

SUPERPIXEL BASED K-MEANS CLUSTERING ON CHEMICAL BONDING STRUCTURE ANALYSIS

Asha Jenifer and N. R. Raajan*

School of Electrical and Electronics Engineering, SASTRA University, Thanjavur-613401, India

*E-mail: nrraajan@ece.sastra.edu

ABSTRACT

Identification and segmentation of chemical bond images is important in computer aided diagnosis. In this paper, two chemical bonds were secured for superpixel based linear clustering, is proposed. In the clustering, histograms from contrast enhanced image channels and centre surround statistics from centre surround difference bonds are proposed as features to determine each superpixel as information or bonded structures. In the initial step, to handle the unbalanced cluster issue due to the presence of unbalanced images, bootstrapping is done. The proposed method has been tested on two bonds namely propanol and methoxyethane images with its boundaries marked by image processing techniques. The results also show a decrease in overlapping error as the reliability score is reduced. The method can be used in computer aided diagnosis systems to enhance the bonding structure for clinical deployment of the automatic segmentation.

Key words: superpixels, K-means clustering, segmentation.

©2015 RASĀYAN. All rights reserved

INTRODUCTION

In many clinical and industrial applications there is a need for analysing chemical components to determine the elements present in it. As a basic step to the analysis, segmentation is necessary to determine the details of each and every pixel. In this paper a superpixel based linear clustering which uses K-means clustering for segmenting the chemical bonding structure is proposed. Superpixels are used to provide a meaningful representation of an image. They may be used as primitive units for successive image processing tasks like object detection or image segmentation⁴.

Superpixels usually cover the entire image and they are distributed according to the characteristics of the image. The pixels of same colour and texture are grouped together in the pixel grid. The purpose is to decrease the redundancy in the image⁴. The over segmentation results are said to be superpixels. They are highly useful in reducing the further image processing levels. It can be generated by many algorithms each one having its own merits and demerits which can be used for a particular application. The size is usually expected to be small, with little variations. The boundary of superpixels must be appropriate with the boundary of the data in the input image.

Superpixel based linear clustering

Clustering is a basic tool used for analysing data which produces corresponding groupings based on the given input data. It is mostly used in image segmentation, data mining and medical images³. The difficult task in this method is determining the number of clusters present for a given input data. The next task is to decide the grouping of input data points according to the clusters. The number of clusters is chosen based on the requirement. When the number of clusters is known, the grouping of the input data points can be done easily. There are two properties for the clusters obtained with the help of this algorithm: the points belonging to the same cluster should be same and the points belonging to different clusters should be different³.

In this algorithm the pixels are clustered based on the colour similarity to generate superpixels. This algorithm basically uses K-means clustering in an iterative manner to obtain the superpixels¹. The boundary of superpixels should match with the natural boundary of the image. The desired number of superpixels is given as input to this algorithm. It enables the compactness with the help of a separate

distance measurement and provides uniform superpixel shapes. This algorithm can be used for both grayscale and colour images. The clustering of pixels is done based on colour similarity over the image plane. As the algorithm basically uses K-means algorithm, K cluster centres are selected at the initial stage of the algorithm and colour space is used for clustering the pixels. The complexities in the distance measurements are reduced in this algorithm and the only parameter used in this algorithm is N i.e the number of superpixels desired. Superpixel based linear clustering algorithm is simple to use.

EXPERIMENTAL

Algorithm

During the training process, bootstrapping is done to handle the unbalanced cluster size present in the image and then the segmentation is carried out based on the superpixel classification. As the K-means algorithm is used as the basic concept, there are four main steps in this algorithm:

- The input to the algorithm is K, it is the number of cluster centres with the help of which superpixels are formed.
- The Euclidean distance is calculated between each pixel and the cluster centre, the pixel is then assigned to the cluster based on the minimum Euclidean distance obtained.
- By averaging the pixels in the superpixel, the superpixel centres are re-computed.
- The second and third steps are repeated until the algorithm converges.

Assuming the image consisting of M pixels, the superpixel has M/K approximate size. There will be a superpixel centre at an interval $C=\sqrt{(M/K)}$. After each iteration, the superpixel centres will be updated due to the re-computation of the superpixel centres².

