

An Efficient Algorithm for Human Cell Detection in Electron Microscope Images Based on Cluster Analysis and Vector Quantization Techniques

Mohammad V. Malakooti
Hamidreza Naji
Ahmad Pahlavan Tafti



Titles:

- **Abstract**
- **Introduction To Human Cell Detection Through Image Processing Techniques**
- **Cluster Analysis**
- **Vector Quantization**
- **Proposed Algorithm**
- **Experimental Results and Discussion**
- **Conclusion**

Abstract

- Automatic detection of human cell is one of the most common investigation methods that may be used as part of a **computer aided medical decision making system**.
- In this paper we present an efficient algorithm, based on the **cluster analysis** and the **vector quantization** techniques for human cell image detection.
- Our proposed algorithm is applied on **two sample datasets** from our research laboratory and also **Imamreza laboratory** in Mashhad which contain 196 number of normal electron microscope images. Experimental results show that this model is both accurate and fast with a **detection rate of around 86.69 percent**.
- Our proposed method **does not require any under segmentation**.

Introduction To Human Cell Detection Through Image Processing Techniques

- Humans contain about 100 trillion cells, each working together in a complex symphony of interactions to perform everything from providing structure and stability and providing energy and means of reproduction for an organism.
- A typical cell size is $10\ \mu\text{m}$ and a typical cell mass is 1 nanogram.

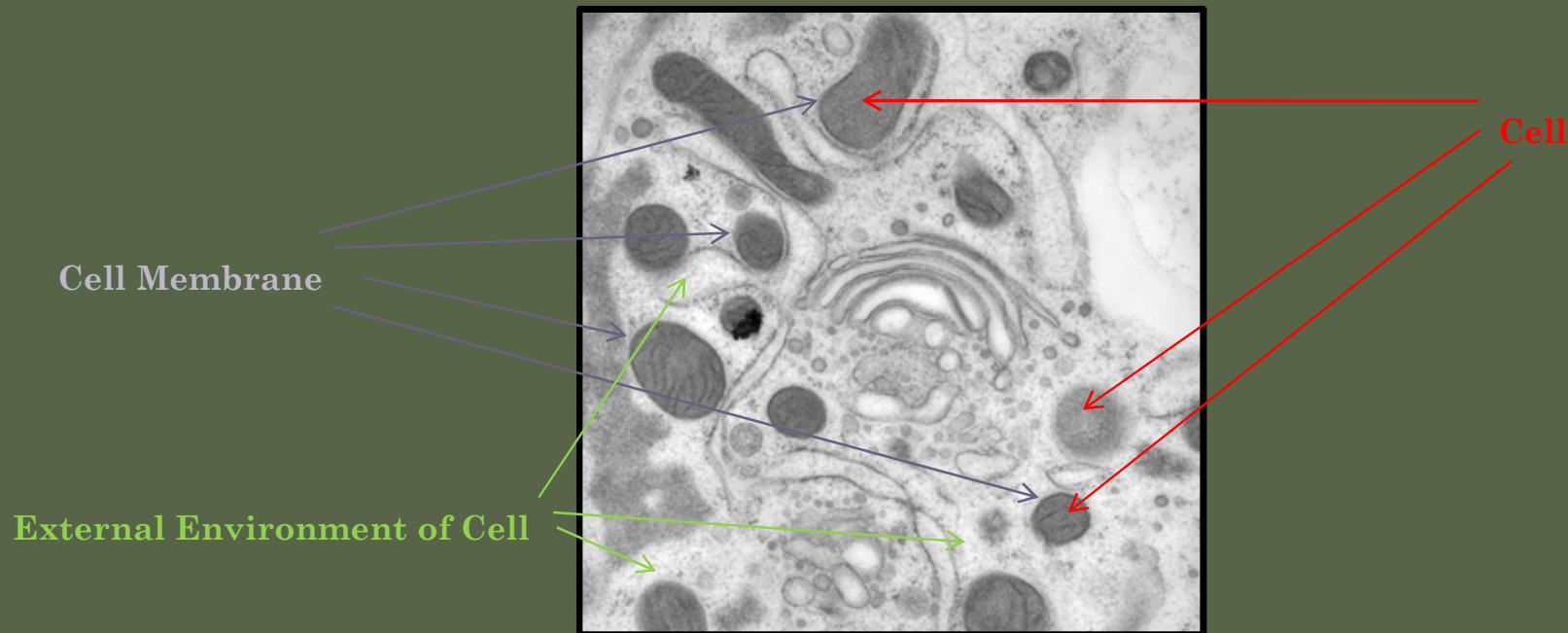


Figure 1. Human Cell in Microscopic Image

Introduction To Human Cell Detection Through Image Processing Techniques

- **Image segmentation** and **object detection** plays an important role in many medical imaging systems, yet in complex circumstances it is still a challenging problem [1]
- Some common digital image processing techniques that are developed for human cell detection are based on image **thresholding**, which could be implemented by **pixel classification** [2]
[3]
- There are other approaches that use region based information, but they are not applicable because darker background regions may be misclassified as cells and lighter cell regions may be classified as background [4]

Introduction To Human Cell Detection Through Image Processing Techniques

- In this paper we proposed an efficient algorithm to detect the human cell in an accurate manner.
- Then we use **vector quantization technique** to **clustering the elliptical approximation of human cells**.
- Our proposed algorithm is applied on two sample datasets from Imamreza laboratory and also our research laboratory in Mashhad which contain 196 number of normal electron microscope images. Experimental results show that this model is both accurate and fast with around 86.69 percent of detection.

