Improving Kraken's character segmentation using Energy Minimization

Berat Kurar-Barakat and Nachum Dershowitz

01.08.2022

We use the energy minimization framework [1] to improve Kraken's rough segmentation. We assume that each character segment from Kraken corresponds to one and only one character but, each character might consist of several connected components. Minimum of the energy function corresponds to a good extraction which urges to assign components to the label of the closest character segment while straining to assign closer components to the same label (Fig. 1). Let $\mathcal{L}$ be the set of character segments from Kraken, and $\mathcal{C}$ be the set of components in the binary document image. Energy minimization finds a labeling f that assigns each component $c \in \mathcal{C}$ to a label $l_c \in \mathcal{L}$, where energy function $\mathbf{E}(f)$ has the minimum.

$$\mathbf{E}(f) = \sum_{c \in \mathcal{C}} D(c, \ell_c) + \sum_{\{c,c'\} \in \mathcal{N}} d(c, c') \cdot \delta(\ell_c \neq \ell_{c'}) \tag{1}$$

The term D is the data cost, d is the smoothness cost, and δ is an indicator function. Data cost is the cost of assigning component c to label l_c. $D(c, \ell_c)$ is defined to be the Euclidean distance between the centroid of the component c and the centroid of the nearest character segment from Kraken l_c. Smoothness cost is the cost of assigning neighbouring elements to different labels. Let $\mathcal{N}$ be the set of nearest component pairs. Then $\forall \{c, c'\} \in \mathcal{N}$

$$d(c, c') = \exp(-\beta \cdot d_c(c, c')) \tag{2}$$

where $d_c(c, c')$ is the Euclidean distance between the centroids of the components c and c', and β is defined as

$$\beta = (2 \langle d_c(c, c') \rangle)^{-1} \tag{3}$$

$\langle \cdot \rangle$ denotes expectation over all pairs of neighbouring components [2] in a document page image. $\delta(\ell_c \neq \ell_{c'})$ is equal to 1 if the condition inside the parentheses holds and 0 otherwise.

Finally, a component c that traverses through several character segments is split by assigning each pixel in c to the label of the closest character segment.

References

1. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions on pattern analysis and machine intelligence **23**(11), 1222–1239 (2001)
2. Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In: ICCV. vol. 1, pp. 105–112. IEEE (2001)

Fig. 1. An example fragment image shown on the Kraken's character segmentation (yellow), the segments that contain a character (red), centroids of the character segments (green) and the resultant pixel labels from the energy minimization.