A Literature Review of Image Segmentation Techniques and Matting for the Purpose of Implementing “Grab-Cut”

by: Matthew Marsh
email: g02m1682@campus.ru.ac.za
4 Montague Road
Grahamstown

Abstract

This literature review attempts to provide a brief overview of some of the most common segmentation techniques, and a comparison between them. It discusses the “Grab-Cut” technique, and reviews some some of the common matting techniques. The graph cut approaches to segmentation can be extended to 3-D data and can be used for segmenting 3-D volumes. Other segmentation techniques use either contour or edge segmentation to perform segmentation. The graph cut techniques use both contour and edge information. The main matting techniques are Poisson matting and probabilistic alpha matting using colour statistics. Poisson matting works directly on the alpha matte of the image, and is interactive. The statistical approach uses Gaussians to model colour statistics in the image and is not interactive.

1 Introduction

Image segmentation is the process of separating or grouping an image into different parts. As partial fulfilment of my degree I intend to implement the method of image segmentation presented in the “Grab-Cut” paper. I intend to do this by creating a plugin for the open source graphics package GIMP which will perform image segmentation using the methods presented in the “Grab-Cut” paper. There are currently many different ways of performing image segmentation, ranging from the simple thresholding method to advanced graph-cut methods. “Grab-Cut” is a new innovative segmentation technique that uses a graph-cut approach to minimize an energy function, and in doing so obtain a segmentation of an image. “Grab-Cut” uses a graph to represent an image, and then segments this graph by using a Min-Cut/Max-Flow algorithm. “Grab-Cut” also makes use of a matting technique for regions that do not have clear boundaries. In order to obtain a good grounding in the workings of “Grab-Cut” this document reviews the literature relating to:

1. Image Segmentation, and in particular “Grab-Cut”, as well as some of the energy minimization techniques that make use of graph-cut.

2. Matting - The different techniques that can be used to pull an alpha matte from an image.

2 Image Segmentation

Image segmentation is the process of separating or grouping an image into different parts. These parts normally correspond to something that humans can easily separate and view as individual objects. Computers have no means of intelligently recognising objects, and so many different methods have been developed in order to segment images. The segmentation process in based on various features found in the image. This might be colour information that is used to create histograms, or information about the pixels that indicate edges or boundaries or texture information (Ballard and Brown, 1982).

2.1 Thresholding

Thresholding is a very simple form of segmentation. A threshold is defined, and then every pixel in an image is compared with this threshold. If the pixel lies above the threshold it will be marked as foreground, and if it is below the threshold as background. The threshold will most often be an intensity or colour value. Other forms of thresholding exist where the threshold is allowed to vary across the image, but thresholding is a primitive technique, and will only work for very simple segmentation tasks.

Most graphics packages come with some form of thresholding segmentation. “Magic Wand” is such a tool, and is included in Adobe Photoshop 7. With this tool the user will select a seed or multiple seed pixels and set some form of tolerance level. The segmentation is then performed by testing all pixels against the set tolerance level(INCORP, 2002). This technique is easy to use, but the results are most often unsatisfactory and finding the correct tolerance level can be cumbersome, and sometimes even impossible (Rother et al., 2004).

2.2 Live-Wire Boundaries

This technique uses dynamic programming to solve a 2-D graph searching problem and, in doing so, detect boundaries in an image. For this technique a 2-D graph is created, where pixels in the image are represented by nodes in the graph. The edge weights between nodes are defined by cost functions that depend on features found in the image. The goal of the 2-D graph search is to find the minimum cost path between a start and goal node.

1. Mortensen and Barrett (1995) have developed an approach to creating an interactive tool called Intelligent Scissors that uses a live-wire boundary to perform segmentation. When the user moves the mouse near a boundary the live-wire snaps to this boundary. This algorithm greatly reduces sensitivity to noise, and selects the mathematically optimal boundaries nearby. It also incorporates on the fly training so that the live-wire snaps to the current type of edge being followed rather than just the strongest one in the neighbourhood.

2. Four years later Mortensen and Barrett wrote another paper which improves on their previous approach. This approach entails over-segmenting the image into regions using a technique called tobogganing and then imposing a weighted planar graph on top of the resulting region boundaries. Tobogganing creates a region-based graph rather than a pixel-based graph. This is much smaller than a pixel-based graph and so it provides much faster graph searches allowing quicker results, and so better user interaction. It also provides a representation of possible live wire boundaries and allows the user to choose which one to use as the object is segmented (Mortensen and Barrett, 1999).

