Image informatics (CSC8628) - Technical Project

200709330 - Hammad Mir

01. Aim

The aim of this technical project is to design, implement and test an image informatics approach for automatic colon cancer cells segmentation in microscopy imagery using image informatics methods/techniques covered in the Image Informatics (CSC8628) module’s lectures and practicals only. To assess the accuracy and the quality of the cells (nuclei and cytoplasm) segmentation approach, we compare the nuclei and cytoplasm segmentation results with the golden standard manual nuclei and cytoplasm segmentation results, using Jaccard Score metric.

02. Introduction

2.1. Image Segmentation

Image segmentation, also known as image regions or image objects, is a technique in digital picture processing and computer vision that divides a digital image into several image segments (sets of pixels). The purpose of segmentation is to make a picture more intelligible and simpler to examine by simplifying and/or changing its representation. In most cases, image segmentation is used to find objects and boundaries (lines, curves, and so on) in pictures. Picture segmentation, to put it another way, is the process of giving a label to each pixel in an image such that pixels with the same label share certain properties.

Image segmentation produces a set of segments or contours that encompass the whole image. In terms of some characteristic or calculated feature, such as colour, intensity, or texture, each of the pixels in a region are comparable. With regard to the same attribute, adjacent patches have dramatically different colour (s).

03. Image Segmentation: Purpose

Cancer has traditionally been a fatal disease. Even in today's technologically advanced world, cancer may be lethal if it is not detected early enough. Early detection of malignant cells has the potential to save millions of lives.

The morphology of the malignant cells is crucial in evaluating cancer's aggressiveness. Object detection will not be particularly useful here, even if we have put the parts together. Only bounding boxes will be generated, which does not assist us in determining the form of the cells.

Image segmentation techniques have a significant influence in this case. They enable us to take a more detailed approach to the problem and provide more useful outcomes. It's a win-win situation for everyone involved in the healthcare business.

Some of the practical applications of image segmentation are:

- Content-based image retrieval

- Machine vision
- Medical imaging, including volume-rendered images from computed tomography and magnetic resonance imaging.
    - Locating tumors and other pathologies, Measuring tissue volumes, Diagnosis and study of anatomical structure, etc.
- Object detection
    - Pedestrian detection, Face detection, Brake light detection, Locating objects in satellite images (roads, forests, crops, etc.)
- Recognition Tasks
    - Face recognition, Fingerprint recognition, Iris recognition, etc.
- Traffic control systems
- Video surveillance
- Video object co-segmentation and action localization

Image segmentation has spawned a slew of general-purpose algorithms and approaches. In order to properly tackle the domain's segmentation difficulties, these strategies must often be paired with domain-specific information.

04. Image Segmentation: Approaches

$$\text{Segmentation based on: } \begin{cases} \text{(1) Discontinuities} \begin{cases} \text{(1) Isolated points} \\ \text{(2) Lines} \\ \text{(3) Edges} \end{cases} \\ \text{(2) Similarity} \begin{cases} \text{(1) Thresholding} \\ \text{(2) Region growing} \\ \text{(3) Region splitting/merging} \end{cases} \end{cases}$$

4.1. Similarity approach:

The similarity method detects similarities between picture pixels to produce a segment based on a threshold. This sort of method to picture segmentation is used by ML algorithms such as clustering.

4.2. Discontinuity approach:

The discontinuity technique is based on the image's pixel intensity levels being discontinuous. This strategy is used by Line, Point, and Edge Detection algorithms to produce intermediate segmentation findings that may then be processed to create the final segmented picture.

05. Image Segmentation: Techniques

There are a slew of image segmentation techniques that can be used to segment images. The most common techniques are:

- Threshold Based Segmentation
- Edge Based Segmentation
- Region-Based Segmentation
- Clustering Based Segmentation
- Artificial Neural Network Based Segmentation

For the scope of this project, we will be using the Threshold and Clustering-based Segmentation technique.

5.1. Threshold Based Segmentation

Thresholding is an image segmentation technique in which we modify the pixels of a picture to make it simpler to examine. Thresholding is the process of converting a colour or grayscale image into a binary image, which is just black and white. Most commonly, we utilise thresholding to identify regions of interest in an image while disregarding the sections we don't care about. There are various thresholding methods, which are used to determine the threshold value for image segmentation.