In this algorithm, a condition for stopping the algorithm is proposed i.e when there will be no more pixel change 'u' and when the number of iterations reaches 'v'. The algorithm stops when either of the conditions is met.

The size of all the superpixels obtained is generally equal in size. Also, in the process of obtaining superpixels, there is no loss in spatial and colour information. The colour and intensity information of the connected pixels in each superpixel are uniform in nature.

Propanol

The chemical structure used in this paper for superpixel based linear clustering is propanol. 1-Propanol is commonly called as propanol, is an alcohol with the structural formula $\text{CH}_3\text{CH}_2\text{CH}_2\text{OH}$ and molecular formula $\text{C}_3\text{H}_8\text{O}$. It is a colourless liquid having a smell like alcohol. It can be mixed with water and it is highly soluble in alcohol.

It is a stable and highly flammable chemical. Propanol is also used in manufacturing pharmaceuticals, perfumes, cosmetics and other chemical products. It can cause extreme irritation when comes in contact with eyes. In vapour form it causes irritation in nose and throat.

Methoxyethane

The other name of methoxyethane is ethyl methyl ether which belongs to ethyl group with the structural formula $\text{CH}_3\text{CH}_2\text{OCH}_3$ and molecular formula $\text{C}_3\text{H}_8\text{O}$ ⁵. It is present in the form of colourless gas, water soluble and it is highly flammable. It can cause dizziness on inhaling. It can form salts on mixing with acids and when mixed with air it forms an explosive.

RESULTS AND DISCUSSION

Propanol

Figure- 1a represents propanol which is given as an input. When the number of superpixels is given as 500, the output is obtained with some overlapping errors as shown in figure-1b. As the number of superpixels is reduced to 200, the output is obtained with less overlapping errors, given in Fig.-1c.

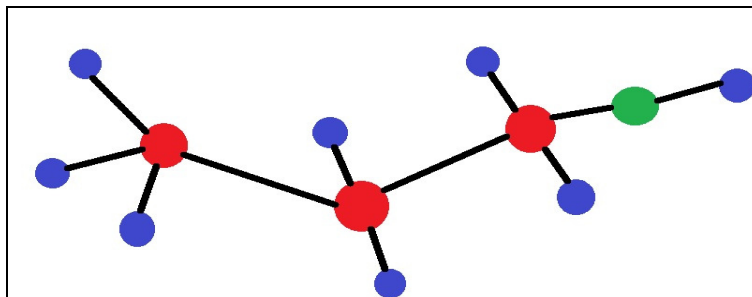


Fig.-1a: Input image

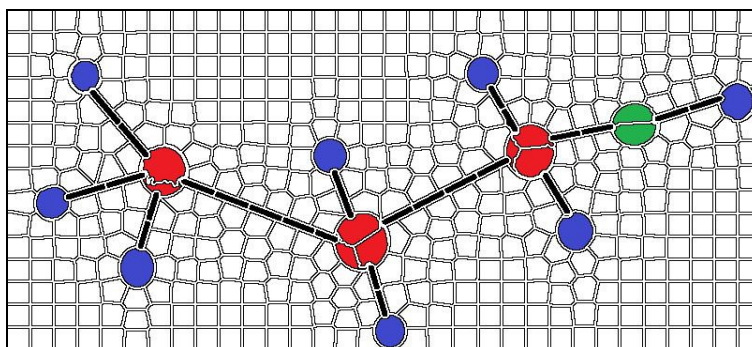


Fig.-1b: Number of superpixels: 500

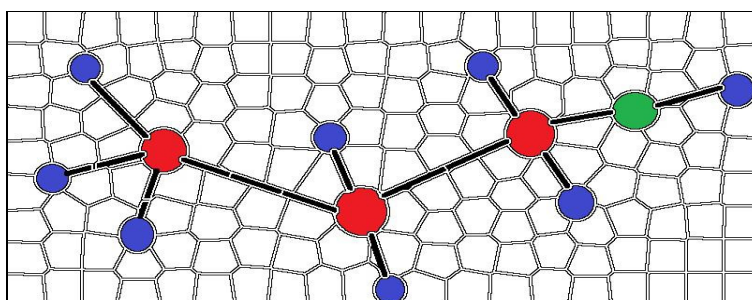


Fig.-1c: Number of superpixels: 200

Methoxyethane

At the same time for Methoxyethane shown in Figure-2a, an input image. When the number of superpixels is given as 200, the output is obtained with some overlapping errors (Fig.- 2b) and when the number of superpixels is reduced to 100, the output is obtained with less overlapping errors (Fig.-2c).

