2.1 Rewritten by Han

Image segmentation is the process of separating an image into multiple parts of interests. This process can also be regarded as grouping adjacent pixels based on their similarity, the edges in the image and some other prior information. It is essential for many computer vision tasks, including object recognition, detection and tracking. Over the previous decades, a large number of algorithms have been proposed to get an appropriate segmentation of an image. The goal was to get segments of image that are as close as human perceive and understand. Obtaining the contour closure provides a reasonable method towards this issue. It runs by firstly getting the contours and then linking some contours together to get the closure, which at the same time, segments the area of interest [6], [10]. However, linking a few contours together to form a closure still remains an open question.

According to [2] and [6], early researchers tried to incorporate the prior information to constrain the contour grouping process, but prior information is difficult to encode quantitatively. As Shi and Malik [11] pointed out, the prior information can also be subjective for different human users. Therefore, there is no standard about how to use the prior information for the contour closures. Moreover, there were a few researchers, who tried to use shape features to group contours, and basically, they utilized the geometrical features like parallelism and symmetry to accomplish it [6]. However, geometrical features also include too many complexities. For example, there are other features like collinearity, orthogonal, and other shape change patterns. Thus, the difficulty embedded in this kind of method is also not a feasible option. Later, some other researchers simplified the shape features into weak shape priors, such as continuity and proximity; and they brought a concept named boundary gap, which is a measurement of missing edges along a possible closed contour [6]. Taking advantage of probability theory, Elder and Zucker [4] tried to compute the probability of a link between any two adjacent contour fragments and found contour cycles using a shortest path algorithm. On the other hand, Wang et al. [5] applied the ratio cut method to optimize a measure of average gap. However, all the above previous methods were too complicated to implement to cope with all possible challenges in real applications. But, Levinshtein et al. [2] took advantages of superpixels and made collections of them to overcome these computational complexity limitations. Thus, Levinshtein’s work became a starting point for our project.

Then, we also want to briefly introduce superpixels here. A superpixel can be thought as a group of adjacent pixels, and it can be the higher-level abstraction of pixels. The researchers in the superpixel field [2], [6], [10], [11] have showed that superpixels are conveniently useful for applications such as depth estimation, image segmentation, skeletonization, body model estimation and object location. There are many methods to generate superpixels for an image, and they generally use different algorithms and have various efficiencies. According to Achanta et. al. [10], there are two categories of superpixels: graph-based and gradient-ascent-based. In our project, we chose one graph-based method (Ncuts) and two gradient-ascent-based methods (SLIC and TurboPixels). Their details will be introduced in the section 3. After getting the superpixls for an image, Levinshtein et al. [6] reformulated the grouping superpixels as a mathematical optimization problem, and the outer boundary of the selected superpixels will be the contour closure for an object. Based on this result, we further the research by implementing it recursively for all possible objects’ boundaries in multiple levels. Moreover, we also intended to build a tree structure to show the relationships among the objects in an image, and we found that this idea was quite similar to some motivations of building a hierarchical structure for objects in an image once proposed by Arbelaez et. al. [8] and Shi and Malic [11]. The following part will introduce the basic framework of our project.