ISBN: 978-981-18-1791-5
2021 the 11th International Workshop on Computer Science and Engineering (WCSE 2021)
doi: 10.18178/wcse.2021.06.010

Integrating Segmentation and Association Relationship for Image Recognition

Xin Xu+

Science and Technology on Information System Engineering Laboratory
Nanjing Research Institute of Electronic Engineering
No. 99 Houbiaoyin Road, Nanjing, 210007, China

Abstract. Image segmentation generally refers to partitioning of an image into a set of regions that cover it. It has been believed that these regions may represent meaningful areas of the image, such as buildings, roads, forests, crops, animals and so on. The regions are usually composed of sets of pixels with similar colors. For interesting targets in the foreground, the regions may even take the forms of particular shapes, such as circles, eclipse or rectangles. Inspired from the concept of image segmentation above, it's interesting to further explore the relationship of segments for the targets of interest according to some association rules, i.e., the spatial association. In this way, we could detect objects from the images satisfying the criterion of both the segment and association relationships. For instance, people may probably prefer to query a scenery image complying with a certain style instead of containing a certain object. In this work, we propose an efficient object detection method integrating information of both segmentation and association relationship. Experimental results indicate that our method has more semantic flexibility for image recognition.

Keywords: contour discovery, image segmentation, image recognition; association.

1. Introduction

Image recognition refers to the process of recognizing the objects from an image. Image segmentation, on the other hand, refers to the process of partitioning an image into various regions according to certain criteria such that each region might represent some meaningful characteristic.

The tasks of image segmentation and image recognition are highly correlated with each other [1]. Some experts in computer vision have believed that image recognition is driven by image segmentation and that the image segmentation serves as a pre-processing step for image recognition. They suggested that the image segmentation help focusing on the relevant object features and pruning the redundant features outside the object [2, 3]. It was also thought that the image segmentation can extract shape information and reduce the background noise and thus the image recognition is facilitated [2].

However, semantic information of images has not been utilized much yet in existing work. The semantic information could be applied for region detection along with the raw image data. For this reason, we explore the integration of segmentation and association relationship for image recognition. Specifically, the image segmentation is performed first to discover the interested regions and the spatial association rules apply later to confirm the association relationships between the regions.

2. Related Work

Our work is closely related with the image segmentation and object detection techniques.

- Image segmentation

The image segmentation techniques have been widely used in various applications, such as medical image processing, face recognition, vehicle license plate recognition, hand-writing recognition and so on. There are generally five categories of image segmentation algorithms, including the threshold-based

+ Corresponding author. Tel.: + 8625 84285481.
E-mail address: xinxu_nriee@163.com.

segmentation, the regional growth segmentation, the edge detection segmentation [4], the clustering-based segmentation and the weakly-supervised CNN [5-6].

- Object detection

Object detection refers to the process of locating the objects of interest with a bounding box and assigns them a class label for each object. There are several benchmark state-of-the-art deep learning models to address the object detection problem. The most famous are RCNN [7-8] and YOLO [9]. The main problem of the above deep learning models is the high cost on both image samples and training.

Few semantic information between either the regions or the objects has been utilized in the above two categories.

3. Method

3.1 Template Construction

- Segmentation Template

The segmentation template is composed of the mean and standard deviation of HSV values and also the m number of Hu statistics for each segment. Table 1 illustrates the segmentation template for n segments:

Table 1: Segmentation template

	S_1	S_2	……	S_n
HSV mean	m_1	m_2	……	m_4
HSV Stand deviation	std_1	std_2	……	std_4
Hu Statistic 1	Hu_{11}	Hu_{21}	……	Hu_{n1}
Hu Statistic 2	Hu_{12}	Hu_{22}	……	-
……	……	……	……	……
Hu Statistic m	Hu_{1m}	Hu_{2m}	……	Hu_{nm}

- Association Relationship Template

We define two different segments *u* and *v* are *adjacent* to each other given a minimum ratio threshold *r*, if and only if there are at least $\min(|S_u|, |S_v|) \times r$ pixel members in segment *u* whose direct upper, down, left, right, upper left, upper right, lower left and lower right neighbors are within the segment *v*, and vice versa.

