A Study of Retinal Blood Vessel Clustering for Finding the Main Vessel Convergence

1Pongsate Tangseng  2Nittaya Muangnak
3Pakinee Aimmanee  4Stanislav S. Makhanov  5Bunyarit Uyyanonvara
Sirindhorn International Institute of Technology
Thammasat University, Thailand

Email: 1pongstate1@gmail.com  2csnym@ku.ac.th
3pakinee@siit.tu.ac.th  4stanislav@siit.tu.ac.th  5bunyarit@siit.tu.ac.th

Abstract
This paper proposes a new algorithm for clustering the main blood vessels in retinal images. The new algorithm involves detecting main vessels and clustering. With an assumption that the location of optic disc (OD) is the centroid of the main branches of vessels, we evaluate our algorithm by using accuracy of the centroid of vessels clusters obtained from the experiment being inside of the area of the optic disc. Our new algorithm can obtain the accuracy as high as 92.59%.

Keywords: Data clustering, retinal blood vessel clustering, main vessel convergence

1 Introduction

Data clustering is a process of grouping similar items based on a similarity of patterns and is often one of the first step in data mining analysis [1]. Typical clustering process involves the following steps: pattern representation, similarity measurement, clustering or grouping, data abstraction, and assessment of output.

Many clustering algorithms such as Fuzzy c-means [2], Gaussian(EM) [3], Quality threshold [4] and K-mean have been proposed. However, each algorithm has its own merits and demerits and cannot work for all real situations. For example, a well known clustering technique, K-mean clustering [5], is used to classify the series of data items into \( k \) clusters for which the within cluster sum of squares is minimum. An example of clusters resulting from K-mean where \( k \) equals to 2, 4, 6 and 7 are shown in Figure 1.

However, a major drawbacks of the \( k \) means is that the number of clusters must be predefined. Since in this work, we would like to reconstruct main branches of vessels from the list of broken vessels which are obtained from vessel segmentation of the original retina image and the number of clusters (main branches of vessels) are normally unknown, K-mean is not a good choice for our work.

Grouping vessels that are relatively close and have similar properties such as intensity and thickness are key concepts in our clustering algorithm. The purposed method aims to increase the correctness of vessel clustering, which is used to find the location of OD. In the next section, the proposed method is explained in more detail. Section 3 describes the collection of data that is used in the experiment, setting of parameters, and evaluation scheme. Section 4 shows results and discussion and finally the last section summarizes this work.

2 Proposed Method

Our proposed vessel clustering algorithm is described in Algorithm 1. Figure 3 - Figure 7 show the results obtaining from each step in the algorithm. In Figure 2a, retinal blood vessels are extracted using the efficient vessel detection approach based upon using wavelets and edge location refinement [6]. The useful vessel data such as vessels center line, diameter, and length are also obtained from the process of vessel extraction. Figure 2b shows the results after the vessel extraction and segmentation. Since we only care vessels that are relatively close, we also ignore vessels that are too far away from the rest. Figure 3 shows an example of results after removing outbound vessels. In this figure, there are several tiny lines which come from noise in the image. These tiny vessels are then removed from the vessels. Figure 4 shows the example of results after tiny vessels are removed. Then we group vessels that are close. In this step, there are still many clusters. Figure 5 and Figure 6 show the results after we cluster the vessels and
Figure 1: Examples of k=2,4,6 and 7 clusters obtained from k-mean clustering of blood vessels at different k’s

select only the first top five vessels that are the thickest. Figure 7 we re-merge vessels to get the final cluster.

3 Experiment and Evaluation Results

We use the collection of images collected from the STructured Analysis of the Retina (STARE)

Algorithm 1 Algorithm for finding the clusters of vessels

Require: 1. Vessel data including vessel number, x,y position, thickness and intensity of green channel (G) of each point from vessel extraction. 2. Original image Let \( S \) ← vessel data; where \( S_i \leftarrow (\text{vessel number, x, y, thickness, G}) \) \( \text{img} \leftarrow \text{green channel of original image} \)

Step 1: Define high contrast vessels using Difference in Green (DG)

for all \( S_i \in S \) do
    \( \text{window} \leftarrow \text{img}\{x-\text{thickness, x+thickness}\}\{y-\text{thickness, y+thickness}\} \)
    \( \text{thresh} \leftarrow \text{GRAYTHRESH(window)} \)
    \( \text{meanBG} \leftarrow \text{MEAN(window,thresh)} \)
    \( \text{DG} \leftarrow \text{G of } S_i - \text{meanBG} \)
end for

