio a partir de un conjunto de prototipos y la estructura resultante es el árbol. Un árbol está formado por nodos, nodos internos y nodos terminales.

Nodos internos: cada nodo interior contiene una pregunta sobre un atributo concreto (node = <attribute, value>) y da lugar a dos hijos, uno por cada posible respuesta, clasificación o decisión.

Nodos terminales (hojas): son los que están asignados a una única clase, aquellos en los que termina el árbol. La complejidad del árbol viene determina por su número de hojas.

La construcción de un árbol es la fase de aprendizaje del método. Consiste en analizar el conjunto de prototipos disponibles, el conjunto de atributos, y obtener unas reglas lógicas que se ajustan a la clasificación conocida, a las clases que están asignadas a cada vector (a1,… as).

El proceso de construcción es recursivo:

1. Se analizan todas las posibles particiones (attribute, value) y se toma de todas ellas la que da lugar a una mejor separación.
2. Se aplica la separación óptima.
3. Se repite el paso 1 con los nodos hijos, únicamente con los que no sean hojas.

Para cada uno de los atributos se toman todos los valores de sus observaciones. Las particiones se definen tomando como corte el punto medio entre cada dos de esos valores.

Una de las grandes ventajas de los árboles de decisión es que son muy intuitivos y fáciles de aplicar.

La aplicación del árbol se conoce como fase de clasificación y consiste en asignar la clase que corresponde a un patrón x, independiente del conjunto de aprendizaje.

Una vez creado es sencillo encontrar la clase desconocida de un nuevo dato x con características (a1,… as). Basta con contestar las preguntas planteadas en cada nodo y seguir el camino impuesto por el árbol, hasta encontrar un nodo hoja. Se predice la clase c de x como la clase mayoritaria en la hoja a la que pertenece x.

Por defecto fitctree utiliza el algoritmo CART estándar para crear árboles de decisión, realizando los siguientes pasos:

1. Comienza analizando todos los datos de entrada y examina las posibles divisiones binarias en cada predictor.
2. Selecciona una división con el mejor criterio de optimización.
3. Realiza la división.
4. Repite de forma recursiva para los nodos secundarios.

Esto se realiza teniendo en cuenta los criterios de optimización y la regla de parada.

Predict

Esta función MATLAB predice la salida de un modelo identificado a partir de datos de entrada y salida, devolviendo un vector de etiquetas de clase predichos para los datos de predicción de la tabla o matriz X, basados en el entrenado, árbol de clasificación completo.

2.4 USO DE TECNOLOGÍA.

2.4.1 Herramienta MATLAB.

Para el desarrollo de la línea base se utilizó el MATLAB que es una herramienta interactiva basada en matrices para cálculos científicos y de ingeniería (de hecho, el término MATLAB procede de matrix laboratory). Desde el punto de vista del control, MATLAB se puede considerar un entorno matemático de simulación que puede utilizarse para modelar y analizar sistemas. Permitirá el estudio de sistemas continuos, discretos, lineales y no lineales, mediante descripción interna y externa, en el dominio temporal y frecuencial. (Bishop, 1993). Para el caso de manipulación de imágenes se emplea el toolbox “Image Processing”.

En la representación de imágenes, el MATLAB almacena las imágenes como vectores bidimensionales (matrices), en el que cada elemento de la matriz corresponde a un solo pixel. Trabajar con imágenes en el MATLAB es equivalente a trabajar con el tipo de datos matriz.

2.5 CONCLUSIONES PARCIALES.

En este capítulo se definió como propuesta el desarrollo de una línea base capaz de describir a través del SURF las imágenes, para posteriormente ser representadas a

través del Bag of Words y por último se clasificaron a través de las funciones MATLAB Fitctree y Predict. Se realizó la descripción de los pasos del proceso que realiza la propuesta y para la implementación se definió al MATLAB como la herramienta a utilizar, que es explicada detalladamente en este capítulo.

CAPÍTULO 3: PRUEBAS

Las pruebas de software se realizan para controlar la calidad del software y verificar la correcta implementación que las funcionalidades establecidas por el cliente. Para esto se realiza la ejecución de casos de prueba que permiten validar la propuesta implementada.