For instance, given a minimum ratio threshold *r* of value 0.5, segment *u* and *v* in Table 2 are not adjacent, since the number of pixel members in segment *v* whose direct neighbors are within segment *u* is two and that in segment *u* is three, both below the threshold.

Table 2: An instance of non-adjacent segments u (elements denoted as □) and v (elements denoted as Δ) given a minimum ratio threshold r of 0.5

□	□	□		Δ
□	□	□		Δ
	□	Δ	Δ	Δ
			Δ	Δ

We made use of the adjacency matrix to indicate the spatial relationships between different segments. For instance, a value of one would be assigned for two segments adjacent to each other and zero other vice. In more complicated cases, an alternative of multi-values could be applied, i.e., one for upper left, two for upper right, three for left, four for right, five for lower left and six for lower right. For example, Figure 1 illustrates an illustrating example of three segments, c_1, c_2 and c_3, where segment c_1 is on top of segment c_2 and segment c_3 is on the right side of segment c_2.

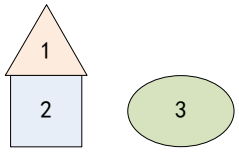


Fig.1: An illustrating example of three segments c_1, c_2 and c_3

The association relationships between the three segments are summarized in Table 3.

Table 3: An illustrating example for association relationships between three segments

	c_1	c_2	c_3
c_1	-	top	left
c_2	-	-	left
c_3	-	-	-

3.2 Image Segmentation

- Evaluation of Color Similarity

We first transformed the images into HSV (Hue, Saturation, Value) color space and then applied a hue threshold ρ to evaluate the color similarity between the potential image segments and the segmentation template. After the hsv transformation, we obtain the hue value of each pixel in range [0, 180]. We define the colors of two pixels p and q as similar when their color difference satisfies the equation below:

$$\min(\text{abs}(\text{hsv(p)} - \text{hsv(q)}), 180 - \text{abs}(\text{hsv(p)} - \text{hsv(q)})) \leq 180\rho \quad (1)$$

- Core Object Merging

An image segment related is assumed to be consisted of a set of core objects which are not only consistent in color similarity but also adjacent to each other in space. For this reason, we further applied a spatial threshold delta as well as the hue threshold ρ to detect the core object. Given the spatial threshold delta in the range of (0, 1] and the image size [w, h], we further define that Δ w = delta $\times$ w and Δ h = delta $\times$ h and a sliding window with size [2 Δ w+1, 2 Δ h+1]. We put forward the formal definition of core objects below:

Definition Core Object

Given a spatial threshold delta and a minimum ratio threshold minr, a pixel p is defined to be a core object in an image segmentation if and only if the number of pixels which are within Δ w distance away from p in X axis and within Δ h distance away from p in Y axis,

$$|x_p - x_q| \leq \Delta w \quad (2)$$

$$|y_p - y_q| \leq \Delta h \quad (3)$$

and whose color are similar to *p* w.r.t. the hue threshold ρ, is greater than $(2\Delta w + 1)(2\Delta h + 1) \times \text{minr}$:

$$\text{Sum(q|Eq(1)\&Eq(2)\&Eq(3))} > (2\Delta w + 1)(2\Delta h + 1)\text{minr} \quad (4)$$

- Segment Discovery based on Core Object Reachability

We further define that two core objects c_1 and c_2 are reachable to each other if and only if c_1 and c_2 are close in both X and Y axes w.r.t. threshold *delta*, $|x_1\text{-}x_2| \leq \Delta w$ and $|y_1\text{-}y_2| \leq \Delta h$, and the hue values of core objects c_1 and c_2 are similar w.r.t. threshold ρ. The formal definition of core object reachability is given below:

Definition Core Object Reachability

Suppose c_1 and c_2 are two core objects, then c_1 and c_2 are reachable from each other if and only if the spatial distances between them along the X and Y axes are within [-Δw, Δw] and [-Δh, Δh] respectively and the color difference in HSV color space satisfies that

$$\min(\text{abs}(\text{hsv}(c_1) - \text{hsv}(c_2)), 180 - \text{abs}(\text{hsv}(c_1) - \text{hsv}(c_2))) \leq 180\rho \quad (5).$$

Starting from a randomly selected core object in the image as the initial cluster, the image segmentation algorithm iteratively merges the current cluster with all the core objects which are reachable from any core objects within the cluster. The final cluster *fc* is formed when no more core objects are available for merging and the number and output as a candidate image segment if the ratio of the cluster size to the image size is above a pre-specified threshold ξ, $|fc| > \xi wh$, then another candidate image segment is explored from the remaining areas of the image.

