International Journal of Computing and Business Research (IJCBR)
ISSN (Online) : 2229-6166
International Manuscript ID : 22296166V10I3202001
Volume 10 Issue 3 September - October 2020

K-Means and Image Processing Use in Histopathological Image Analysis

Faiez Musa Lahmood Alrufaye

Technical Institute/ Kut,
Middle Technical University, Baghdad, Iraq
Faiez.Alrufaye@mtu.edu.iq

Abstract- Image processing has entered most areas of life, and it would be negligent not to take advantage of these techniques in the most important of those areas, which have a direct link to humans and their health.

These techniques can distinguish human cells and detect the change in them through the processing of these images, passing them through several stages.

In our current research we highlight these images using image processing techniques, from the initial stages of that processing through pre-treatment, improved by removing unintended distortions and then analyzing to extract their important characteristicsfor the purpose of use in the comparison of those images.

We used the K-means algorithm for the purpose of grouping those images using its grouping function.

Keywords- *K-Means,Image Processing, Histopathological Image, Image analysis*

I. INTRODUCTION

The use of computer technologies has become common in most areas of life because the techniques offer easy access to accurate, fast results. The most important of these techniques is image processing technology.

Given the advances of modern medicine, it has become possible to use image processing techniques in this area.

Histology is a science that deals with the cellular structure of organisms. Knowledge of the biological structure is important in the knowledge and diagnosis of disease.

In conventional methods, the biological composition of the cell is determined by examining it under a microscope.

Today, due to the great advances in image processing techniques and the emergence of sophisticated image scanners to obtain pathological images, digital images are available. By using digital image processing algorithms, computer-assisted diagnosis is much easier thanusing conventional methods. Image processing techniques and different image processing signals can be used to analyze images for diagnosis and prognosis. Therefore, histology provides the scientific basis for clinical research, education, and practice.

The aim of this research is to use and develop powerful and accurate computer-assisted image analysis algorithms for pathological anatomy images. Different image processing techniques will be applied to classify image tissue and determine cell type or classification and to derive quantitative measurements of disease features from histological images and to determine whether the tissue is normal or there is a change in it, indicating the presence of a disease.

II. LITERATURE SURVEY

Several theories have been proposed to explain image recognition techniques, including animage processing algorithm designed by Belsare and Mushrif, developedfor cancer detection and classification[1]. Histopathology image processing by using different colour models (RGB, HSI, CMYK, CIEL*a*b*, and HSD) applied to very large microscopic image analysis, is a model developed by Déniz, Salido, Fernández, Vállez and García-Rojo in 2013[2].

In 2015, Vaishali, R. Ramesh and J. Anita Christaline detected cancer based on a huge collected database and use of 2D autoregressive (AR) model in automated cancer diagnosis based on histopathology[3].

Hagerty, Stanley and Stoecker from Missouri University of Science and Technology developed a model by using deep learning to analysis Medical Images, where they depended on neural networks to get perfect results, that was in 2017[4].

In 2018, Ramakrishnan Mukundan designed and implemented a system to explore different types of Characteristics measurements that are suitable for the automated scoring of human epidermal growth factor receptor 2 (HER2) in histology slides[5].

In the same year, the researcher Kevin Kiambe developed a model by using a neural network to Support Vector Machine (SVM), k-Nearest Neighbor (k-NN) and Logistic Regression (LR) models for classifying breast histopathological images[6].

In our research, we will depend on image processing techniques to analyse medical images to get accurate results.

III. K-MEANS AND IMAGE PROCESSING SYSTEM

Image processing models are divided into four primary phases: image formation, enhancement, visualization and analysis[7].Since there are many sources that have talked about the first, second and third phases, we will limit our current research to the fourth phase only. Figure (1) illustrates the phases of Image processing models.

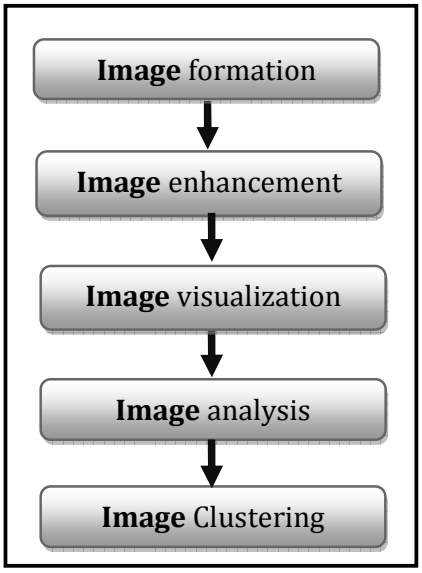

Figure 1: Phases of Image processing models.

1. Image Analysis:

At this stage, the system uses a set of algorithms to apply a variety of techniques such as extraction, segmentation and classification.The results of the system depend mainly on the results of this stage. Where the results of this phase are accurate the results of the system are also accurate.This stage consists of a set of steps namely Preprocessing, Segmentation, Feature extraction, Classification and Interpretation[8].

