Automatic Texture Segmentation Based on k-means Clustering and Co-occurrence Features

Lucas Bastos, Aura Conci
Universidade Federal Fluminense
Instituto de Computação
Rua Passo da Pátria, 156, Niterói, RJ, Brazil
lbastos@ic.uff.br, aconci@ic.uff.br

Abstract

This work presents a method for automatic texture segmentation based on k-means clustering technique and co-occurrence texture features. A set of features was extracted from 256 gray-level co-occurrence information. These features were used to segment image regions regarding the textural homogeneity of its areas. As the process of calculating co-occurrence information demands the majority of computational time required, we propose a new methodology based on a gray-level co-occurrence indexed list (GLCIL) for fast element access, highly optimizing this step in the algorithm. Besides that, we compare the efficiency of the proposed method against other well known algorithms. The experiments show that GLCIL is the most efficient method in terms of computational time. Additionally, traditional Brodatz textures and other literature examples were tested to evaluate the appropriateness and robustness of the method.

1. Introduction

Image segmentation has a large number of applications in several fields. Its importance concerns basically in distinguishing different objects or regions inside the image. The process of automating image segmentation is, however, very complicated. This is often unfeasible, not only because of technical limitations but because there is no indication of what is meaningful for an intended user. Natural scenes, for example, are rich in details and tonal variations, which are better detected when the algorithm considers additional features. Therefore, texture based methods tend to be more efficient than tone based methods.

One of the most used approaches to image segmentation is based on region detection. Region based methods consist in localizing areas by analyzing pixels statistics. Several authors have worked in finding descriptors and features for texture identification. Among these, Haralick features [2] are the most widely used. In his work, Haralick [2] suggested the use of gray-level co-occurrence matrices (GLCM) to extract texture features from an image. Since then, GLCMs became widely used for image texture features extraction in many types of applications. The major drawback of the use of GLCMs is that it is computationally very intensive. Besides that, because co-occurrence matrices are generally sparse, most of the calculations are done over unnecessary zero frequencies. To deal with this problem, Shokr [3] suggested the use of a Grey-level Co-occurrence Linked List (GLCLL) to store only the non-zero frequencies. Other techniques were suggested as a faster alternative to co-occurrence matrices calculation [1].

In this paper, we propose a method for image segmentation based on co-occurrence features and k-means clustering. For the co-occurrence calculation phase, we introduce a novel algorithm based on a Grey Level Co-occurrence Indexed List (GLCIL) structure for fast element access and total computational time reduction. Some of the most common Haralick features are extracted. Finally, a k-means clustering algorithm is used for texture segmentation.

2. Proposed method

The proposed method uses a $n \times n$ sliding window for co-occurrence calculation and features extraction. For each input image pixel p, the window is positioned so that its first cell matches the position of p in the image. The co-occurrence information is calculated inside the window and associated to pixel p. Moreover, the features are extracted from these co-occurrences and also associated to p. For co-occurrence and features calculations we propose the gray-level co-occurrence indexed list (GLCIL) algorithm. The algorithm uses a list L of non-zero frequency (i, j) gray-level pairs associated with a matrix M of indexes. So, the position of each element in L can be accessed instantly by reading its index stored in the element (i, j) in M. This is important in the frequency count phase to avoid searching elements inside L. The feature extraction phase, takes advantage of the generally low dimension of L in comparison to M. The next step, in the proposed scheme, is the feature classification or clustering which generates the segmentation of input image. For this step, a k-means algorithm is used.

3. Experimental results

Three algorithms were developed for the co-occurrence and features calculation: the GLCM algorithm, the non sorted version of the GLCLL algorithm and the GLCIL algorithm. The purpose of the GLCM and GLCLL implementations is to evaluate the efficiency of the proposed GLCIL algorithm. A standard k-means implementation was used to accomplish the clustering phase. The experiments were divided in two sequences of tests. In the first sequence, the performance of the co-occurrence calculation algorithms is tested and, in the second sequence, the quality of segmentation is evaluated. For the co-occurrence calculation performance test, a Brodatz set of images was used. Figure 1 shows a percent comparison between the GLCIL algorithm computational time divided by the GLCM algorithm computational time.

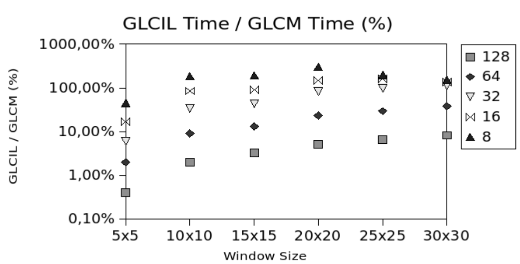

Figure 1. GLCIL time / GLCM time (%).

The x axis represents the window size and the y axis represents the percent value of the expression GLCIL Time / GLCM Time. Each data sequence corresponds to the results for a different quantization level value. The quantization levels used were 128, 64, 32, 16 and 8 gray-levels. Note that, for higher quantization levels, the superiority of GLCIL algorithm is considerable but reduces slowly as the window size increases. Moreover, for small quantization levels, GLCM performance increases. Finally, for the values of GLCIL Time / GLCM Time above 100% the GLCM algorithm performs better than GLCIL algorithm.

Next, we compare GLCIL algorithm against GLCLL algorithm. The results of this comparison are shown in Figure 2. In this case, the GLCIL algorithm performs better for all tested quantization levels and for all window sizes. Additionally, the advantage of GLCIL algorithm increases when window size grows. The quantization level presented low influence over final results but higher values tend to increase GLCIL algorithm advantage.

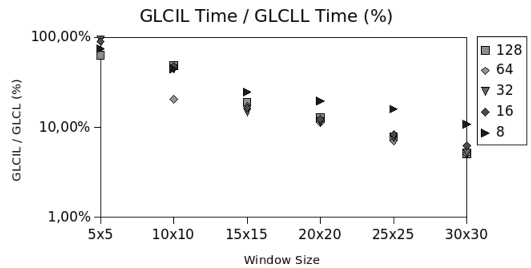

Figure 2. GLCIL time / GLCLL time (%).

The segmentation quality for a given image was highly influenced by the feature selected. That is, some features yield good results for certain images but produces very poor quality segmentation for other images. However, it was possible to find at least one feature that produced a good quality segmentation for each image tested.

4. Conclusion

In this work, a method based on co-occurrence calculation and k-means clustering for automatic image segmentation is presented. A new algorithm for fast co-occurrence calculation and two new statistics for texture extraction were proposed. Experimental results showed that the GLCIL algorithm is, in most cases, extremely faster than GLCM and GLCLL algorithms. Moreover, experimental results have demonstrated that proposed method achieves very good segmentation quality for a variety of situations and texture types. As a disadvantage, in some cases, it is necessary a trial and error process to find an appropriate set of features to get a good segmentation. So, the study of best features for these cases is a possibility of future research.

References

[1] D. A. Clausi and Y. Zhao. Rapid co-occurrence texture feature extraction using a hybrid data structure. *Computers and Geosciences*, 28:763–774, 2002.

[2] R. M. Haralick. Statistical and structural approaches to texture, 1979.

[3] M. E. Shokr. Evaluation of second-order texture parameters for sea ice classification from radar images. *Journal of Geophysical Research*, 96:10625–10640, 1991.