3.3 Template Matching

- Segment Matching

Given the color information of the target to be detected, we iteratively merge the adjacent candidate segments and match with the invariant moments of the target segments. The contour of target object is identified when all or a majority of the invariant moments are matched.

The central moments of the target image are defined as shown Equation (6):

$$u_{pq} = \iint (x - \bar{x})^p (y - \bar{y})^q p(x, y) dx dy \tag{6}$$

The orthogonal invariants by Hu method include $u_{20} + u_{02}$, $(u_{20} + u_{02})^2 + 4{u_{11}}^2$, $(u_{30} - 3u_{12})^2 + (3u_{21} - u_{03})^2$, $(u_{30} + u_{12})^2 + (u_{21} + u_{03})^2$, $(u_{30} - 3u_{12})(u_{30} + u_{12})[(u_{30} + u_{12})^2 - 3(u_{21} + u_{03})^2] + (3u_{21} - u_{03})(u_{21} + u_{03})[3(u_{30} + u_{12})^2 - (u_{21} + u_{03})^2]$, $(u_{20} - u_{02})[(u_{30} + u_{12})^2 - (u_{21} + u_{03})^2 + 4u_{11}(u_{30} + u_{12})(u_{30} + u_{12})(u_{21} + u_{03})$.

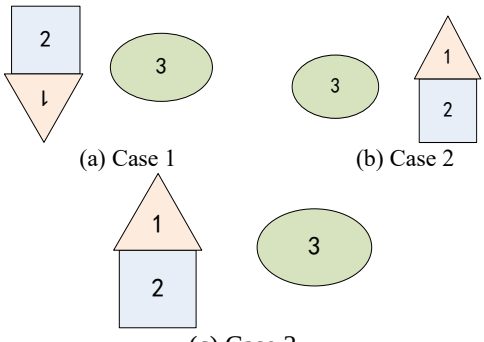
Suppose there are N number of objects in the template, denoted as o_1, o_2, ……, o_N respectively. And each object o_i has n_i different segments, which may have overlapping between each other. The total number of segments in the template is denoted as n, where $n = \sum_{i=1}^{N} N n_i$. We denote the number of Hu invariants of the segment template as m and the corresponding Hu invariants for segment j as h_{ij1}, h_{ij2}, …, and h_{ijm}. Upon that, we can compare each Hu invariant $h_{ijk}(s)$ of the segment sample against that of the segment template h_{ijk}, where $1 \le k \le m$. We define the sample s and template segment j of object i are matched if the average of their difference ratios is above 90%, as illustrated in Equation (7):

$$matched(i, j, k) = \begin{cases} true & if \ \frac{1}{m} \sum_{k=1}^{m} \min(\frac{h_{ijk}}{h_{ijk}(s)}, \frac{h_{ijk}(s)}{h_{ijk}}) \ge 0.9 \\ false & otherwise \end{cases} \tag{7}$$

- Association Matching

According to the definition of adjacency, the spatial association relationships between the matched segments are further evaluated with an adjacency matrix. A value of one would be assigned if the two segments c_i and c_j are associated, i.e., $Rel(c_i, c_j) = 1$ if segment c_i is at the right hand side of c_j, 2 if segment c_i is at the left hand side of c_j, 3 if segment c_i is at the top of c_j, 4 if segment c_i is at the bottom of c_j, and zero otherwise, $Rel(c_i, c_j) = 0$.

Suppose the set of segments matching segment template s_u, where $1 \le u \le n$, is denoted as $MSet(s_u)$. Then the task of association matching is to find a set of segments $\{c_1, c_2, \ldots, c_n\}$ such that $c_1 \in MSet(s_1)$, $c_2 \in MSet(s_2)$, ……, $c_n \in MSet(s_n)$ and $\forall 1 \le i, j \le n, i \ne j$, $Rel(c_i, c_j) = Rel(s_i, s_j)$.



