Automating Grab-Cut Selection for Single-Object Foreground Images

Eugene Weiss
Stanford University
eweiss@sbcglobal.net

Abstract

Many photographs consist of a single subject against a background, with the subject located toward the center of the image to provide a focus. Using this knowledge, it should be possible to automate the creation of a selection box for the Grab-Cut algorithm to use for separating the foreground object from the background.

1. The Problem

Segmentation and outlining is a significant problem in computer vision. A library of properly indexed object, separated from their backgrounds, is important for building object identification learning sets. Sometimes, human segmentation has been used for this, but automated approaches are obviously preferred. One approach that has shown promise is the Grab-Cut algorithm, pioneered by Rother, Kolmogorov, and Blake (2004), with further work done by others, including Cui, Yang, and Wen (2008). Grab-Cut method requires some human input; the user must place a rectangular outline around the foreground object. This project hopes to automate this process for a subset of images.

Many photographs consist of a single object at the focus of the picture. If one restricts oneself to such images, the geography of the image contents can be used to roughly locate the foreground image, and create the separation lines needed for the Grab-Cut technique to be effective.

This project is conceived as creating one component among several that would allow the vast stores of images which have been collected and sorted by linguistic tags by search engines to be used by computer vision researchers without manual selection or preprocessing. It would best be used preceded by an algorithm that determines the likelihood that the image in question is a single-object foreground image, and succeeded by an estimate of the quality of the segmentation. While these additional problems are beyond the scope of this project, the software developed will output data that may be useful in solving them.

2. Datasets

The project will use a dataset comprised of a subset of the Berkeley Segmentation Dataset, and a set of images selected from Google's image database, chosen from the single-word search terms, such as "flowers", "trees", "persons", etc. All images are hand selected for having a single foreground object, sized such that an appropriate bounding box would have a large proportion of the total pixels in the image, with an approximate minimum being 30% and for most images being 50% or greater. 40 such images have been chosen from the Berkeley Segmentation Dataset, and another 40 from Google, with five images taken for each search term, with the exception of the term 'persons', where ten have been selected, five being full-body images, and five being head and shoulder portrait-type images.

The images chosen represent a good challenge for the algorithm. All have realistic backgrounds. (ie. no pre-segmented images with a white or black background) The majority contain objects completely contained within the image, but some have objects which are partially cropped, such as the head and shoulder portraits discussed above. These images are important to challenge the algorithm to deal with cases where at least one edge of the bounding box should be at the image boundary.

3. Method

The proposed method will iteratively take vertical and horizontal pixel slices from the edges of the image. These will be presumed to be background, and will be evaluated by a weighted list of criteria including energy, color and line-density. They will be compared to the remainder of the image, presumed to contain the foreground object, and will be progressively thickened until they begin to converge with the foreground data. At that point, the rectangular box defined by the slices will be used to run the Grab-Cut algorithm, as defined in the papers listed above. Jahangiri and Heesch (2009), have proposed a modified Grab-Cut for simple images with little variation in background color and relatively high contrast between foreground and background. The interactive trimap generation central to the original formulation of GrabCut is replaced by a tentative approximation of the background using active contours. I propose instead to use the original Grab-Cut, with automated box location, with the hope of achieving success with more complex backgrounds.

To deal with the cropped foreground images discussed above, a second run will be taken with the centers of the horizontal slices compared to their vertical ends, and vice versa, with the centers of the slices also compared to the central foreground values.

After examining several implementations of the Grab-Cut algorithm that are publicly available, I have chosen to use the implementation in the OpenCV library, version 2.1. This implementation seems much faster, while providing results that are as good or better than any of the Matlab-based implementations.

4. Readings

Rother, C., Kolmogorov, V., and Blake, A. (2004). Grab-Cutinteractive foreground extraction using iterated graph cuts. ACM Transactions on Graphics (Proc. SIGGR2004), 23(3):309314.

Cui, J., Yang, Q., Wen, F., Wu, Q., Zhang, C., Van Gool, ductive object cutout. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, AK.

Jahangiri, M.; Heesch, D., Modified grabcut for unsupervised object segmentation, 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo

A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. Interactive image segmentation using an adaptive GMMRF model. Proc. Eur. Conf. on Computer Vision, ECCV (2004). . PDF

V. Gulshan, V. Lempitsky, and A. Zisserman. Humanising GrabCut: Learning to Segment Humans Using the Kinect. Workshop on Consumer Depth Cameras for Computer Vision, Barcelona, 2011

A. Levin and Y. Weiss. Learning to Combine Bottom-Up and Top-Down Segmentation. Proc. of the European Conference on Computer Vision (ECCV), Graz, Austria, May 2006.

Shoudong Han, Wenbing Tao, Desheng Wang, Xue-Cheng Tai, Xianglin Wu: Image Segmentation Based on GrabCut Framework Integrating Multiscale Nonlinear Structure Tensor. IEEE Transactions on Image Processing 18(10): 2289-2302 (2009)

Cui, J., Yang, Q., Wen, F., Wu, Q., Zhang, C., Van Gool, ductive object cutout. In IEEE Computer Society Connference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, AK.

Jingyu Cui, Qiong Yang, Fang Wen, Qiying Wu, Changshui Zhang, Luc J. Van Gool, Xiaoou Tang. Transductive object cutout. In Proceedings of CVPR'2008.

Jahangiri, M.; Heesch, D., Modied grabcut for unsupervised object segmentation, 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo

4.1. Evaluation

Results will be compared against human-aided Grab-Cut, as well as against the human-segmented data from the Berkeley data set. A numerical error rate calculation will be done, based on the one in Rother, Kolmogorov, and Blake. The comparison will be between the hand-selected segmentation data for the images taken from the Berkeley dataset, and for both sets between Grab-Cut using hand-selected bounding boxes, and the automated selection using the project's technique. In no case is it anticipated that this technique will produce better quality results than hand-selected bounding boxes, so the delta between the two approaches will be considered error. The Grab-Cut algorithm is itself imperfect, and it is anticipated that images that do poorly with it will do even more poorly with the automatically generated bounding boxes. Among the images in the datasets, hand-selected Grab-Cut does worst with images from the Berkeley dataset which consist of an animal in an environment in which it's skin color and pattern are evolved for natural camouflage. The problem here is not surprising, as millions of years of evolution have gone into making this difficult, but these images have been retained in the dataset precisely because the largest deltas will be found, and hence provide the best opportunity for optimizing the algorithm.