

Segmentación de Células Cervicales en Imágenes del Conjunto SIPaKMeD

Camilo Moreno, Juan Muñoz, Santiago Mendivelso

September 2024

1 Introducción

El análisis de imágenes biomédicas se ha convertido en una herramienta crucial para la detección temprana de enfermedades, incluyendo el cáncer cervical. Los avances en técnicas de procesamiento de imágenes han permitido desarrollar algoritmos que segmentan y clasifican células en imágenes microscópicas con mayor precisión. En este proyecto, se aplicarán técnicas de segmentación de imágenes para aislar células cervicales utilizando el dataset SIPaKMeD, contribuyendo al desarrollo de métodos que faciliten la identificación de anomalías celulares. [1]

2 Segmentación

La segmentación es el proceso de dividir una imagen en sus partes o regiones constituyentes. El objetivo de la segmentación es simplificar o cambiar la representación de una imagen para que sea más significativa y más fácil de analizar. En términos simples, se trata de la partición de una imagen en regiones que son homogéneas con respecto a ciertas características, como el color, la intensidad o la textura. [2]

2.1 Tipos de segmentación

Algunos de los tipos de segmentación usados en imágenes y mencionados en el libro guía son:

- Umbralización: Se basa en separar los objetos del fondo utilizando un valor de umbral, donde los píxeles de una intensidad mayor o menor que el umbral pertenecen a diferentes regiones.
- Segmentación basada en bordes: Este enfoque se centra en la detección de discontinuidades en la imagen para identificar las fronteras entre regiones.
- Segmentación basada en regiones: Se agrupan píxeles que tienen propiedades similares, como la intensidad, para formar una región continua.

- Segmentación por agrupamiento: Técnicas como K-means agrupan los píxeles en distintas categorías basadas en sus características. [2]

3 Contexto del problema

El cáncer cervical sigue siendo una de las principales causas de mortalidad en mujeres a nivel mundial. Para su detección, se realizan exámenes microscópicos de células cervicales, los cuales demandan tiempo y son propensos a errores humanos. El dataset SIPaKMeD contiene imágenes microscópicas de cinco tipos de células: parabasal, intermedia, superficial y koilocíticas (estas últimas están asociadas a la presencia del virus del papiloma humano, que está relacionado con el cáncer cervical). [1]

Sin embargo, la identificación precisa de estas células depende en gran medida de una segmentación efectiva que permita aislar las células del fondo, lo cual es el principal desafío que aborda este proyecto.

4 Problema

El principal problema es la segmentación automática de células cervicales en imágenes biomédicas, que presenta dificultades debido a las características visuales similares de las células y el ruido de fondo. Se busca implementar un método eficiente de segmentación que separe las células del fondo, mejorando la precisión de la clasificación posterior.

5 Descripción de Imágenes a Utilizar

Las imágenes a utilizar en este proyecto pertenecen al conjunto de datos SIPaKMeD, el cual contiene imágenes microscópicas de células cervicales en formato de 8 bits y 128x128 píxeles. A continuación se describen algunas de las características clave de las imágenes:

Imágenes en escala de grises y color: Las imágenes pueden estar en color o en escala de grises, dependiendo del preprocesamiento. Cada imagen contiene una o más células que deben ser segmentadas. Diversidad de tipos celulares: Las imágenes muestran distintos tipos de células, incluidas células anormales (koilocíticas), cuya detección es clave para el diagnóstico temprano del cáncer cervical. [1]

6 Análisis del Dataset

Para el análisis del dataset SIPaKMeD, se considerarán varios aspectos clave en las imágenes celulares. Estos aspectos incluyen:

- Análisis del pigmento: Se analizarán los tres canales de color (RGB) en cada imagen para estudiar la distribución del pigmento en las células. Este

análisis es crucial, ya que las variaciones en la intensidad de los canales pueden proporcionar información morfológica importante para distinguir entre diferentes tipos celulares.

- Detección de bordes: Se emplearán técnicas de detección de bordes para identificar con precisión los contornos celulares. La correcta detección de los bordes permitirá definir los límites de las células, lo cual es fundamental para la segmentación y clasificación.
- Análisis del punto central: Se analizará el punto central de cada célula, que generalmente corresponde al núcleo celular, evaluando características como la densidad, textura, y sobre todo el tamaño del núcleo. El tamaño del núcleo es un factor relevante que puede ayudar a determinar la clasificación celular, ya que puede presentar diferencias entre tipos de células.
- Tamaño y forma de la estructura externa: Se evaluará el tamaño y la forma de las estructuras externas al núcleo, principalmente el citoplasma. Este análisis ayudará a identificar variaciones morfológicas entre diferentes tipos celulares.

Este enfoque integral basado en el análisis del color, los bordes, el núcleo y la morfología celular externa tiene como objetivo mejorar la precisión en la clasificación de las células.

7 Posible resultado

Para realizar nuestro objetivo de segmentar las células utilizaremos los siguientes métodos:

- Umbralización Adaptativa: Este método calcula un umbral local para cada pequeño bloque de la imagen, lo que permite separar las células del fondo a pesar de las variaciones de iluminación. La imagen binaria resultante puede mejorar la detección de células con bordes claros. [2]
- Algoritmo Watershed: Este algoritmo es eficaz para separar células que están superpuestas o muy cercanas entre sí. Se basa en un enfoque de detección de contornos mediante la simulación de un flujo de agua que "inunda" la imagen, separando las células en regiones distintas. [3]
- Filtros de Detección de Bordes (Canny o Sobel): Los filtros Canny o Sobel pueden utilizarse para resaltar los bordes de las células, ayudando a mejorar la detección de los contornos celulares. Estos bordes pueden luego combinarse con el método Watershed para una segmentación más precisa. [2]
- Postprocesamiento (Operaciones Morfológicas): Para eliminar pequeños fragmentos de ruido y mejorar las regiones segmentadas, se aplicarán operaciones morfológicas como dilatación y erosión, que refinan la forma de las células segmentadas. [2]

8 Referencias Bibliográficas

1. Plissiti, M. E., Dimitrakopoulos, P., & Nikou, C. (2018). SIPaKMeD: A New Dataset for Feature and Image-Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images [Dataset]. IEEE Dataport. <https://doi.org/10.21227/H29W-9H39>
2. Gonzalez, R. C., & Woods, R. E. (2017). *Digital Image Processing* (4th ed.). Pearson Education.
3. Roerdink, J. B., & Meijster, A. (2000). *The Watershed Transform: Definitions, Algorithms and Parallelization Strategies*. Fundamenta Informaticae, 41(1-2), 187-228.