Snapshot of Our Model

Input:

Electron Microscope Images of Cell

Pixel Values Determination of Cell, Cell Membrane and External Environment of Cell

Calculating Median, Mean and Variance of Those Particular Pixels

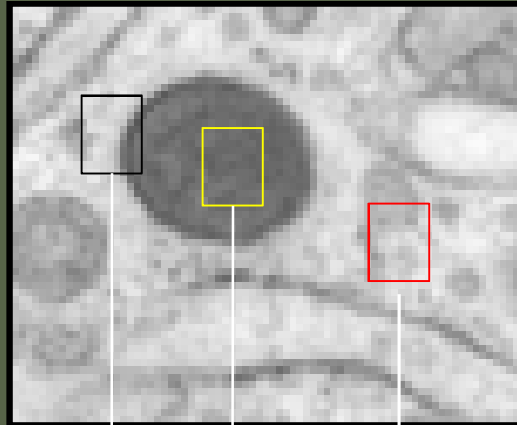
Performing Vector Quantization on Pixel Values

Generating **CodeBook** and **CodeVector** based on Pixel Values on Cell, Cell Membrane and External Environment of Cell

Storing the
CodeBook and
CodeVector
in a Database

Figure 2. Our Model Snapshot

Our Sample Dataset ; Training Data



Cell

Cell Membrane

External Environment of Cell

Pixel Values

	Cell	Cell Membrane	External Environment of cell
Sample Image 1	74	58	129
Sample Image 2	81	53	133
Sample Image 3	70	49	121
Sample Image 4	71	51	117
Sample Image 5	66	59	117
Sample Image 6	71	59	123
Sample Image 7	62	46	119
Sample Image 8	69	43	130
Sample Image 9	83	59	128
Sample Image 10	81	41	119
Sample Image 11	71	44	126
Sample Image 12	67	53	139
Sample Image 13	69	57	137
Sample Image 14	81	55	126
Sample Image 15	75	43	119
Sample Image 16	79	52	134

What is Clustering ?

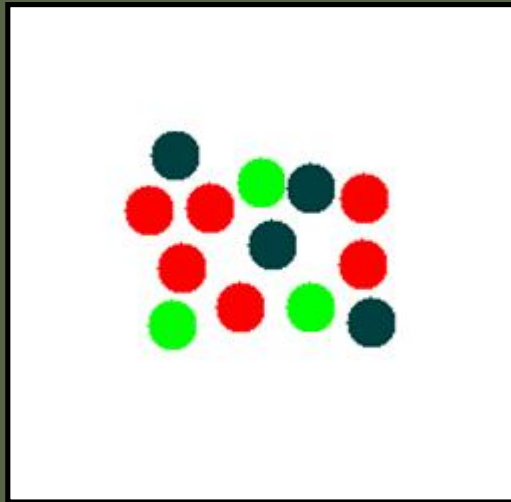


Figure 3: Original Data

Clustering

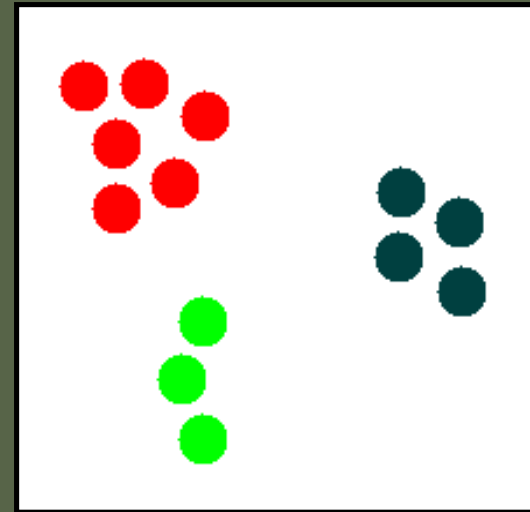


Figure 4: Clustered Data

Cluster Analysis

- Cluster analysis is the task of assigning a set of objects into groups.
- The objects in the same cluster are more similar to each other than to those in other clusters [5].