Como se describió en el capítulo anterior, la programación se desarrolló en la herramienta MATLAB cuya interfaz de desarrollo se evidencia en la imagen siguiente:

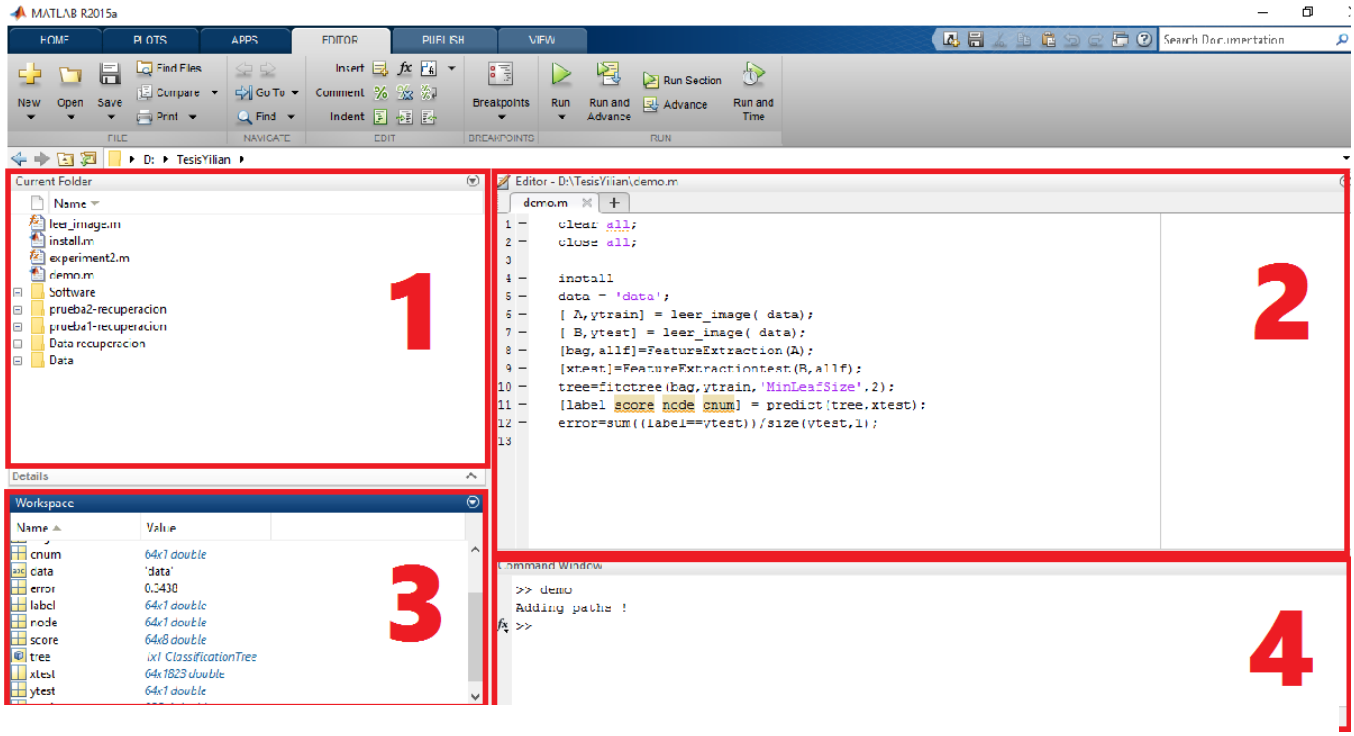


Ilustración 7 Interfaz de desarrollo de la herramienta MATLAB.

Donde el área uno muestra las clases que contiene el código y las bases de imágenes utilizadas, el área 2 es el área de desarrollo, en el área tres se encuentran los resultados obtenidos, donde cada paquete representa una tabla con dichos resultados y por último se encuentra el área número cuatro que muestra si la programación corre correctamente, de tener errores muestra un mensaje en rojo con la información de la línea de código y la clase que contiene el error.

Para la validación y verificación del correcto funcionamiento de la línea base se realizaron pruebas del falso positivo y falso negativo para analizar el porcentaje de aciertos mediante el método de validación cruzada con K iteraciones, también llamado en inglés como "k fold cross validation". Esta técnica consiste en dividir las imágenes en varias dataset, seleccionando uno para testear y el resto para entrenar el árbol de

decisión. Este proceso se realiza repetidamente hasta testear con cada set de datos. Los resultados son guardados en una tabla para analizar la eficiencia de la predicción. En el grafico siguiente se muestra un ejemplo para K=3.