Step 2: Remove low contrast vessels

for all \( \text{numberOfVessels} \) do
    \( \text{vesselDG} \leftarrow \text{mean of DG in a vessel} \)
end for
Find mean and standard deviation of vesselDG
Remove vessels that vesselDG < \( \text{param1} \)

Step 3: Remove short vessels

for all \( \text{numberOfVessels} \) do
    \( V \leftarrow S \); where vessel number = i
    Calculate distance between every 2 points in the V
    Create a weighted complete graph G where each v in V are nodes and weight of edge between each 2 v are distance between them.
    Find minimum spanning tree MST of G
    vessels length is summation of weight in MST
end for
Find mean and standard of vessels length
Remove vessels that vessels length < \( \text{param2} \)

Step 4: Group adjacency vessels
Find distance between vessels
Group vessels that the distance between them is less than \( \text{param3} \). The group of vessels called cluster.

Step 5: Find length and thickness of clusters

for all \( \text{clusters} \) do
    Create a weighted complete graph G by consider vessels in cluster as a node and distance between vessel is a weight of edge between nodes
    Find minimum spanning tree MST of G
    clusters length is summation of weight in MST
    clusters thickness is summation of thickness of every point in the cluster
end for

Step 6: Select 5 thickest clusters & group adjacency clusters
Find distance between clusters
Group clusters that the distance between them is less than \( \text{param4} \). The group of clusters is final answer.
\return finalClusters
Figure 2: Segmented retinal blood vessel result prior to step 1

Figure 3: After delete outbound Vessels of DG in step 1

Figure 4: After delete short Vessels in step 3

Figure 5: All possible Clusters from step 4

Figure 6: Top five thick Clusters in step 6

700x605. These images can be classified into two groups: healthy and unhealthy. The numbers of images in the healthy and unhealthy groups are 50 and 31, respectively. For the healthy group, the images are taken from the normal retina. The optic disc boundary and vessels in this set are clearly detected. On the other hand, the optic disc boundary and vessels in the unhealthy group are faint. The boundary of OD and blood vessels are unclear and having close intensity with the background.

First, an image resource was collected from the STructured Analysis of the Retina (STARE) research project [7]. The eighty-one color images are 700x605 pixels in size, and were divided into two groups which are 31 images of healthy retinas and 50 images of diseased retinas. In the images of a healthy group, a clearly identified
optic nerve has been defined by property contribution of the nerve such as shape, intensity, size, and convergence; meanwhile, the diseased retinal images have been shown anomaly in retina particles that caused the absent of the nerve. The latter makes the retinal vessel detection difficult. Examples of images which are healthy and unhealthy retina are shown in Figure 8a and Figure 8b.

In our experiment, we set the threshold of the difference in green (DG) value be the difference between mean and standard deviation value of DG value, the distance between two vessels to be considered to be in the same cluster be 10, and the length of the vessels to be ignored be the difference between mean and normalized standard deviation value of vessel length where the normalized standard deviation is defined by the standard deviation divided by a constant value.

The selected results obtained from using the proposed methods are shown below in Figure 9 - Figure 12.

Using the assumption of that the point that is closest to all cluster is the center of the OD [8], we evaluate our vessel clustering [9] by finding the accuracy of getting the location of OD. The accuracy is calculated by finding the ratio of the total number of images whose centroid obtained from the clustering algorithm is inside of the OD to the total number of images. We find that the accuracy of healthy group and unhealthy group are 96.77% and 90.00% , respectively.

4 Conclusion

In this paper we proposed a new clustering algorithm which is designed specifically to group the main blood vessels into clusters. These clusters represent main branches of blood vessels. The proposed method uses length, intensity, and thickness of the vessels to detect main vessels
and uses distances between vessels to cluster two vessels or clusters that are close together. With the assumption that the centroid of the cluster is the location of OD, we find that the accuracy of the algorithm in getting the location of the OD is 92.59%.

Acknowledgment
We acknowledge STructured Analysis of the Retina (STARE) research project for useful standard image database. Thanks to National Research University Project of Thailand Office of Higher Education Commission and Thai Royal Government Scholarship for funding this project.

References