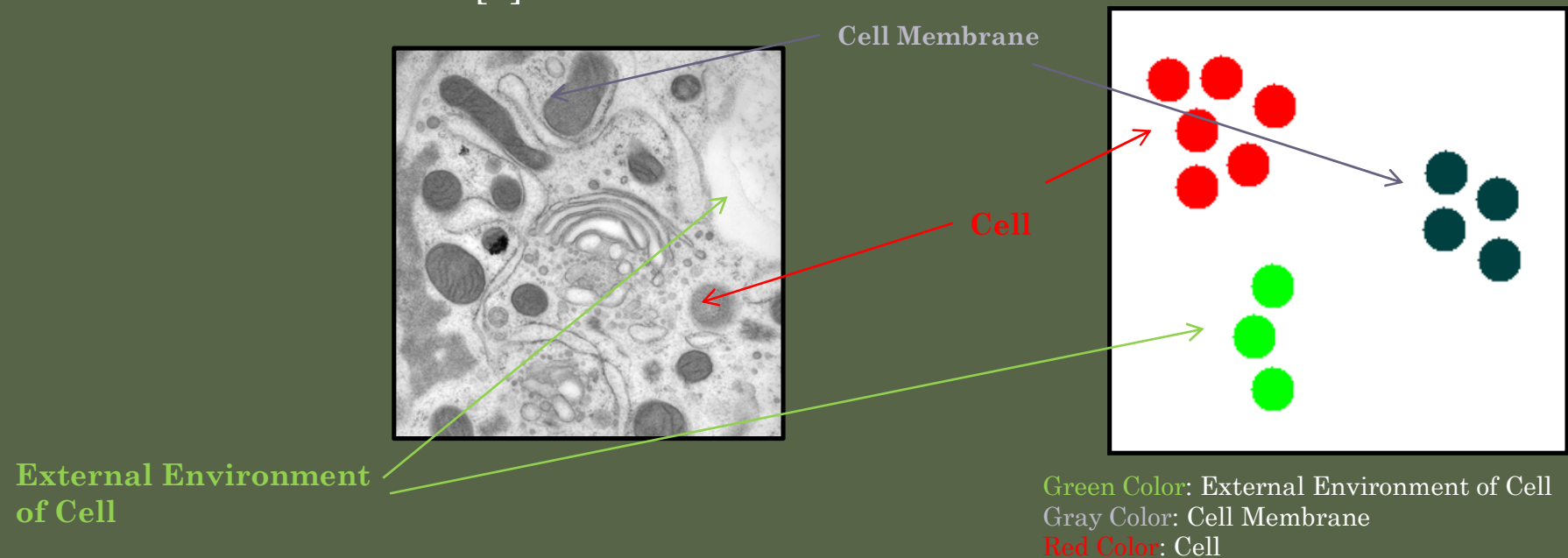


Figure 5: Example of **Clustering** on Cell Electron Microscope Image

Cluster Analysis

- Cluster analysis can be achieved by various algorithms, but popular notions of clusters include groups with low distances among the cluster members, dense areas of the data space or particular statistical distributions.
- It will be necessary to modify preprocessing and parameters until the result achieves the desired attributes [6].
- In **medical imaging**, cluster analysis can be used to **differentiate between different types of tissue in an electron microscope image**. In this type of applications, actual position is not important, but the object intensity is considered as a vector.
- Clustering techniques are mainly classified into hard clustering and soft clustering approaches. While in the hard clustering, each object belongs to a cluster or not, in soft clustering each object belongs to each cluster to a certain degree [7].

Vector Quantization

- Vector quantization (VQ) is a **classical quantization technique** from **signal processing** which allows the modeling of **probability density functions** by the distribution of prototype vectors [9].
- It could divide a large set of points (vectors) into groups having approximately the same number of points closest to them [8].
- A vector quantization is nothing more than an approximator. The idea is similar to that of rounding-off (say to the nearest integer) [9].
- Vector quantization has been very popular in different types of research fields such as **image segmentation** [10], **speech recognition** [11], **content based image retrieval** [12] and **face recognition** [13].

Vector Quantization

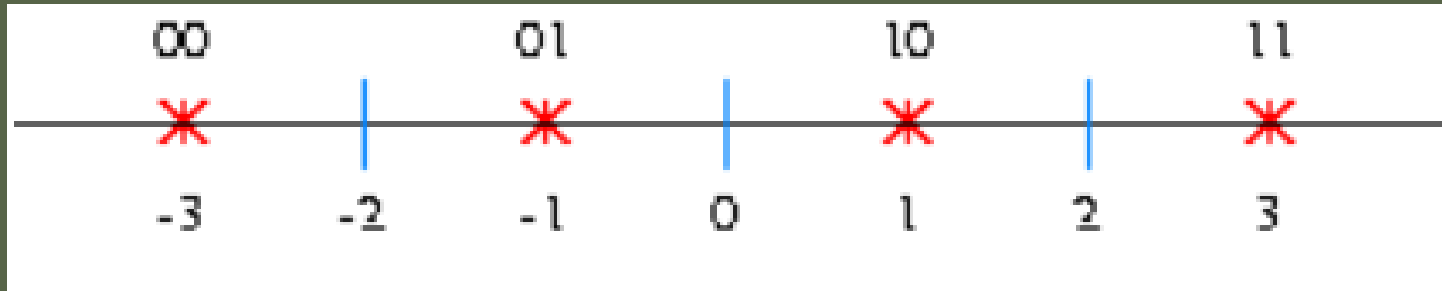


Figure 6: Example of one Dimensional VQ.

In this figure, every number less than -2 are approximated by -3. Every number between -2 and 0 are approximated by -1. Every number between 0 and 2 are approximated by +1 and every number greater than 2 are approximated by +3.

Note that the approximate values are uniquely represented by 2 bits (00, 01, 10, and 11). This is one dimensional, 2-bit vector quantization.

In this figure, the red stars are called *codevectors* and the regions defined by the blue borders are called *encoding regions*. The set of all codevectors is called the *codebook* and the set of all encoding regions is called the *partition* of the space [8].

Vector Quantization

- To design a VQ we need a source vector with its statistical properties, the number of codevectors, find a codebook (the set of all red stars) and a partition (the set of all blue lines) [8].

- Assume that, we have a training sequence consisting of n source vectors as follow:

$$\mathbf{T} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$$

This training sequence can be obtained from a database. For example, if the source is a digital image, then the training sequence can be obtained by recording the statistical properties of that. Assume that the source vectors are k dimensional.

$$\mathbf{X}_n = (\mathbf{X}_{n,1}, \mathbf{X}_{n,2}, \dots, \mathbf{X}_{n,k}), \quad n=1,2,\dots,N.$$

Vector Quantization; An Example

Source Vector
CodeVectors

0 0 01 0 0 1 01 1 0 01 01 0 1




CodeBook	
	0
	1
	01

Figure 7: Example of Codevectors and Codebook

Proposed Model

- The main purpose of the proposed algorithm is to **detect the human cell in electron microscope images**.
- We explain our detection algorithm by **Linde Buzo and Gray segmentation method**. It consider as a **VQ method to generating both codevectors and codebook**. These codevectors and codebook are based on image's **textual properties** from grayscale features such as **median, mean** and **variance**.
- The LBG VQ (Linde Buzo and Gray Vector Quantization) algorithm is an iterative method [9].

Proposed Model

- For the purpose of explaining our algorithm for human cell detection, we are considering a 4×4 digital image (Figure 8).