Para determinar la base de datos (imágenes) a utilizar se realizó un estudio bibliográfico teniendo como resultado que una de las bases de datos más utilizada en los últimos años es la base de imágenes ETH80, además de ser una de las más variadas debido a la cantidad de clases que contiene. Es Utilizada en artículos como “Extracting Structures in Image Collections for Object Recognition” de los autores Sandra Ebert, Diane Larlus y Bernt Schiele, pertenecientes al Departamento de Ciencias de la Computación, TU Darmstadt, Alemania. Ademas está el artículo “Memory Organization for Invariant Object Recognition and Categorization“del autor Guillermo S. Donatti perteneciente al Instituto de Cálculo Neuronal, Universidad de Ruhr Bochum, Alemania. Esta base de datos está conformada por 400 imágenes, divididas en 8 clases de objetos, cada clase tiene 50 vistas de diferentes ángulos. Teniendo en cuenta las características anteriores se escogió la ETH80 para realizar la prueba de validación.

Para la división de los datos se utilizó el método de selección aleatoria estratificada, que consiste en dividir la base de imágenes en grupos, seleccionando aleatoriamente los sujetos finales de cada grupo en forma proporcional en función de características determinadas, para este caso tendremos en cuenta que los sujetos no se superpongan y que sean lo más variado posible geográficamente.

Para el experimento se realizó la validación cruzada con K=3, para cada k se realizó la selección de 8 vistas diferentes de cada clase. Teniendo en cuenta que son 8 clases por las 8 vistas seleccionadas, se obtiene un grupo de 64 imágenes para el test y, al restarle este grupo a las 400 imágenes que conforma la ETH80, quedan 336 imágenes para entrenamiento. En cada iteración se realiza con grupos diferentes a la iteración anterior.

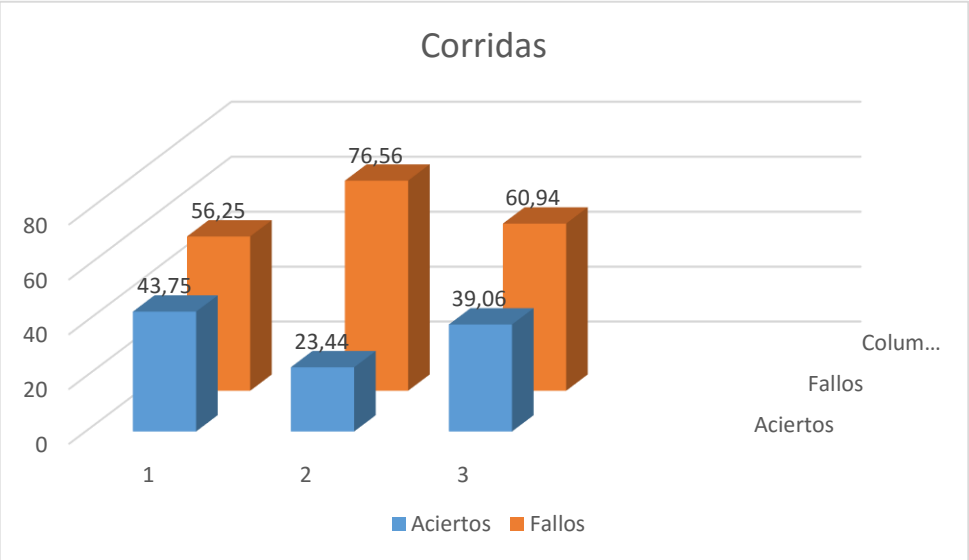
Tabla 1 Resultados favorables de las corridas.

Corridas	Total de aciertos	% de aciertos
1	28	43.75
2	15	23.44
2	25	39.06

Tabla 2 Resultados desfavorables de las corridas.

Corridas	Total de fallos	% de fallos
1	36	56.25
2	49	76.56
2	39	60.94

Los resultados anteriores se recogen en la siguiente grafica para tener una mejor visibilidad de los resultados:



Se puede observar la cantidad de aciertos y fallos arrojados por cada una de las corridas realizadas, donde el porcentaje de fallos esta entre el 55 y 80 %, mientras que el porcentaje de aciertos se encuentra entre el 20 y 45 %, lo que significa que, a pesar que la solución implementada cumple con las funcionalidades establecidas, no alcanza un buen porcentaje de aciertos debido al efecto de esparcidad, pues se evidencia un gran aparición de eventos nulos en el vector de características.

