

SISTEMA PARA LA RECUPERACIÓN POR CONTENIDO EN UN BANCO DE IMÁGENES MÉDICAS

**Angel Cruz¹, Clara Spinel², David Seligmann³, Eduardo Romero¹, Fabio Gonzalez¹,
Gloria Diaz¹, Jessica Forero³, Juan Carlos Caicedo¹, Natalia Revelo⁴**

Universidad Nacional de Colombia

¹ Grupo de Investigación Biolngenium ² Facultad de Biología

Ciudad Universitaria, Bogotá D. C, Colombia

fagonzalezo@unal.edu.co

³ **Institución Universitaria Politécnico Grancolombiano**

Facultad de Ingeniería y Ciencias Básicas

Calle 57 # 3 – 00 Este, Bogotá D.C., Colombia

dseligma@poligran.edu.co

⁴ **Centro Internacional de Física**

Edificio “Manuel Ancízar”, Ciudad Universitaria, Bogotá D. C., Colombia

cmspinelg@unal.edu.co

Resumen

El objetivo de este proyecto es desarrollar e implantar un repositorio de imágenes histológicas que será accesible a través de Internet 2 gracias a la interconexión provista por RENATA. El sistema propuesto permitirá almacenar y administrar imágenes de histología para estudiar la estructura de los cuatro tejidos fundamentales en biología. Estas imágenes son un recurso invaluable para la docencia y la investigación, teniendo en cuenta que las imágenes a publicar han sido capturadas con técnicas especiales que facilitan la identificación de ciertas estructuras biológicas que normalmente no se encuentran en libros o artículos científicos. Además de la valiosa recopilación de la colección de imágenes, este proyecto incorporará en el sistema de acceso a las imágenes un módulo de recuperación de imágenes por contenido, que se compone de los métodos y algoritmos para buscar imágenes similares teniendo en cuenta las propiedades visuales de la imagen. En este artículo se presenta la colección de imágenes, la arquitectura del sistema, los modelos propuestos y los avances obtenidos en el proyecto.

Palabras Clave

Imágenes Médicas, Histología, Micro-preparados histológicos, Recuperación de Imágenes por Contenido, Procesamiento de Imágenes.

Abstract

The aim of this project is to develop a histology image retrieval system that is going to be accessible through Internet 2 using the RENATA connection. The proposed system will allow to store and manage histology images, which are useful to study the structure of four fundamental tissues in biology. Those images are an invaluable asset for academic and research activities since they have been acquired using special techniques for marking some biological structures that are usually not found in books or articles. In addition, the system will include a content-based image retrieval module, composed of methods and algorithms to search similar images based on visual properties. This paper presents the image collection, the system architecture, the proposed models and several project advances.

Introducción

El estudio de la histología en la docencia de ciencias biomédicas es fundamental, teniendo en cuenta que los tejidos son la base del estudio de todos los seres vivos al estar constituidos por células que son las unidades de la vida. En investigación los tejidos son básicos y esenciales, por ser los controles que sirven para estudiar cualquier pregunta respecto a biología celular, biología molecular, genética, comportamiento animal o humano. En pocas palabras es la base de la vida y conocer, aprender y estudiar sus variables han hecho que la histología sea curso obligado en todas las profesiones biomédicas. Adicionalmente, la ingeniería de tejidos es un área de investigación creciente en el mundo que se ocupa en lograr la organización de los tejidos y órganos en cultivo, para lo cual se debe conocer muy bien la histología del tejido y la función que se quiere reproducir.

Existen diferentes métodos para estudiar y caracterizar los tejidos biológicos, siendo uno de los más comunes y aceptados las imágenes adquiridas por microscopía. Para poder observar las estructuras del tejido en estas imágenes se utilizan diferentes técnicas de tinción que permiten resaltar las propiedades biológicas a estudiar. Una colección de imágenes de histología con un gran número de ejemplos de los tejidos fundamentales es un importante material de referencia

para la docencia y la investigación en las ciencias biomédicas. En la docencia, la colección de imágenes permite a los profesores seleccionar ejemplos clave para contrastar los conceptos de estudio. De la misma forma los estudiantes pueden utilizar la colección de imágenes para repasar y fortalecer las nociones de histología. Para la investigación una colección de imágenes permitiría a los investigadores comparar imágenes normales con imágenes patológicas y contrastar estructuras obtenidas en los experimentos.