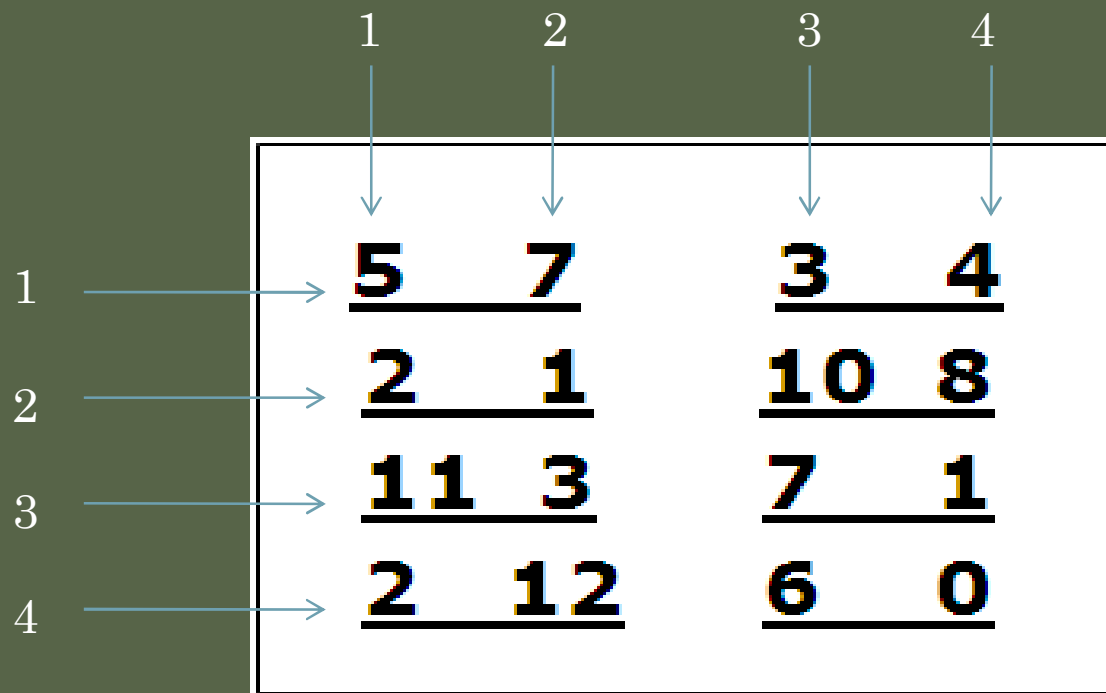


Figure 8: A sample 4×4 digital image.

Proposed Model

- Now, we are designing two dimensional vector space as shown in figure 9.