5.2. Clustering Based Segmentation

Clustering is the process of separating the population (data points) into a number of groups so that data points in the same group are more comparable to other data points in the same group than data points in other groups. Clusters are the name given to these groups.

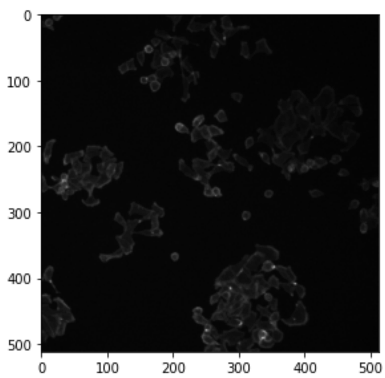
There are various clustering methods used for image segmentation. The most popular being k-means clustering. aside from k-means clustering, there are other clustering methods such as the Birch algorithm, agglomerative clustering, mean-shift and DBSCAN, which are used for image segmentation.

06. Dataset

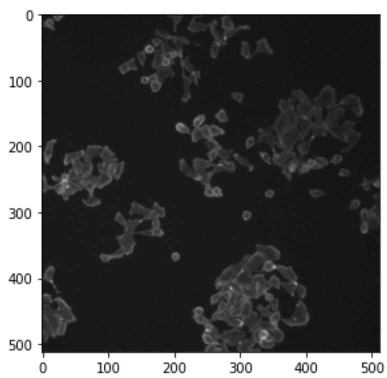
The dataset consists of separate cytoplasm and nuclei colon cancer cells microscopic images. There are 12 images for each of the two classes and the golden standard manual nuclei and cytoplasm segmentation results (masks) are also provided. Each image is a 512x512 pixel grayscale image.