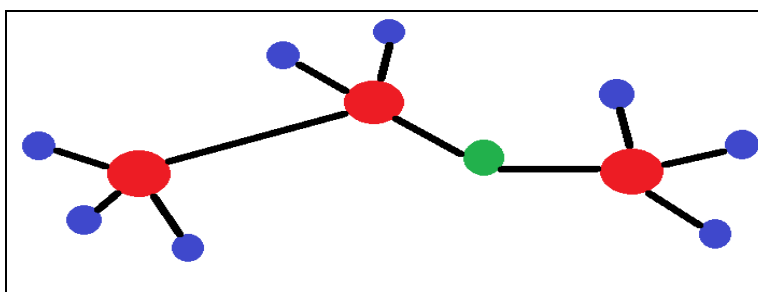


Fig.-2a: Input image

Thus, the overlapping error is decreased when the reliability score is reduced and helps in identifying the chemical components and the bonding efficiently.

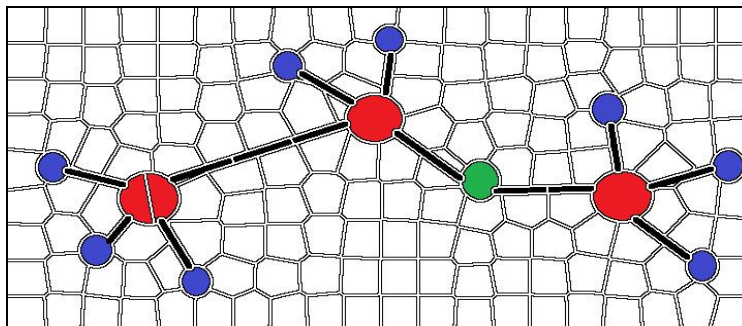


Fig.-2b: Number of superpixels: 200

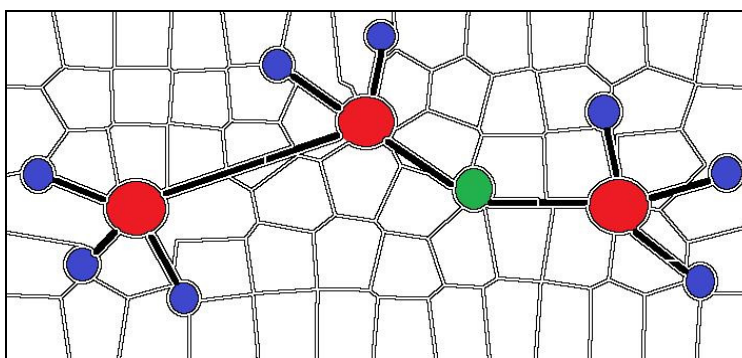


Fig.-2c: Number of superpixels: 100

CONCLUSION

Superpixels have become a crucial tool for the computer vision applications. In this paper, we proposed a new method for generating superpixels based on k-means clustering for analysing chemical bond structures, which has been shown to outperform in terms of performance, segmentation speed and boundary adherence. Although our experiments are effective, they come with a note of caution. There are certain limitations which have to be resolved from superpixels, however this may adversely affect their performance.

REFERENCE

1. Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, *Journal of Latex Class Files*, **6(1)**, 1 (2011).
2. Pritesh Vora, Bhavesh Oza, *International Journal of Science and Modern Engineering*, **1(3)**, 24(2013).
3. Trung Huy Duong, *New Techniques for Data Clustering and Color Image Segmentation*, UMI 3630062 Published by ProQuest LLC (2014).
4. Wanda Benesova, Michal Kottman, *In Proceedings of the International Conference of Machine Vision and Machine Learning Prague, Czech Republic*, Paper No. 67, (2014).
5. S. Syed Shafi and S. Senthilkumar, *Rasayan J. Chem.*, **7(4)**, 370(2014).

[RJC-1229/2015]