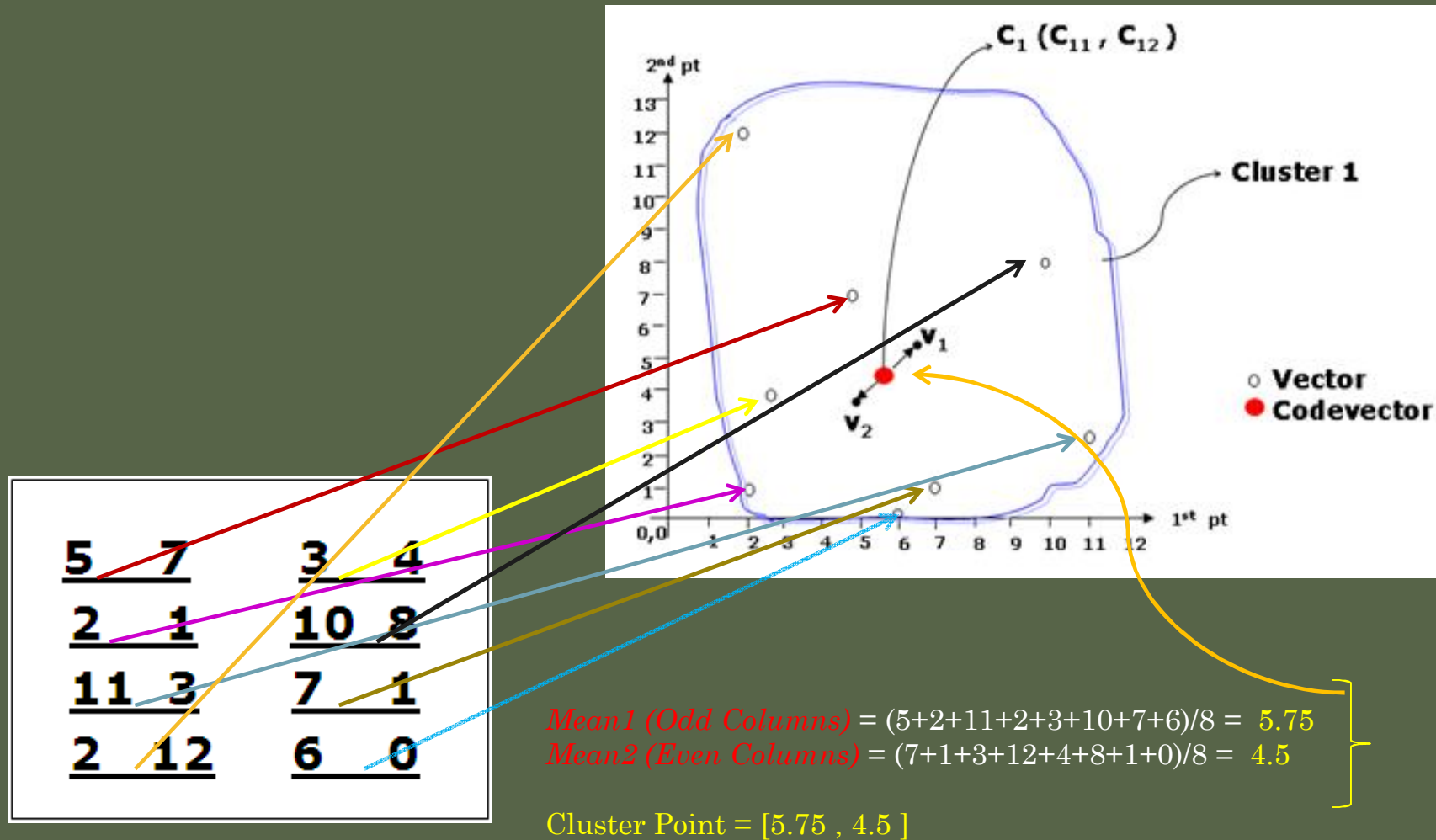


Figure 9: Two dimensional vector space

Proposed Model

Clustered
Line

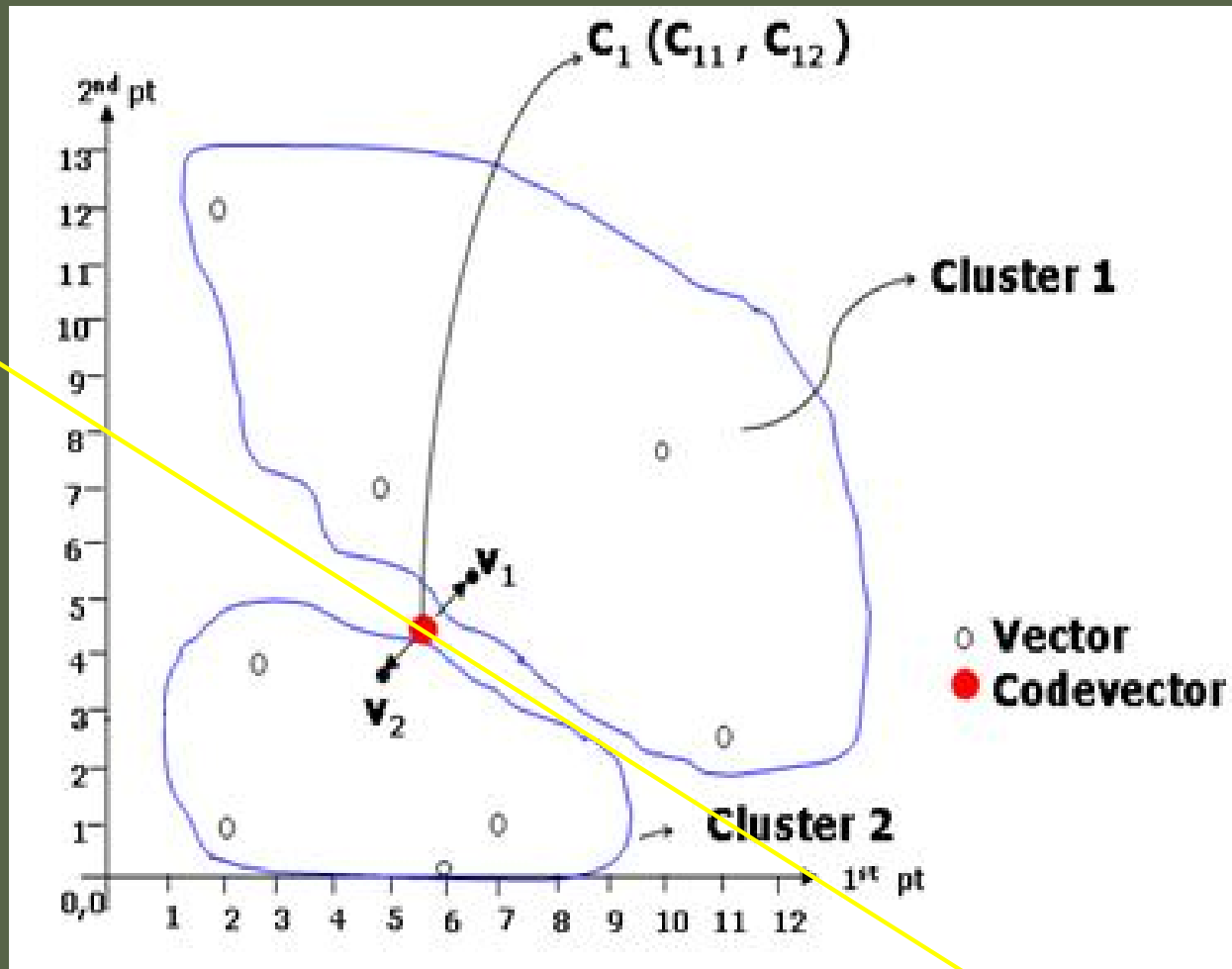


Figure 10: 1st iteration of LBG for this two dimensional vector space.