07. Our Approach

7.1. Image Preprocessing



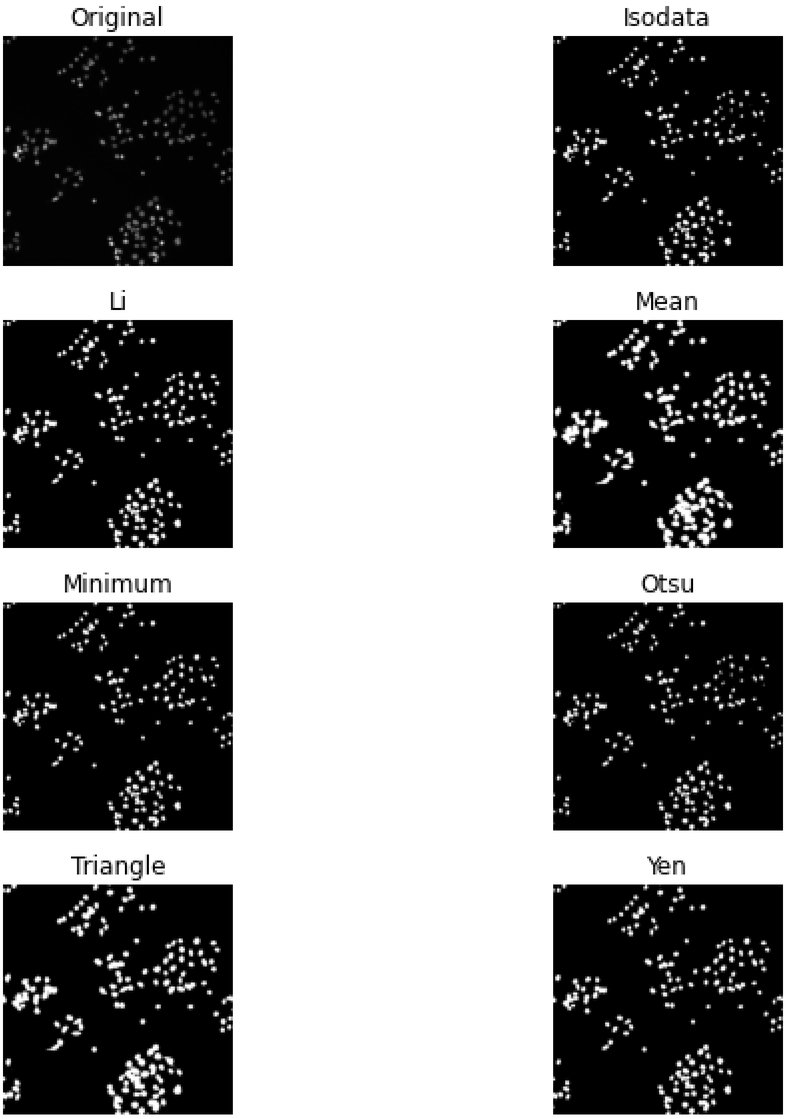
Original Image



Preprocessed Image

The images were preprocessed for better visualization. We explored the effect of various image preprocessing techniques on the raw images, including image normalization, image contrast enhancement, image smoothing, image noise removal, and image histogram equalization. We found that image normalization and image contrast enhancement using gamma adjust using a gamma of 0.6 to be the most effective for image visualization and segementation.

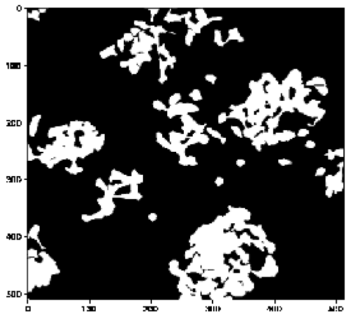
7.2. Image Segmentation: Thresholding



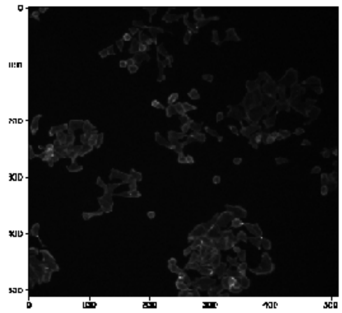
We began our image segmentation process by applying a basic recursive threshold to the images. Following this, we applied various other thresholding techniques to the images. We found that the thresholding technique with the highest accuracy was the Mean and Multi Otsu's methods.

7.3. Image Segmentation: K-means Clustering

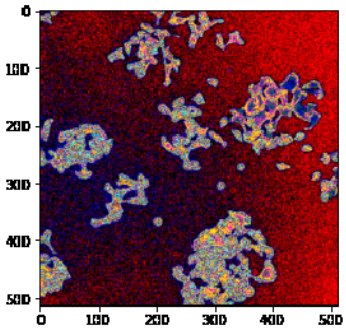
We further explored the effect of clustering techniques on the images such as k-means clustering, mean-shift segmentation, watershed segmentation and graph-cut clustering. We found that k-means clustering was the most effective for all the segmentation methods. We decided on k-means clustering to segment the images for the final comparison.



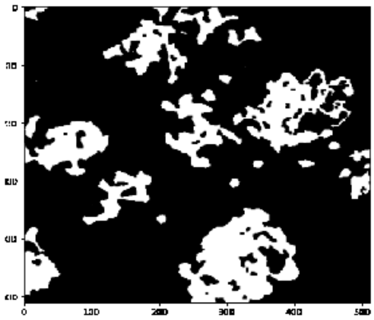
Ground Truth



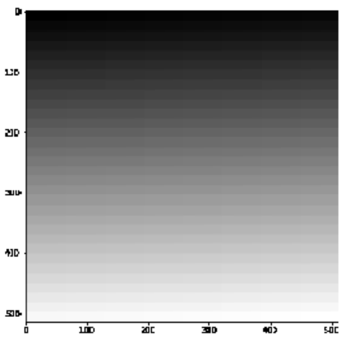
Original image



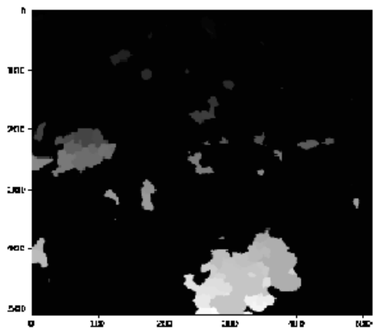
Mean-shift



k-means



graph-cut



watershed

08. Image Post-Processing

After the segmentation, we post-processed the images to remove noise and smooth the edges with median filtering and region filling. We found that these operations did not improve the overall segmentation.

09. Results

Upon comparing the segmentation results of the two methods on cytoplasm and nuclei images separately, we found that the mean thresholding on preprocessed images worked the best for cytoplasm images with an average jaccard score of 0.86. Consecutively, the multi Otsu's thresholding on preprocessed image worked the best for nuclei images with the average jaccard score of 0.88, beating the average jaccard score of 0.76 from k-means clustering.