FINDING IMAGE STRUCTURE BY HIERARCHICAL SEGMENTATION

Bo QIU1, Daniel RACOEANU2

1Institute for Infocomm Research (I²R), Singapore, 119613
2Image Perception, Access & Language, UMI CNRS 2955, Singapore
{qiubo, visdaniel}@i2r.a-star.edu.sg

ABSTRACT

Image segmentation has been studied for many years. But what factors influence segmentation results indeed? Why some images are easy to be handled while the others are not? In this paper we put forward the so-called ‘image structure constant’ and ‘image structure map’ to judge the complexity of an image. They can be applied on any image. ‘Structure constant’ can be found by a hierarchal segmentation method based on k-means and gray histogram, which is processed by increasing the clustering centers’ number of k-means step by step and tracing the regions’ change. At the same time its structure map can be formed reflecting the relationship between pixel gray values and image regions. With the structure constant and structure map we can dissect an image is easy to be segmented or not, quantitatively. Furthermore, a Neighbor-Matched-Region (NMR) graph is designed to judge an image’s complexity. Experiments show that the proposed concepts and the relevant algorithms are useful tools in analyzing images.

1. INTRODUCTION

Image segmentation keeps on being an important field for many years. A comprehensive review of segmentation methods can be referred to [1]. Generally speaking, a segmented image is a piecewise ‘smoothed’ or piecewise constant image. Its objects or regions are non-overlapping, homogeneous, and connected [2]. Within the regions, intensity is smooth while at the borders it is abrupt. Most of the segmentation algorithms can be related to Mumford-Shah segmentation model [1], in which a cost functional or energy functional is defined to evaluate how well a proposed segmentation $J[R]$ explains the structure of a given image $I$ [3]. The ‘correct’ segmentation will reach the minimum of $E$ globally. A definition of $E$ is:

$$E(I,J,[R_i]) = \int (I - J)^2 + \sum_i \| \nabla J \| + \text{length}(\bigcup \partial R_i).$$  (1)

In this formula, $J$ is called the cartoon of the full image $I$, and is an idealization of $I$, in which clutter and noise have been stripped away. The details can be referred to [1] and [3].

However, the formula (1) is difficult to be computed in reality owing to the variety and complexity of images. Many different kinds of segmentation methods having been developed, generally summarized as four groups: edge-based, neighborhood-based, histogram-based and cluster-based [4].

Among all of the methods, a difficult problem still remains how to effectively compare them. This is owing to the general lack of standardization of performance measures. In addition, there is also no standard database for comparison of segmentation methods [5].

Instead of comparing the segmentation methods, in this study we propose another view angle, which is to compare images. In fact, the characteristics of the image determine the really effective segmentation method, since different methods are chosen and adjusted according to different images.

In this paper we develop an algorithm to judge the complexity of images, which can help us to know more about the image nature. Through the hierarchal segmentation by $k$-means clustering, the factors which influence the segmentation results are explored and the structures of images are formed. Based on the factors and structures, we can conclude that some images are initially easy to be segmented while others are difficult. Besides, we also can have a reasonable segmentation result based on the hierarchal segmentation process. The system flowchart is shown in Figure 1.

As a beginning of algorithm development, the system only considers gray images. All color images are thus converted into gray level ones before processing.

![Figure 1 Flowchart of our system](image-url)

2. HIERARCHICAL SEGMENTATION BY K-MEANS

Our method combines the histogram-based and cluster-based segmentation: firstly, based on the histogram of an image, $k$-means is applied and image segmentation clusters are generated; secondly, the $k$ is increased step by step (the word ‘layer’ is used to present different $k$), and the clusters are compared between layers. By tracing and analyzing the changes of clusters from $k$-means, the method can find a tree structure for each image. This structure describes the complexity of an image.

2.1. K-means

Generally, $k$-means algorithms assigns each point to the cluster whose center (also called centroid) is nearest [6][7][8]. Its process is roughly as follows:

- Choose the number of clusters, $k$;
Generate $k$ seeds;
Assign each point to the nearest cluster center and re-compute the new cluster centers;
Repeat until some convergence criterion is met.

For the input gray level image, its histogram is computed at first. Then the histogram vector ($256 \times 1$) is regarded as $Y$, and the corresponding gray values (from 0 to 255) are taken as vector $X$. Thus, the data points set $(X, Y)$ is constructed.

In the following, $k$-means is applied on the $(X, Y)$. For the distance calculation, when using city block distance, it should be

$$\text{dist}(i, j) = |x_i - x_j| + |y_i - y_j|.$$  \hspace{1cm} (2)

In this paper, we use

$$\text{dist}(i, j) = |x_i - x_j|.$$  \hspace{1cm} (3)

because $X$ and $Y$ have totally different meaning, and only the gray values ($X$) are meaningful to the clustering. However, there is no change in the re-computation of centroids:

$$x_{\text{centroid}} = \frac{\sum x_i}{n}, \quad y_{\text{centroid}} = \frac{\sum y_i}{n}.$$  \hspace{1cm} (4)

In fact, except the calculation of distance, all the other parts are the same as in the traditional $k$-means. Our experiments show that when traditional distance is used, even when $Y$ is normalized, the results are very poor.