Proposed Model

Clustered
Line

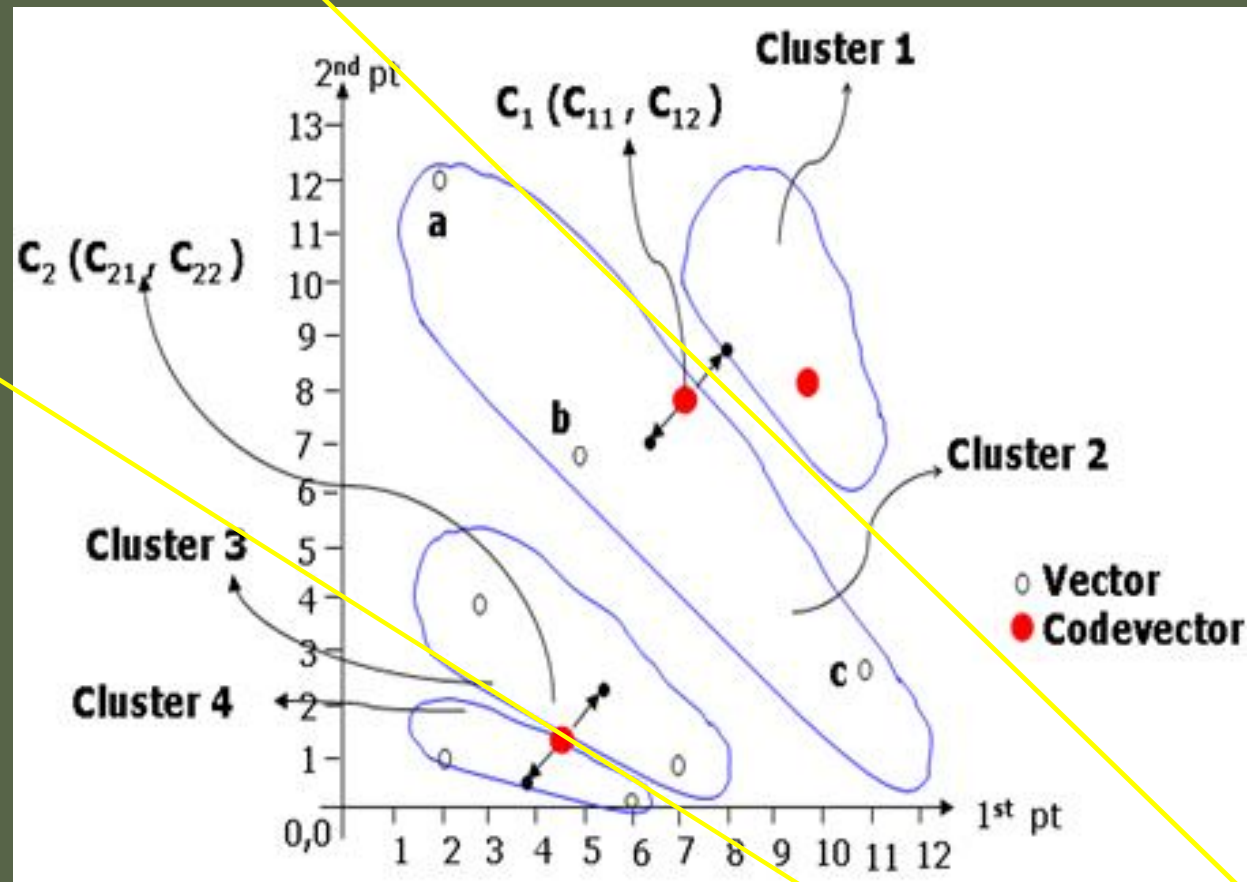


Figure 11: 2nd iteration of LBG for this two dimensional vector space.

Proposed Model

- The iteration is repeated for all new clusters until the specified Mean Square Error (MSE) is reached. MSE measures the average of the squares of the "errors." The error is the amount by which the value implied by the estimator differs from the quantity to be estimated [14].
- In our proposed method to human cell detection, and based on our sample dataset, a codebook of size 128 is generated for the given images. Codevectors are clustered in 8 clusters using the same LBG VQ algorithm.
- Median** is described as the numeric value separating the higher half of a sample, a population, or a probability distribution, from the lower half [14], **Mean** is a method to derive the central tendency of a sample space [14] and In probability theory and statistics, the **variance** is a measure of how far a set of numbers is spread out [14].

Proposed Model

Input:

Electron Microscope Images of Cell

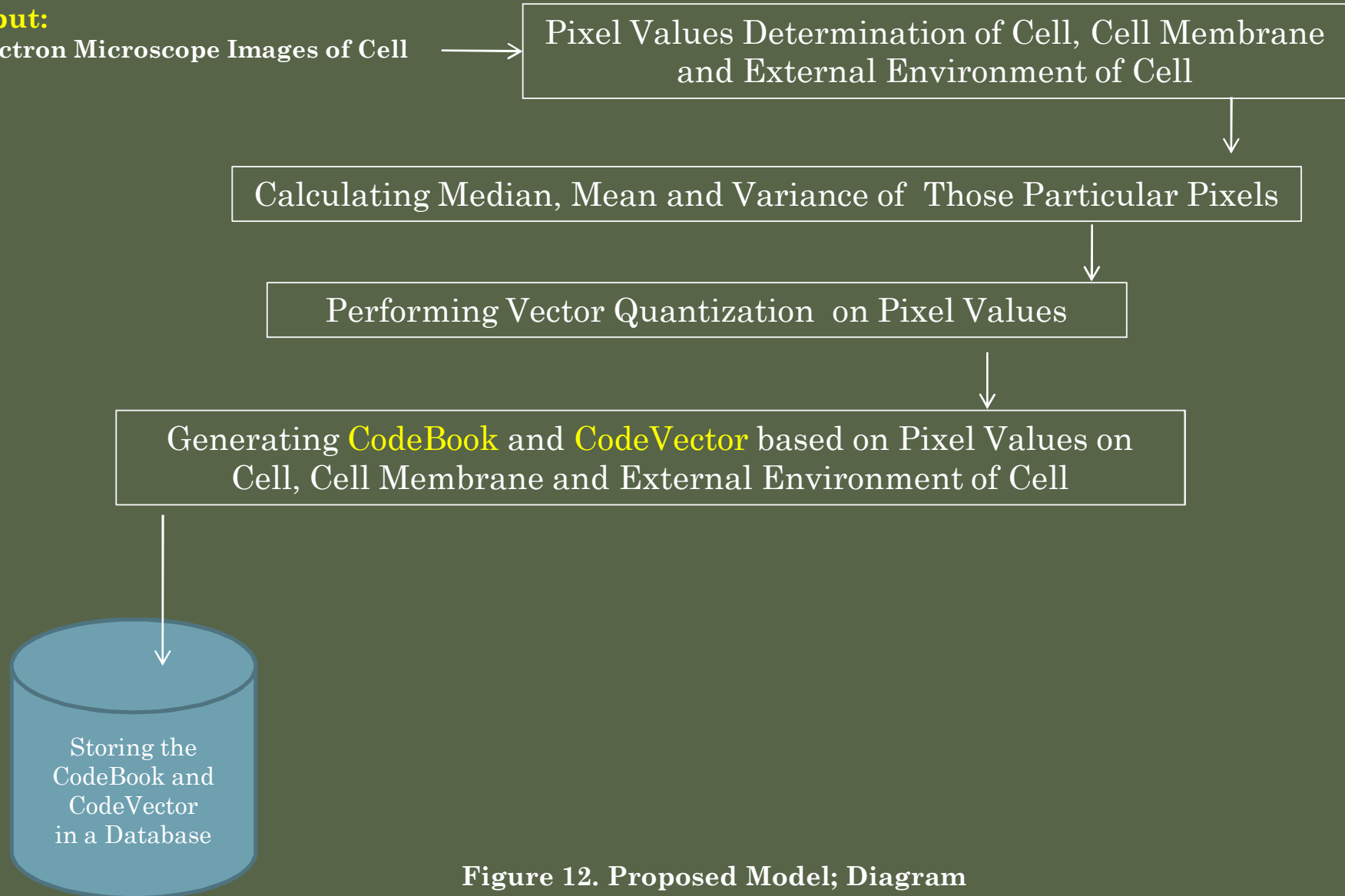


Figure 12. Proposed Model; Diagram

LBG VQ Algorithm

1. Determine the number of codewords, N , or the size of the codebook.
2. Select N codewords at random, and let that be the initial codebook. The initial codewords can be randomly chosen from the set of input vectors.
3. Using the Euclidean distance measure clusterize the vectors around each codeword. This is done by taking each input vector and finding the Euclidean distance between it and each codeword. The input vector belongs to the cluster of the codeword that yields the minimum distance.