Para ofrecer a la comunidad académica una colección de imágenes que pueda ser aprovechada en la docencia y la investigación, se requiere un sistema que permita acceso efectivo y eficiente a un gran número de imágenes. Los sistemas de información de hoy en día están concebidos para administrar documentos con datos en formatos alfanuméricos, como fechas, direcciones, nombres y textos. Sin embargo, las imágenes contienen más información de la que puede describirse con palabras, por lo cual se requieren métodos de procesamiento automático para identificar patrones, objetos y conceptos en las imágenes. Un sistema que procese y administre las imágenes con este tipo de tecnologías, permitirá a los usuarios buscar imágenes que sean visualmente similares o que compartan determinados conceptos, haciendo más fácil la tarea de encontrar una imagen relevante para las tareas de docencia e investigación.

En este proyecto se plantea el estudio, diseño e implementación de un sistema de búsqueda de imágenes de histología, cuya operación ofrecerá servicio a través de la red RENATA. El primer reto de este proyecto es la captura y digitalización de una colección de 20,000 imágenes de histología que comprenda ejemplos de los cuatro tejidos fundamentales con las estructuras biológicas asociadas. El segundo reto es la construcción de un sistema de búsqueda de imágenes que automáticamente identifique imágenes similares basándose en su contenido visual. Este artículo presenta los principales avances en el desarrollo de este proyecto y algunos conceptos importantes para su desarrollo.

Las imágenes de histología

La histología estudia la morfología de los tejidos que conforman los órganos de un individuo, y su distribución permite distinguir los órganos en sistemas. La histología permite estudiar y definir un tejido sano y compararlo con las modificaciones que sufren cuando presentan anomalías, es decir tejidos patológicos, que se estudian de forma similar a los tejidos normales gracias a los métodos de histotécnica.

Un tejido es el conjunto de células que tiene una morfología similar y ejercen la misma función. Los cuatro diferentes tejidos del cuerpo son: epitelial, conectivo, muscular y nervioso. Estos tejidos se observan bajo el microscopio utilizando los micro-preparados o láminas histológicas de las cuales se pueden capturar imágenes en diferentes aumentos para estudiar sus propiedades. Las imágenes capturadas bajo el microscopio constituyen la colección de imágenes en este proyecto. Estas imágenes no deben tener artificios para evitar una mala comprensión o aprendizaje de los tejidos y órganos a estudiar.

La captura de imágenes se hace teniendo en cuenta que presenten las características que se quieren resaltar y definir en la descripción. La colección de imágenes en este proyecto incluye diferentes imágenes de los cuatro tejidos fundamentales y para cada uno de ellos se resaltan las propiedades relevantes a saber:

- Tejido epitelial: se evidencian células unidas las unas con las otras. El epitelio se divide en dos: de revestimiento y glandular.
- Tejido conectivo: las células están separadas por el material que ellas mismas secretan:

fibras y sustancia fundamental.

- Tejido muscular: se caracteriza por formar fibras musculares y dependiendo de su aspecto se dividen en dos: las fibras de aspecto estriado y las fibras de aspecto liso.
- Tejido nervioso, sus células conforman el sistema nervioso, y de acuerdo a como se organizan se divide en dos: sistema nervioso central y sistema nervioso periférico.

Sistema para Búsqueda de Imágenes por Contenido

Un sistema de recuperación de imágenes por contenido es aquel que utiliza la información visual contenida en las imágenes para decidir los resultados de una búsqueda. El sistema propuesto debe ofrecer esta funcionalidad para facilitar a los usuarios encontrar una imagen relevante entre las 20,000 que componen la colección. El proceso de búsqueda en este sistema seguirá el paradigma de consulta con una imagen de ejemplo, que permite a los usuarios interactuar mediante consultas del tipo “*Busque imágenes similares a esta*”.

El diseño del sistema seguirá una estructura modular que permita extender su funcionalidad y mantener la aplicación en operación. Los principales módulos que componen este sistema son:

- Almacenamiento: Las imágenes deben ser almacenadas e indexadas en el sistema, sin importar el formato particular en el cual hallan sido archivadas.
- Administración: Este módulo contiene las herramientas necesarias para configurar el sistema, datos de los usuarios y permisos de acceso.
- Interacción: Comprende las funcionalidades relacionadas con el manejo de las imágenes, tales como cargar, descargar, describir y eliminar imágenes.
- Búsqueda: Incluye las operaciones asociadas al procesamiento de imágenes y los algoritmos de búsqueda por similitud y despliegue de resultados.
- Visualización: Este módulo incluye interfaces de usuario para ver las imágenes y los datos asociados a ellas.

Uno de los componentes novedosos de este sistema es el módulo de búsqueda que incluye métodos para el procesamiento de imágenes y algoritmos de búsqueda basados en el contenido de las imágenes. Este módulo cumple con dos funciones particulares: extracción de características como colores, bordes y texturas; y cálculo de la similitud utilizada para ordenar los resultados de una búsqueda.