3.1 CONCLUSIONES PARCIALES.

Con la realización de las pruebas, se puede determinar que la solución implementada realiza una correcta recuperación semántica de imágenes, aunque suele ser impredecible para imágenes con objetos similares. Esto demuestra que, con respecto a la descripción tomando características visuales como color, iluminación y sombra, los descriptores de alto nivel aportan mayor información sobre los objetos y acontecimientos de la escena.

A pesar de quedar probado el funcionamiento de la línea base se detectó como deficiencia que el porcentaje de acierto está por debajo del porcentaje de fallos debido al efecto de la esparcidad.

CONCLUSIONES

Para la realización de la tesis se hizo un análisis teórico metodológico de los principales conceptos de la recuperación semántica de imágenes, así como las diferentes etapas por las que pasó, enfatizando la segunda etapa denominada CBIR debido a que la solución se basó en dicha etapa. Dentro de esta etapa se estudiaron varios métodos empleados en las diferentes partes de la arquitectura CBIR, teniendo como resultado que para la descripción se utilizó el método SURF, para la representación se utilizó el método Bag of Word y por ultimo para la clasificación se utilizó las funciones MATLAB Fitctree y Predict.

Se desarrolló una línea base, que integra los métodos antes mencionados para su posterior utilización en el desarrollo de nuevas aplicaciones dedicadas al procesamiento de imágenes digitales, especialmente utilizando la recuperación semántica de imágenes.

Con la realización de la validación de la línea base, a través de la prueba del falso positivo quedó probado el correcto funcionamiento de la solución implementada.

RECOMENDACIONES

Implementar una nueva versión de la línea base desarrollada más robusta teniendo en cuenta el efecto de esparcidad para mejorar el porcentaje de aciertos y ganar en eficacia.

Realizar pruebas con otras bases de datos a parte de la ETH80 para comparar resultados.

REFERENCIAS

1. Almarales, F. R., Santos Martínez, G., & González, H. (2016). *Adaptación del algoritmo LMNN para Problemas de Predicción con Salidas Compuestas.* La Habana.

2. Álvarez, J. (2017). *Sistemas automatizados de identificación.* México.

3. Álvarez, S. P. (2007). *Sistemas CBIR: Recuperación de imágenes por.* Gijón: Trea.

4. B., M. C. (2010). *Recuperación de imágenes y datos de fuentes Heterogeneas.* Cali.

5. Bag of words. (2017). *DEEPLEARNING4J.* Obtenido de https://deeplearning4j.org/bagofwords-tf-idf

6. Bautista, J. O. (2008). *Estudio de métodos de indexación y recuperación en bases de datos de imágenes.* San Sebastián.

7. Bishop, R. (1993). *Modern Control Systems Analysis and Design Using matlab.* AddisonWesley,.

8. Caicedo, J. C., González, F., & Romero , E. (2014). *Prototipo de Sistema para Almacenamiento y Recuperación por Contenido en Imágenes Médicas de Histopatología.* Colombia.

9. Campos, F. R. (2016). *Algoritmos basados en aprendizaje de funciones .* Villa Clara, Cuba.

10. Comeche, J. A. (2013). *La recuperación automatizada de imágenes: retos y soluciones.* Madrid.

11. Comeche, J. A. (2013). *La recuperación automatizada de imágenes: retos y soluciones.* Madrid.

12. Cruz, N. M. (2013). *Desarrollo de un sistema avanzado de asistencia a la conducción en tiempo real para la dirección de patrones en entornos urbanos conplejos.* Leganés.

13. Cuenca, J. S. (2008). *Reconocimiento de objetos .* Barcelona.

14. Cuenca, J. S. (2008). *Reconocimiento de objetos por descriptores .* Barcelona.

15. Feng, L., Brussee, R., Blanken, H., & Veenstra, M. (2007). *Languages for Metadata, en BLANKEN, H.M; VRIES, A.P. De; BLOK, H.E.; FENG, L.* Berlin: Springer-Verlag, pp. 23-51.: Multimedia Retrieval.