2.2. Hierarchical segmentation & image structure

Based on the $k$-means algorithm described in Section 2.1, the hierarchical segmentation can be applied on an image.

As shown in Figure 2, for the initial image in (a), when $k$ is increased step by step from 2 to 6, at each step there will be a new region/object segmented from an old region/object. As we can see, the new regions are purely coming from one region in a higher layer (corresponding to a smaller $k$). For example, in (c) there are 3 regions: the background, the checkmark, and the other two signs as a whole region; in (d) there are 4 regions: the background and the checkmark regions keep the same, and the two-sign region splits into two. In (e) and (f), each time the new region is born from the background region.

![Figure 2](image)

Figure 2: An example of hierarchical segmentation (one color corresponds to the label of one region)

The upwards process can form the image structure as shown in Figure 3(a). For $k=2$, the image is segmented into $R_{2,1}$ (background) and $R_{2,2}$ (two signs as foreground), corresponding to Figure 2(b). For $k=3$, the new region checkmark ($R_{3,2}$) is generated from the old region $R_{2,2}$, while the $R_{3,1}$ (foreground) keeps the same but its name was changed to $R_{3,3}$. For $k=4, 5, 6$, the similar description can be given while separately corresponding to Figure 2(d), (e), and (f).

Of course bigger $k$ can be assigned and the process can go on. Nevertheless, for the image in Figure 2(a), the further segmentation is meaningless and the process should stop at $k=6$. In fact, this corresponds to a criterion for automatic segmentation: when there are no regions changing any more, the segmentation process stops.

Obviously, the image represented in Figure 2(a) is a simple image: all its objects are independent and inside each object, the pixel values keep the same. So it results a very clear image structure as shown in Figure 3(a). However, for many other images, the objects are not so easy to be segmented and their image structures are disordered. In Figure 3(b) there is an example: it is obvious that until $k=5$ there is nearly no order, and most of the regions in different layers have no relations between each other, but are generated from the initial image directly, except the $R_{5,5}$ which is generated from $R_{4,2}$. When $k=6$, there are some hierarchical relationships displayed, e.g. the $R_{6,1}$ is generated from $R_{5,1}$, $R_{6,2}$ is generated from $R_{4,1}$, and $R_{6,3}$ is generated from $R_{4,2}$ etc. However it is not enough. Further segmentation is needed until the regions become totally stable: each region is just coming from its neighbor higher layer and no new regions are generated again.

![Figure 3](image)

Figure 3: Structure image of hierarchical segmentation ($R$: region)

2.3. Structure constant and structure map

For some images, even when $k$ increases, the image structure keeps on being disordered. Obviously, the biggest possible $k$ is 256, the gray values’ superior limit. For most images, when $k$ is big enough, the image structure will become stable. Considering all of the situations, we can say that the biggest $k$ of an image can be taken as the image’s structure constant. The structure constant reflects a kind of structural nature of images: the relationship between pixel
gray values and objects’ independence, since the segmentation process is always tracing the changes of objects/regions. If most of an image’s objects/regions are composed of pixels with a single gray value, this image’s structure constant will probably be small, and the objects are prone to be independent and easy to be segmented.

However, even having the same structure constant, different images may have different structure maps. In Figure 3, if (b) turns stable after $k=6$, it will have the same structure constant $6$ as (a). Obviously, the structure maps are totally different: (a) is very clean and the hierarchical relationship is very clear; but (b) is very disordered and the hierarchical relationship is rather vague. On this meaning, image structure map reflects the complexity of an image. If an image has a clear structure map, it is easy to be segmented; inversely, if an image has a disordered structure, it will be difficult to be processed by any image processing technique. Thus, image structure map can be a figure to judge an image’s complexity.

In a summary, image structure constant and structure map can be used to evaluate an image’s complexity or nature, which influences the feasibility of being segmented or processed by image techniques. In another word, they can give a judgment of an image: it is ‘difficult’ or not. In the following some ‘difficult’ images are shown and their structures are analyzed.

2.4. NMR graph
As described in former sections, the ‘easy’ images are those with discrete/independent objects inside, where the pixels are nearly monochromic (see Figure 4(a)). The ‘difficult’ images are generally composed of too much texture, where an object contains different pixels with varied values (see Figure 5(a)).

We introduce the Neighbor-Matched-Region (NMR) graph, which can be seen in Figure 4(c), (d) and Figure 5(b). In NMR graph, the horizontal axe is $k$, the segmentation layer; and the vertical axe is matched region number, meaning that how many regions in current layer are purely generated from just one region in its neighboring higher layer. The diagonal 45° line shows an ideal situation: all the regions in a same layer are matched with regions in its neighboring higher layer.