Acceso a la Colección de Imágenes

El sistema en operación debe ofrecer a los usuarios un alto grado de interactividad, es decir que las funciones del sistema tengan una respuesta en tiempo inferior a un segundo. Este requerimiento tiene diferentes connotaciones técnicas que incluyen el formato de las imágenes, la carga computacional para realizar la búsqueda y la velocidad de conexión a la red de datos.

Uno de los principales retos de este proyecto es lograr una aplicación interactiva aprovechando la interconexión de datos provista por la red RENATA. El sistema propuesto debe manejar muchos archivos de gran tamaño teniendo en cuenta las características típicas de una imagen médica. Una lámina de histología completamente digitalizada puede ocupar más de 2GB en disco y la transferencia de estas imágenes puede consumir mucho tiempo. De igual forma el procesamiento y los cálculos en el servidor deben implementarse con estrategias eficientes para garantizar un despliegue oportuno de los resultados.

Resultados Preliminares

En esta sección se presentan algunos de los resultados obtenidos en las etapas completadas de este proyecto. Con respecto a la captura y digitalización de imágenes, se han adquirido un total de 11284 imágenes de histología con los cuatro tejidos fundamentales (epitelial, conectivo, muscular y nervioso) y distintas técnicas de tinción. La organización de las imágenes se realiza utilizando una jerarquía de 115 conceptos de histología para la descripción de su contenido.

Para la visualización y anotación de las imágenes histológicas se ha desarrollado una herramienta Web en lenguaje PHP, agrupando las imágenes por Casos para facilitar su exploración. En cada Caso pueden explorarse un conjunto de imágenes particular, mediante miniaturas de 100x67 píxeles y el número de regiones que lo componen para facilitar la visualización de todas las imágenes del caso simultáneamente.

Caso 4

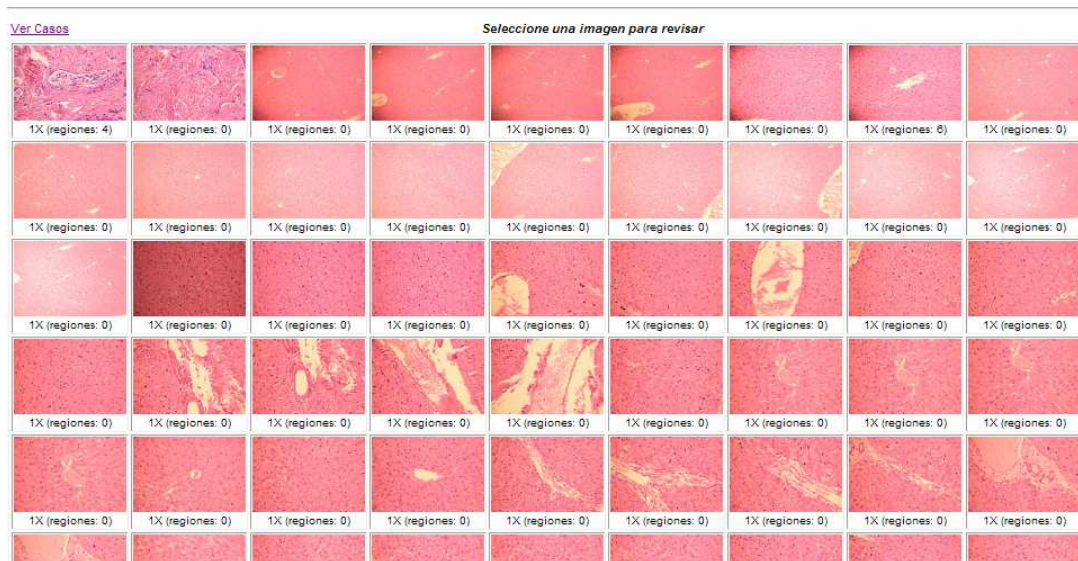


Figura. Interfaz de exploración de imágenes por Caso.

El prototipo inicial para la administración y búsqueda de imágenes histológicas está siendo desarrollado bajo la arquitectura de aplicaciones J2EE en lenguaje Java. La aplicación cuenta con la posibilidad de autenticación segura de distintos usuarios, crear grupos y subgrupos de imágenes para su organización. Además cuenta con la posibilidad de cargar las imágenes individual o masivamente. Finalmente cuenta con las opciones de búsqueda de imágenes por grupo y/o subgrupo o metadato.

Referencias

- [1] Henning Müller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler. A review of content based image retrieval systems in medical applications clinical benefits and future directions. *International Journal of Medical Informatics*, 73:1–23, 2004.
- [2] Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40:262–282, 2007.