(a) Case 1 (b) Case 2

(c) Case 3

Fig.2: Three cases of association templates

Given the three cases in Figure 2, we could infer that case 1 does not match the template in Table 3 since segment 1 is at the bottom of segment 2. Neither does case 2 since segment 3 is on the right hand of segment 1 and 3. Case 3 matches the association template since all the pairwise segments match.

4. Application

We applied our method for scenery query. Suppose we would like to find the scenery images complying with a similar style, i.e., high mountains above a big blue lake, while there are very few image samples of the mountains or lakes. In that case, the benchmark deep learning models are difficult to be trained and refined.

With our method, a scenery template could be set up for both segment matching and spatial association matching. Comparatively, our method is much more simple and understandable.

For the mountains, we set up the color variation range and also the Hu statistics for the templates. And for the lake, we loosed the shape constraints as the lake may be shadowed or concealed by trees. Instead, we specified he constraints of area and positions such that the area of the lake is no smaller than 1/4 of the image and the lake is situated below the mountains.

Figure 3 illustrate the image segmentations of two images and Figure 4 indicate the two images both comply with the same template. As we can see, although the mountains and lakes are from different places, they demonstrated similar styles.

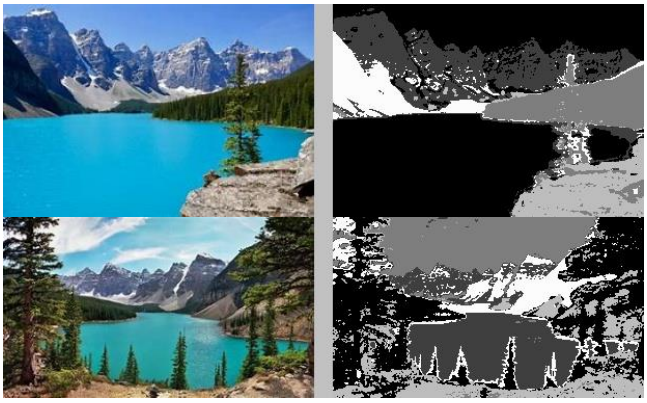


Fig. 3: Image segmentation results of two images

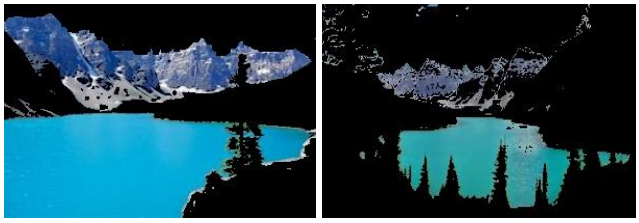


Fig. 4: The matched regions between two images

Different from benchmark object detection algorithms, our method does not require training the model with sufficient large number of samples with high cost. For the above example, we just need to specify the color and area. In combination with the association templates, images with similar styles instead of similar objects could be identified. Queries of similar architecture styles or life styles could be made as well.

5. Acknowledgements

Our work is sponsored by National Natural Science Foundation of China (Grant No. 61771177). This work was partially supported by Collaborative Innovation Center of Novel Software Technology and Industrialization.

6. References

[1] Karan Sharma, The link between image segmentation and image recognition, Dissertations and Theses, Portland State University, 2012.

[2] Malisiewicz, T., & Efros, A. Improving Spatial Support for Objects via Multiple Segmentations. Robotics Institute. Paper 280. http://repository.cmu.edu/robotics/280, 2007

[3] Rabinovich, A., Vedaldi, A. & Belongie, S. Does image segmentation improve object categorization? University of California San Diego Technical Report cs2007-0908, 2007.

[4] Senthilkumaran N, Rajesh R. Edge detection techniques for image segmentation–a survey of soft computing approaches[J]. International journal of recent trends in engineering, 2009, 1(2): 250-254.

[5] Lin D, Dai J, Jia J, et al. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 3159-3167.

[6] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. arXiv preprint arXiv:1606.00915, 2016.

[7] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks [C], NIPS 2015.

[8] Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick. Mask R-CNN, ICCV2017.

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You only look once: unified, real-time object detection, CVPR 2015.