LBG VQ Algorithm

4. Compute the new set of codewords. This is done by obtaining the average of each cluster. Add the component of each vector and divide by the number of vectors in the cluster.

$$y_i = \frac{1}{m} \sum_{j=1}^m x_{ij}$$

where i is the component of each vector (x , y , z , ... directions), m is the number of vectors in the cluster.

5. Repeat steps 2 and 3 until the either the codewords don't change or the change in the codewords is small.

LBG VQ Code

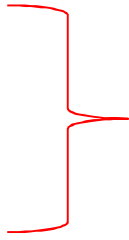
```
function [CODE_n, label, dist] = vq(X, CODE, n_it, QUIET);

%vq      Vector quantization using the LBG algorithm.
%      Use: [CODE_n,label,dist] = vq(X,CODE,n_it)
%      Performs n_it iterations of the algorithm on X, using
%      CODE as initial codebook.

error(nargchk(3, 4, nargin));
if (nargin < 4)
    QUIET = 0;
end;

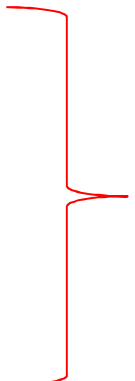
% Dimensions of X
[n,p] = size(X);
% Codebook size
m = length(CODE(:,1));
% Initialize label array
label = zeros(1,n);
% As well as distortion values
dist = zeros(1,n_it);

% Main loop
CODE_n = CODE;
for iter = 1:n_it
    % 1. Find nearest neighbor for the squared distortion
    DIST = zeros(m,n);
    if (p > 1)
        for i = 1:m
            DIST(i,:) = sum((X - ones(n,1)*CODE_n(i,:)).^2);
        end
    end
end
```



LBG VQ Code (Cont.)

```
end;
else
    % Beware of sum when p = 1 (!)
    DIST = (ones(m,1)*X' - CODE_n*ones(1,n)).^2;
end;
[vm,label] = min(DIST);
% Mean distortion
dist(iter) = mean(vm);
% 2. Update the codebook
n_out = 0;
for i = 1:m
    ind = (1:n);
    ind = ind((label == i));
    if (length(ind) == 0)
        % Isolated centroid are not modified
        n_out = n_out + 1;
    elseif (length(ind) == 1)
        % When there is only one nearest neighbor for a given codebook entry
        CODE_n(i,:) = X(ind,:);
    else
        CODE_n(i,:) = mean(X(ind,:));
    end;
end;
end;
% Affichage
if (~QUIET)
    fprintf(1,'Iteration %d:\t%.3f\n',iter,dist(iter));
end;
if (n_out > 0)
    fprintf(1,' Warning : %.0f isolated centroids\n',n_out);
end;
end;
```



Experimental Results

In order to evaluate the quality, accuracy and time complexity of the proposed method, we perform several tests on a sample dataset. Our sample datasets contains 196 normal electron microscope images from Imamreza laboratory and also our research laboratory in Mashhad, Iran. All codes are implemented by C++ (C++ Builder XE2). Figure 13 shows the original image and figure 14 (1 to 8) shows results for eight codevectors with using LBG VQ algorithm.

Experimental Results

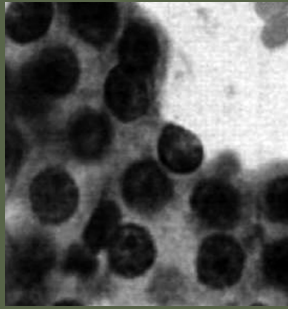
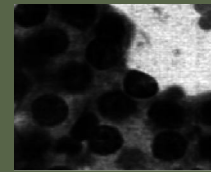
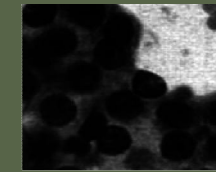


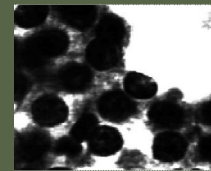
Figure 13: The original image



1



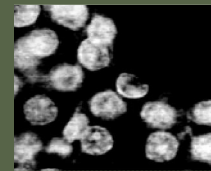
2



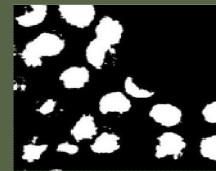
3



4



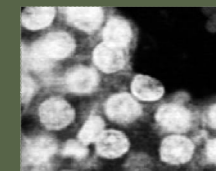
5



6



7



8

Figure 14: (1) Image with first codevector, (2) Image with second codevector, (3) Image with third codevector, ..., (8) Image with eighth codevector.

Experimental Results (Time of Execution)