16. García, Ó. B. (2011). *Etudio comparativo de descriptores visuales para la deteccion de escenas causi-duplicadas.* Madrid.

17. García, R. H. (2014). *Tesis presentada en opcio´n al Grado Cient´ıfico de Doctor en Ciencias T´ecnicas.* La Habana.

18. Gardcía, R. H. (2014). *Modelo de representación de n-gramas.* La Habana.

19. Gonzalez, R., & Woods, R. (2002). *Digital image processing.* N.J.: Upper Saddle River.

20. GONZALEZ, R., & WOODS, R. (2008). *Digital Image Processing.* New Jersey: Pearson Education: Third ed.

21. Gravano, A. (2014). *Modelo del Lenguaje. N-gramas.* La Habana.

22. Juan Carlos Caicedo R. (s.f.). *Recuperación de Imágenes Médicas por Contenido: arquitectura, técnicas y aproximaciones.* Colombia.

23. (2014). *Kernel Regression.* Obtenido de Whant is Kernel Regression.

24. Larin-Fonseca, R., Garea-Llano , E., & Chacón-Cabrera , Y. (2012). *Semántica y Descriptores Invariantes para el Procesamiento de Imágenes de Teledetección.* La Habana.

25. Mikolajczyk, K., & Schmid. , C. (2005). A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions* , 27(10):1615 –1630.

26. Monroy, A. P., Montes y Gómez, M., Jair Escalante , H., & González, F. (2016). *Categorizacióndeimágenes mediantetécnicasdemineríadetexto.* México.

27. Mosquera González, A., Carriera Nouche, M., & Hónzalez Penedo, M. (2011). *Recuperación de información: Un enfoque práctico y multidisciplinar.* Madrid: Ra-Ma: “Recuperación de imagen”, en Cacheda Seijo, F. et al.(Eds.).

28. MUller, H., Clough, P., Deselaers, T., & Caputo, B. (2010). *ImageCLEF: Experimental Evaluation in Visual Information Retrieval.* Berlin: Springer-Verlag.

29. PIÑAR, B. A. (2012). *Detección automática de objetos en imágenes mediante el uso de puntos característicos .* Universidad Autónoma de Madrid .

30. R., J. M. (2004). *Aplicaciones clínicas del procesamiento digital de imágenes médicas .*

31. Rodriguez, J. M. (12 de 10 de 2012). *Modelos de Lenguaje: Introducción a N-Gramas .*

32. Sarriá, N. M. (2010). *Sistema de clasificación automática de imágenes basado en contenido.* MADRID.

33. Soto, R. V. (2014). *Implementación de un sistema reconocedor de eventos en videos, con clasificador K-NN.* SANTIAGO DE CHILE.

34. *Tf-idf weighting.* (2008).

35. Weinberg, K., & Saul, L. (2009). *Distance metric learning for large margin nearest neighbor classificatio.* Journal of Machine Learning Research 10, 207-244.

36. Wu, H., Luk, R., Wong, K., & K., K. (2008). *"Interpreting TF-IDF term weights as making relevance decisions".* ACM Transactions on Information Systems.

37. Yarza, J. T. (2013). *Multi-clasificación Discriminativa de Partes Corporales basada en Códigos Correctores de Errores .* Barcelona.

BIBLIOGRAFÍA

1. Almarales, F. R., Santos Martínez, G., & González, H. (2016). Adaptación del algoritmo LMNN para Problemas de Predicción con Salidas Compuestas. La Habana.

2. Álvarez, J. (2017). Sistemas automatizados de identificación. México.

3. Álvarez, S. P. (2007). Sistemas CBIR: Recuperación de imágenes por. Gijón: Trea.

4. B., M. C. (2010). Recuperación de imágenes y datos de fuentes Heterogeneas. Cali.

5. Bag of words. (2017). DEEPLEARNING4J. Obtenido de https://deeplearning4j.org/bagofwords-tf-idf

6. Bautista, J. O. (2008). Estudio de métodos de indexación y recuperación en bases de datos de imágenes. San Sebastián.

7. Bishop, R. (1993). Modern Control Systems Analysis and Design Using matlab. AddisonWesley,.

8. Caicedo, J. C., González, F., & Romero , E. (2014). Prototipo de Sistema para Almacenamiento y Recuperación por Contenido en Imágenes Médicas de Histopatología. Colombia.