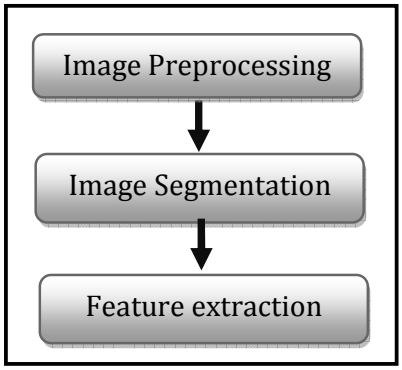

Figure 2: Techniques of image analysis.

A. Image Preprocessing

This technique is very important as it can make the image more suitable for later stages by removing unintended distortions or clarifying the image further.

B. Image Segmentation

The most important step in the image processing technique is Image segmentation. Its main goal is to segment an image into parts that have a strong correlation with objects or areas of interest. Segmentation can be improved by three different techniques: threshold, edge-based segmentation, and region-based segmentation.

C. Feature extraction

Feature extraction is the basis of our system; where the system depends entirely on the previously mentioned stages. Image features can be classified as a general or specific domain, and usually include color, texture, shape, drawing, spatial characteristic, and distortion. Beyond that, is any specific domain. Each feature may have multiple representations; for example, both colored graph and colored moments are considered as an image color property.

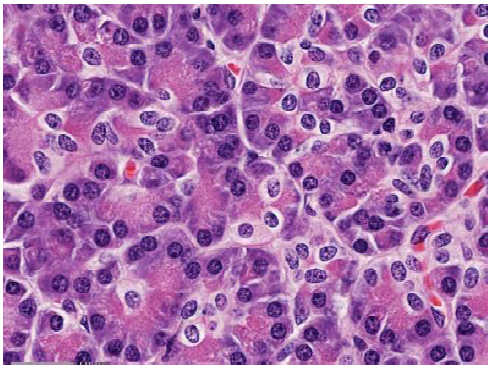

Figure 3: Histology image of normal human pancreas

2. Image Clustering

Clustering is an Unsupervised Machine Learning method. It collects images together which have similar features, in clusters, using a Similarity Function. The number of clusters is generated automatically depending on the nature of the input image. The inputs to this stage are a set of features. They are collected accordingly to create Clusters. In this stage, the distance between the images is calculated by summing the weight of these feature, multiplied by the incompatibility of the feature between the first image and the next It is calculated by the following equation[9]:

$$\text{Dist(NPi,NPj)} = \sum_{f \in F} I_f * incompatibility_f(NPi, NPj)$$

When:

NPi represents the first image. NPj represents the next image.C a set of features used to determine the distance, I_f represents the weight of each characteristic.

Incompatibility is represented by 0 if there is no incompatibility between the first image and the next, else it will represented by 1.

The weight and the incompatibility are very important to minimizing or maximizing the distance between them. The distance relationship with similarity is an inverse relationship:

(when the distance maximizes then the similarity will minimize. The reverse is also correct).

IV. CONCLUSION

Image processing techniques in analysis are crucial to obtaining critical results in order to analyze those images and obtain their features that can then used in the classification and clustering phase. Using traditional methods of image processing, the results were fairly good, but with an algorithm of K-means, the results were very accurate - an outcome of the power of comparing the features of the input images to the system, using the power of K-means to compare those features. We may get a stronger comparison using natural intelligence techniques such as neural networks or fuzzy logic.

REFERENCES

[1] A. D. Belsare and M. M. Mushrif, "*Histopathological Image Analysis Using Image Processing Techniques: An Overview*", Signal & Image Processing : An International Journal (SIPIJ) Vol.3, No.4, August 2012.\
[2] Oscar Déniz, Jesús Salido, M. Milagro Fernández, Noelia Vállez and Marcial García-Rojo, "*Colour Model Analysis for Histopathology Image Processing*", Color Medical Image Analysis pp 165-180, Springer 2013.
[3]D. Vaishali, R. Ramesh, J. Anita Christaline, " *Histopathology Image Analysis and Classification for Cancer Detection Using 2D Autoregressive Model*", International Review on Computers and Software (IRECOS) Vol.10, No.2, (2015).
[4] Jason Hagerty, R. Joe Stanley, William V. Stoecker, "*Medical Image Processing in the Age of Deep Learning Is There Still Room for Conventional Medical Image Processing Techniques?*", The 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017), pages 306-311, 2017.
[5] Ramakrishnan Mukundan, "*Analysis of Image Feature Characteristics for Automated Scoring of HER2 in Histology Slides*", 22nd Conference on Medical Image Understanding and Analysis, Southampton, UK, pages 9–11 July 2018.
[6]Kevin Kiambe, "*Breast Histopathological Image Feature Extraction with Convolutional Neural Networks for Classification*", ICSES Transactions on Image Processing and Pattern Recognition (ITIPPR), Vol. 4, No. 2, June 2018.
[7] Dr.R.Uma Rani and Miss.P.Amsini, " *Image Processing Techniques Used In Digital Pathology Imaging: An Overview*", International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 5, Issue 1, January 2018.
[8] Bosoon Park and Renfu Lu, " *Hyperspectral imaging technology in food and agriculture*", Springer, New York, Pages 9-56, NY 2015.
[9] Dhyaa Sh. Al-Azzawy and Faiez M. Alrufaye, " *Arabic Words Clustering by Using K-Means Algorithm*", Annual Conference on New Trends in Information & Communications Technology Applications-(NTICT'2017) Pages 263-267, 2017.