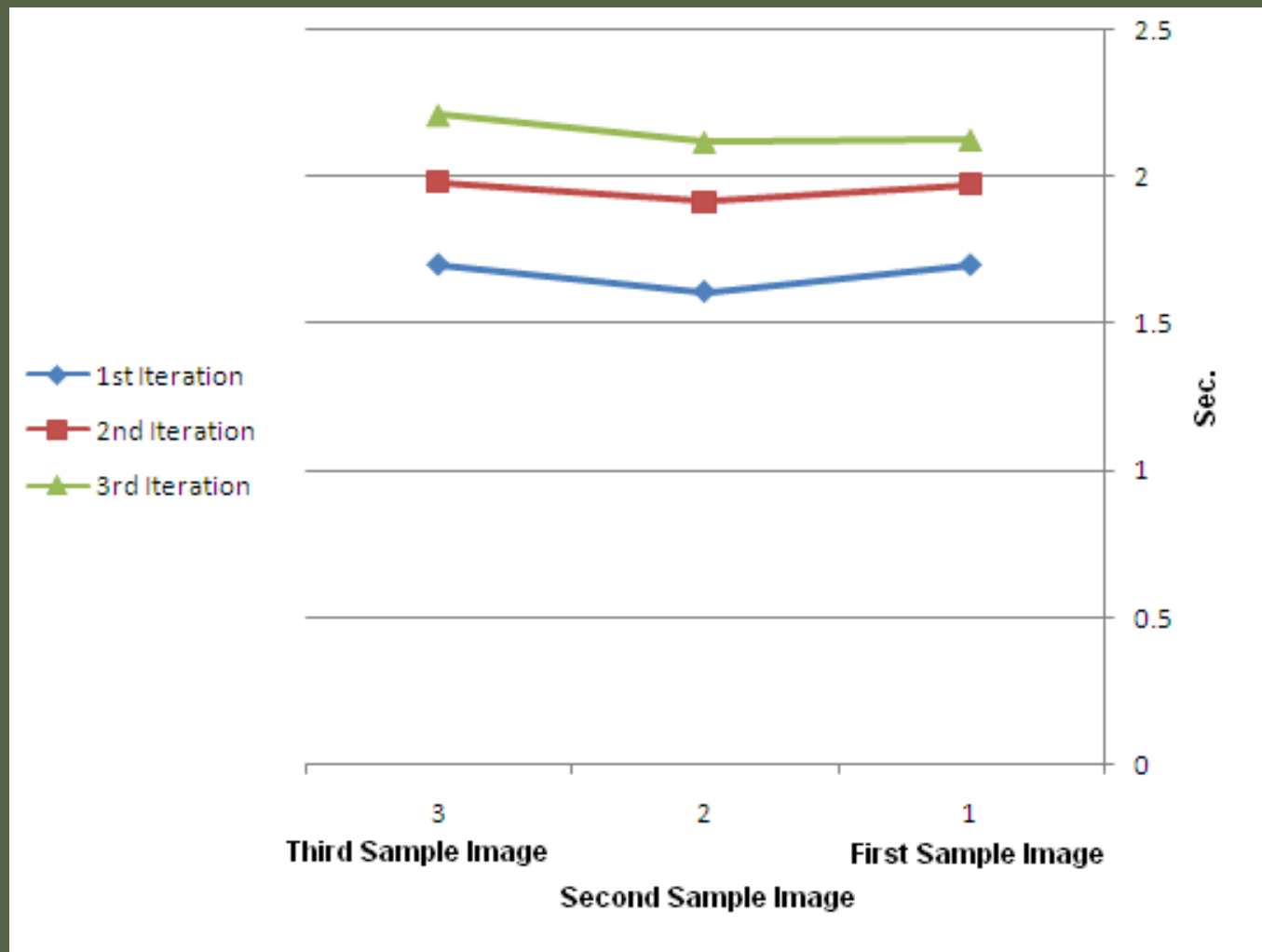


Figure 15: The line chart of time complexity for our proposed algorithm which performed on 3 sample images. We tested our code on a computer with CPU Pentium IV, 3.20 GHZ.

Comparison with other active works

Method	Developed By	Rate of Accuracy
Edge Detection and Elliptical Equation	F.Ebrahimnejad, N. Razmjoo	77.6%
Modified Cross Point Number and Neural Network Technique	A. M. Aibinu	85.36%
Deformable Templates	N. PeHrez de la Blanca	81.44%
Our Proposed Model	Mohammad V. Malakooti Hamidreza Naji Ahmad Pahlavan Tafti	86.69%

Conclusion

- In this work, we have used vector quantization as a clustering algorithm for human cell detection in microscopic images.
- Our proposed algorithm has been applied successfully for human cell detection in this type of images.
- The proposed algorithm has been shown on a number of experiments to address the accuracy and efficiently. Experimental results obtained from our algorithm shown that this model is both accurate and fast with around 86.69 percent of detection and 2 seconds for the third iteration.
- Future works will be adding the other properties such as texture features like saturation and intensity to improve our cell detection algorithm.
- Our algorithm may not only be used for microscopic images, but also can be applied on the other types of digital images.

References

- [1] Z. Yu, L. Wang and C. Pan. Medical image segmentation based on novel local order energy, Proceedings of the 10th Asian Conference on Computer Vision, pages 148-159, Queenstown, New Zealand, 2010.
- [2] R. Richard and L. Shapiro. Image segmentation techniques, Computer Vision, Graph and Image Process 29, 100-132, 1985.
- [3] P. K. Sahoo, A. K. Saltani and Y.C. Chen. A Survey of thresholding techniques, Computer Vision, Graph and Image Process 41, 233-260, 1988.
- [4] A. Garrido and N. P. La Blanca. Applying deformable templates for cell image segmentation, Elsevier , Pattern Recognition 33, 821-832, 2000.
- [5] B. Auffarth. Clustering by a genetic algorithm with biased mutation operator, WCCI, IEEE, 2010.
- [6] R. Huth. Classification of atmospheric circulation patterns, Ann. N. Y. Acad. Sci., 105-152, 2008.
- [7] E. Achtert, C. Bohm and P. Kroger. Mining hierarchies of correlation clusters, 18th SSDBM, 119-128, 2006.
- [8] R. M. Gray. Vector quantization, IEEE ASSP, 4-29, 1984.
- [9] Y. Linde, A. Buzo and R. M. Gray. An algorithm for vector quantizer design, IEEE Transactions on Communications, 702-710, 1980.
- [10] H. B. Kekre, K. Sarode and B. Raul. Color image segmentation using Kekre's algorithm for vector quantization. IJCS, Vol. 3, 287-292, 2008.
- [11] C. C. Chang and W. C. Wu. Fast planner oriented ripple search algorithm for hyperspace VQ codebook, IEEE Transactions on Image Processing, 1538-1547, 2007.
- [12] H. B. Kekre, K. Sarode and S. D. Thepade, Image retrieval using color texture features from DCT on VQ codevectors obtained by Kekre's fast codebook generation, GVIP, Vol. 9, 1-8, 2009.
- [13] H. B. Kekre and K. Sarode. Speech data compression using vector quantization, WASET IJCISE, Vol. 2, 251-254, 2008.
- [14] R. Grimaldi. Discrete and Combinatorial Mathematics: An Applied Introduction, Prentice Hall, 2002.

**Thank you very much
for your time and kind
consideration**