9. Campos, F. R. (2016). Algoritmos basados en aprendizaje de funciones . Villa Clara, Cuba.

10. Comeche, J. A. (2013). La recuperación automatizada de imágenes: retos y soluciones. Madrid.

11. Comeche, J. A. (2013). La recuperación automatizada de imágenes: retos y soluciones. Madrid.

12. Cruz, N. M. (2013). Desarrollo de un sistema avanzado de asistencia a la conducción en tiempo real para la dirección de patrones en entornos urbanos conplejos. Leganés.

13. Cuenca, J. S. (2008). Reconocimiento de objetos . Barcelona.

14. Cuenca, J. S. (2008). Reconocimiento de objetos por descriptores . Barcelona.

15. Feng, L., Brussee, R., Blanken, H., & Veenstra, M. (2007). Languages for Metadata, en BLANKEN, H.M; VRIES, A.P. De; BLOK, H.E.; FENG, L. Berlin: Springer-Verlag, pp. 23-51.: Multimedia Retrieval.

16. García, Ó. B. (2011). Etudio comparativo de descriptores visuales para la deteccion de escenas causi-duplicadas. Madrid.

17. García, R. H. (2014). Tesis presentada en opcio´n al Grado Cient´ıfico de Doctor en Ciencias T´ecnicas. La Habana.

18. Gardcía, R. H. (2014). Modelo de representación de n-gramas. La Habana.

19. Gonzalez, R., & Woods, R. (2002). Digital image processing. N.J.: Upper Saddle River.

20. GONZALEZ, R., & WOODS, R. (2008). Digital Image Processing. New Jersey: Pearson Education: Third ed.

21. Gravano, A. (2014). Modelo del Lenguaje. N-gramas. La Habana.

22. Juan Carlos Caicedo R. (s.f.). Recuperación de Imágenes Médicas por Contenido: arquitectura, técnicas y aproximaciones. Colombia.

23. (2014). Kernel Regression. Obtenido de Whant is Kernel Regression.

24. Larin-Fonseca, R., Garea-Llano , E., & Chacón-Cabrera , Y. (2012). Semántica y Descriptores Invariantes para el Procesamiento de Imágenes de Teledetección. La Habana.

25. Mikolajczyk, K., & Schmid. , C. (2005). A performance evaluation of local descriptors. Pattern Analysis and Machine Intelligence, IEEE Transactions , 27(10):1615 –1630.

26. Monroy, A. P., Montes y Gómez, M., Jair Escalante , H., & González, F. (2016). Categorizacióndeimágenes mediantetécnicasdemineríadetexto. México.

27. Mosquera González, A., Carriera Nouche, M., & Hónzalez Penedo, M. (2011). Recuperación de información: Un enfoque práctico y multidisciplinar. Madrid: Ra-Ma: “Recuperación de imagen”, en Cacheda Seijo, F. et al.(Eds.).

28. MUller, H., Clough, P., Deselaers, T., & Caputo, B. (2010). ImageCLEF: Experimental Evaluation in Visual Information Retrieval. Berlin: Springer-Verlag.

29. PIÑAR, B. A. (2012). Detección automática de objetos en imágenes mediante el uso de puntos característicos . Universidad Autónoma de Madrid .

30. R., J. M. (2004). Aplicaciones clínicas del procesamiento digital de imágenes médicas .

31. Rodriguez, J. M. (12 de 10 de 2012). Modelos de Lenguaje: Introducción a N-Gramas .

32. Sarriá, N. M. (2010). Sistema de clasificación automática de imágenes basado en contenido. MADRID.

33. Soto, R. V. (2014). Implementación de un sistema reconocedor de eventos en videos, con clasificador K-NN. SANTIAGO DE CHILE.

34. Tf-idf weighting. (2008).

35. Weinberg, K., & Saul, L. (2009). Distance metric learning for large margin nearest neighbor classificatio. Journal of Machine Learning Research 10, 207-244.

36. Wu, H., Luk, R., Wong, K., & K., K. (2008). "Interpreting TF-IDF term weights as making relevance decisions". ACM Transactions on Information Systems.

37. Yarza, J. T. (2013). Multi-clasificación Discriminativa de Partes Corporales basada en Códigos Correctores de Errores . Barcelona.