Automating Grab-Cut Selection for Single-Object Foreground Images

Abstract

This paper investigates a technique to apply the Grab-Cut algorithm to single-subject images without requiring human delineation of a bounding box to separate foreground from background pixels. The proposed technique uses a weighted combination of several known methods of distinguishing pixels in regions of an image. Implemented are means of hue and intensity, Canny line density and Sobel gradients.

1. Introduction

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. Search engines and other aggregators have indexed vast numbers of images; Google alone had indexed 1 billion images already in 2005. The images are indexed semantically, but vary widely in both the accuracy of their indexing and in types of images. 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.

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. 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.

2. Grab Cut

The Grab-Cut algorithm, pioneered by Rother, Kolmogorov, and Blake (2004), uses a minimum-energy Gaussian Mixture Model algorithm to find and discard background pixels in a foreground area delineated by a human-defined rectangular bounding box. Consequently, the most important criteria for a good result is having a good selection of all types of background pixels outside of the bounding box when the Grab Cut algorithm is run. While restricting the bounding box to the tightest possible dimensions around the foreground object is desirable, it is not essential, provided pixels from all of the background features are significantly represented outside of it. If the bounding box cuts through the foreground object, then worse results are to be expected, with some foreground pixels removed. Also, existing implementations of Grab Cut consider all pixels outside of the bounding box to automatically be considered as background, and are hence removed, although one could, conceivable, implement Grab cut in a way that follows the minimum energy snake outside of the box. Grab Cut is iterative, and for this project all examples of Grab Cut usage, both automated and manual, use a two iterations. This number was chosen because a subjective look at manually bounded Grab Cut segmentations using the dataset showed this to give the best result in most instances. The implementation of Grab Cut used in this project was from the function included in the OpenCV library, version 2.2. 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.

3. Datasets

The project used 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, "flowers", "trees", "persons", “trees”, “houses”, and “ambulances”. 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 25% 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.

4. Algorithm

The algorithm used iteratively takes vertical and horizontal pixel slices from the edges of the image. These were presumed to be background, and were evaluated by a weighted list of criteria including the mean of the hue, the mean of the saturation values, Sobel gradients and line-density. The line density calculation used the Canny line detector algorithm, with They were compared to the remainder of the image, presumed to contain the foreground object, and were progressively thickened until they begin to converge with the foreground data. A two-pass algorithm was used, with the first pass being a naive, and comparing all the pixels on the foreground side of the line to all the pixels on the background side.

Image1. The red band below shows the proposed slice; all pixels above are considered potential foreground for the naive pass.

A preliminary bounding box was created by the naive pass, and a second pass was then made. For the second pass, the boundaries for all of the edges other than the working edge being evaluated were considered to belong to the foreground, but the background was only calculated for the area outside of the working edge, and between the two edges perpendicular to it (see Image 2).

Image 2. The grayed out area represents the proposed foreground after the first pass, and the red area shows the background slice for the second pass.

At that point, the rectangular box defined by the slices was used to run the Grab-Cut algorithm, as defined in the papers listed above. The slices are started seven pixels from the image boundaries to avoid edge effects. If the best differentiation was found to be at the initial slice, then the boundary for that edge are pushed out to the edge of the image. All edges are pushed out two pixels to avoid clipping small projections from the foreground object.

5. Results

The results were mixed, but promising for future work. The methodology for determining the error rate was very simple; the pixels of the automated Grab Cut were compared to those of the manually bounded sample. All differences were considered to be errors of the automated version, which was usually, but not always the case. An error of .12064 was found with weightings of for hue of 11, saturation 8 , Sobel gradient 3 and edge density 3. The best performing images were those with the most clearly defined foreground/background separation, as might be expected. The algorithmic performed worst where there was little difference between the foreground and the background, or when the image intersected with one of the edges. The algorithm for handling edge intersections was clearly inadequate.

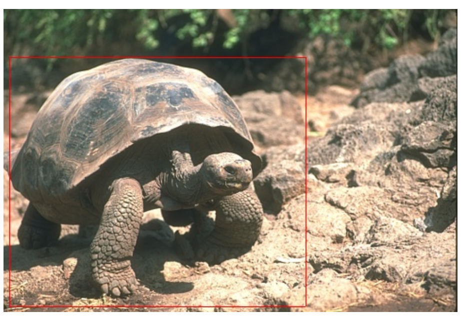

Image. 3. These images show a good result, despite a challenging background.

Image 4. This image illustrates the problem with edge-intersecting object images.

Image 5. With portraits, the algorithm always focused on the face, rather than integrating the torso as well.

6. Conclusions

The error rate for this process at this time still precludes good aesthetic results, but since the most serious problems seem to involve cropping the foreground, rather than leaving the background intact, it may be useful even in its current state for creating feature libraries. Future work to improve this algorithm will include adding additional differentiators to the pixel analysis, especially a better texture descriptor than Sobel gradients. The algorithm can likely also be improved by adding a threshold level for moving the boundary inward; that would both avoid cropping and help with the edge-intersecting object image problem. It would likely also be helpful to compare only the portion of the foreground nearest the working edge during the second pass to create a greater differentiation. Han, Tao, Wang, Tai, and Wu [2009] have integrated an MLNST texture discriminator into the Grab-Cut algorithm. Anything that improves Grab-Cut generally will improve the performance of its automated version. This code creates the opportunity to numerically describe the difference between the interior and exterior of the box. It may be possible to use this description to predict the likelihood that the image is a single subject image, or that it is a good candidate for automated Grab-Cut.

7. References

Rother, C., Kolmogorov, V., and Blake, A. (2004). “GrabCut”—Interactive Foreground Extraction Using Iterated Graph Cuts. ACM Transactions on Graphics (Proc. SIGGR2004), 23(3):309–314.

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.