Chapter XVIII

Unsupervised and Supervised Image Segmentation Evaluation

Christophe Rosenberger, Université d’Orléans, France
Sébastien Chabrier, Université d’Orléans, France
Hélène Laurent, Université d’Orléans, France
Bruno Emile, Université d’Orléans, France

ABSTRACT

Segmentation is a fundamental step in image analysis and remains a complex problem. Many segmentation methods have been proposed in the literature but it is difficult to compare their efficiency. In order to contribute to the solution of this problem, some evaluation criteria have been proposed for the last decade to quantify the quality of a segmentation result. Supervised evaluation criteria use some a priori knowledge such as a ground truth while unsupervised ones compute some statistics in the segmentation result according to the original image. The main objective of this chapter is to first review both types of evaluation criteria from the literature. Second, a comparative study is proposed in order to identify the efficiency of these criteria for different types of images. Finally, some possible applications are presented.

INTRODUCTION

As Confucius said, “A picture is worth a thousand words.” This quotation means that an image contains lots of information. The goal of image analysis is to automatically extract this information. Segmentation is an essential stage in image analysis since it
conditions the quality of the interpretation. This processing either consists in partitioning an image into several regions or in detecting their frontiers. The classical hypothesis is that a good segmentation result guarantees a correct interpretation. This hypothesis makes sense clearly when the gray-level of each pixel is related to the interpretation task. For example, if we consider satellite images, the localization of the different types of vegetation in the image can be achieved with a segmentation method. In this case, the relation between the segmentation and the interpretation is very close. However, much more complicated situations can be encountered. If we have an indoor image containing some objects we want to identify, a good segmentation result will determine the frontier of each object in the image. In this case, a region containing an object is not characterized by a gray-level homogeneity and the level of precision of a segmentation result affects the understanding of the image.

Many segmentation methods have been proposed in the literature in the last decades (Jain, Duin, & Mao, 2000). A major problem in segmentation is the diversity in the types of regions composing an image. Indeed, an image can be composed of uniform, textured or degraded regions. Few segmentation methods provide good results for each type of region. Moreover, the efficiency of a new segmentation method is usually illustrated by only a few segmentation results on benchmark images, such as the Lena image. The problem is that this visual evaluation is still subjective. Thus, the comparison of different segmentation methods is not an easy task. Some techniques have been proposed to facilitate the visual evaluation of a segmentation result by using a colored representation. Furthermore, different metrics have been proposed to quantify the quality of a segmentation result. In order to make an objective comparison of different segmentation methods or results, some evaluation criteria have already been defined and literature is available. Briefly stated, there are two main approaches.

On the one hand, there are supervised evaluation criteria. These criteria generally compute a global dissimilarity measure between the ground truth and the segmentation result. They need two components. The first one is a ground truth corresponding to the best and expected segmentation result. In the case of synthetic images, this ground truth is known. In other cases (natural images), an expert can manually define this ground truth (Martin, Fowlkes, Tal, & Malik, 2001). Even if these images are more realistic, one problem concerns the objectivity and variability of experts. The second component is the definition of a dissimilarity measure between the obtained segmentation result and the ground truth. In this case, the quality of a segmentation result depends on the correct classification rate of detected objects in the image (Huet & Philipp, 1998). This type of approach is based on local processing and is dedicated to a given application.

On the other hand, there are unsupervised evaluation criteria that enable the quantification quality of a segmentation result without any a priori knowledge (Zhang, 1996). The evaluation of a segmentation result makes sense at a given level of precision. The classical evaluation approach is based on the computation of statistical measures on the segmentation result, such as the gray-level standard deviation or the contrast of each region in the segmentation result. The problem is that most of these criteria are not adapted for texture segmentation results (Bartels & Fisher, 1995). This is a major problem as, in general, natural images contain textured areas.

These criteria can be used for different applications. The first application is the comparison of different segmentation results for a single image. We could compare the
Related Content

Mind the Gap: Content-Based Image Retrieval and the User Interface
[www.irma-international.org/chapter/mind-gap-content-based-image/27065/](www.irma-international.org/chapter/mind-gap-content-based-image/27065/)

A Time Efficient Clustering Algorithm for Gray Scale Image Segmentation
[www.irma-international.org/article/time-efficient-clustering-algorithm-gray/78368/](www.irma-international.org/article/time-efficient-clustering-algorithm-gray/78368/)

Learning Aided Digital Image Compression Technique for Medical Application
Kandarpa Kumar Sarma (2016). *Handbook of Research on Emerging Perspectives in Intelligent Pattern Recognition, Analysis, and Image Processing* (pp. 400-422).
[www.irma-international.org/chapter/learning-aided-digital-image-compression-technique-for-medical-application/141645/](www.irma-international.org/chapter/learning-aided-digital-image-compression-technique-for-medical-application/141645/)

Fast and High Capacity Digital Image Watermarking Technique Based on Phase of Zernike Moments
[www.irma-international.org/article/fast-high-capacity-digital-image/68004/](www.irma-international.org/article/fast-high-capacity-digital-image/68004/)

Mobile Platform Challenges in Interactive Computer Vision
[www.irma-international.org/chapter/mobile-platform-challenges-in-interactive-computer-vision/163726/](www.irma-international.org/chapter/mobile-platform-challenges-in-interactive-computer-vision/163726/)