For example, in Figure 4(c), the first point is (3, 2), where 3 means $k=3$, corresponding to the 2nd image in Figure 4(b), and 2 means in the current 3 regions, there are 2 regions matched the regions in its higher layer. As a reference, when comparing the 1st and 2nd images in Figure 4(b), it can be seen that the biggest and smallest regions of layer $k=3$ are purely coming from the 2 regions in the layer $k=2$. The middle region of layer $k=3$ (green part) is composed of pixels from both of the 2 regions in layer $k=2$. The point of Figure 4(c) is (4, 4), meaning that all of the 4 regions in current layer $k=4$ are purely generated from their corresponding regions in higher layer $k=3$.

The NMR graphs in Figure 4(d) correspond to the latter 3 images in Figure 4(a). The NMR graphs in Figure 5(b) correspond to the 3 example images in Figure 5(a). As it can be seen, when most of the points in NMR graphs fall onto the diagonal line, the images are ‘easy’ and their image structures are very clear; when most of the points are far away from the diagonal line, the images are rather ‘difficult’ and their structure is disordered.

3. EXPERIMENTS AND RESULT ANALYSIS
All the images used here have been extracted with Google image search engine.

3.1. Results
For the images in Figure 4(a), their structure constants are 6, 6, 5, and 11. The corresponding structure maps are easy to be given according to the segmentation processes. For the images in Figure 5(a), their structure constants and structure maps are difficult to be found. Higher $k$ must be tried.

For most of the natural images, small $k$ cannot bring the points falling on the diagonal line of the NMR graph. But as we can imagine, with the increment of $k$, sooner or later there will be some points falling on it, at least for the maximal point $k=255$. In this case the complexity can be judged by this way: fit a line through (0,0) and the points in NMR graph, then calculate the angle between this line and the diagonal line. This angle is used to measure the complexity. As for Figure 5(b), the most complex image is the left one and the complexity lowers down to the right because the angles are smaller and smaller.
3.2. Influence factors

As we know, many factors will influence the $k$-means segmentation results: choice of initial points, noise, image sizes, gray or color, and some thresholds.

In our method, the initial points are fixed along the gray value axe: they are equal-distance points from 0 to 255. So the results won’t be influenced by this factor.

Common noises include Gaussian, Poisson and Salt & Pepper. Our method is robust when facing Salt & Pepper, but turns a little unstable when facing Gaussian or Poisson noise (see Table 1).

Table 1 Influence of noises

<table>
<thead>
<tr>
<th>Initial</th>
<th>Gaussian</th>
<th>Poisson</th>
<th>Salt &amp; Pepper</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Initial Image" /></td>
<td><img src="image2" alt="Gaussian Image" /></td>
<td><img src="image3" alt="Poisson Image" /></td>
<td><img src="image4" alt="Salt &amp; Pepper Image" /></td>
</tr>
</tbody>
</table>

Image sizes have no influence on our method. Even when multiplied different scales with image width and height, there is no change for the NMR graph, the structure constant and map.

Color images do behave more complex than gray images. So after transferring a color image into a gray one, even though the initial color image is simple, the transferred gray image may show some complexity. From this viewpoint, color is an influence factor. However this can be improved later by introducing segmentation techniques in a color space.

Figure 6. Influence of threshold

A threshold is defined when comparing regions between different layers. One region A in the current layer is matched to a region B in a higher layer when all the pixels of A are coming from B. But this is an ideal situation. Considering the influence of noises and varied edge pixels (when region edges are not sharp), even between matched regions there are some different pixels. The threshold is defined as a superior limit that how many different pixels at most can be tolerated for two matched regions. Owing to different image sizes, this threshold can be related to a fixed percent of image areas. Influence of the threshold is shown in Figure 6.

In Figure 6(b) it is obvious that the bigger the threshold is, the closer the points are to the diagonal line, which means clearer structure. This is because bigger threshold omits the influence of small regions.

4. CONCLUSIONS AND FUTURE WORK

In this paper, image nature is explored by a hierarchal image segmentation based on $k$-means. Some derived figures are defined like image structure constant, structure map, and NMR graph, which are very useful in describing an image.

Image structure constant reflects where the maximal segmentation layer is. It also reflects whether an image is easy to be segmented or not. Thus, it is a good figure used to judge an image’s complexity.

Image structure map can describe an image’s segmentation structure. It reflects the relationship between pixel gray values and image regions. Clear structure map means image regions are composed of identical gray value pixels.

NMR graph reflects the relationship between two layers in the process of hierarchal segmentation. It is another form of image structure map. When most of its points are close to the diagonal line, it means that the image has a clear structure.

Nevertheless, image region continuity hasn’t been considered in the method by now. As we can see from Figure 2(b), there are 2 regions belonging to a same cluster (blue), which means that they are regarded as one region. This partly explains why the method is not robust when facing noises. Noise points change too much between two layers’ segmentation. Without the limitation of region continuity, it is impossible to erase the effects of random noise points. Future development will solve this problem and color space will also be considered. Besides of $k$-means, single-, complete-, and average-link hierarchal clustering methods may be tested and compared.

5. REFERENCES