2.3 Snakes / Energy Minimizing Splines

Snakes are energy minimizing splines that are guided by external constraints and internal constraints, and are influenced by image forces that pull them towards features like lines and edges. They are active in that they lock onto nearby edges. Snakes are so called due to the wriggling motion they undergo while minimizing their energy functions. They are designed to be interactive, in that the user must give some clues as to where abouts the boundaries are, and then snakes are used to minimize the energy and so trace the contour or boundary (Kass et al., 1987). Snakes work on the assumption that edges are found not only by looking at the local gradient, but also at the long range distribution of the gradient. This is done by using curvature constraints as well as continuity constraints. Snakes have an internal energy function which determines their elasticity and rigidity, and an external energy function based on image information and user interaction.

The classic implementation of snakes by Kass et al. (1987) allows the problem to be reduced to a matrix form. However this puts constraints on the energy functions. Davison et al. (2000) propose a less complicated form of the energy functions, and energy minimization is carried out by adjusting individual vertices on the snakes. This allows for a greater range of energy functions, and the addition of internal energy functions like area and symmetry terms without complicating the minimization process as would be the case with the classic implementation.

2.4 Graph Cut Techniques

2.4.1 Graph Cuts as a energy minimization technique

The use of graphs to solve energy minimization problems has become more and more popular in the context of low level computer vision(Boykov and Kolmogorov, 2004). Many problems in computer vision can be reformulated as an energy minimization problem. Energy minimization has in the past been computed by using dynamic programming (which only works in very simple cases) and simulated annealing (which is very slow). Recently there have been many approaches to energy minimization by using graph-cut techniques, and in most cases each graph has had to be specially created for solving a specific energy minimization problem (Kolmogorov and Zabih, 2004). Kolmogorov and Zabih (2004) look at the types of functions that can be minimized by graph cuts and they give a general construction of the graphs that can solve different classes of problems.

All segmentation techniques using graphs use some form of a graph-cut algorithm to segment the graph into two regions, and in doing so minimize the energy. Boykov and Kolmogorov (2004) provide a comparison between current Min-Cut/Max-Flow algorithms with regards to their efficiency. This paper includes Goldberg-Tarjan style “Push-Relabel” methods and Ford-Fulkerson style “Augmenting Path” methods. It also introduces a new algorithm which they claim works several times faster than all other known methods in most cases. The paper uses energy minimization techniques in the context of image restoration, stereo and image segmentation in order to compare the speed of the different methods

2.4.2 Graph Cuts for image segmentation

1. Boykov and Jolly (2001a) introduce a technique for segmentation using a graph to represent the image, and then using a Min-Cut/Max-Flow algorithm to segment the graph. Pixels in the image are represented by nodes on the graph. The edge weights on the graph are defined by a cost function, which is defined by region and boundary information from

the image. A Min-Cut/Max-Flow algorithm is then used to segment the image by minimizing the cost function. This technique has a well defined cost function and so provides a globally optimal solution.

2. Li et al. (2004) introduce an interactive means of segmentation based on graph minimization techniques very similar to Boykov and Jolly (2001a). It involves two major steps. The first step is the object marking stage in which hard constraints are set so that regions are set to be either definitely background or definitely foreground. They use graph cuts with an over-segmentation technique which greatly increases the speed of segmentation, and can provide segmentation results quickly to the users. The second step is a boundary editing step in which individual vertices on the segmentation can be moved around until the user is satisfied.

3. Interactive segmentation is becoming more and more popular and it is far preferred over fully automatic segmentation methods that are never perfect (Boykov and Jolly, 2001b). There have however been some methods that try to perform automatic segmentation. Blake et al. (2004) describe a method of image segmentation that operates without any hints from the user. Most image segmentation techniques use parameters set by the user. This paper tries to get an algorithm to learn parameters from image data, and then perform the segmentation based on these parameters. They use a database of images with the correctly segmented results to test their approach. The percentage error using interactive approaches was considerably lower than this approach (Blake et al., 2004).

2.4.3 "Grab-Cut"

Rother et al. (2004) present an innovative way of segmenting an image into a foreground and background region. The image to be segmented is represented by a graph which is constructed in such a way that a minimum cost cut on the graph will produce a best segmentation of the image. Pixels in the image are represented by nodes on the graph. The edge weights of the graph are then defined by a cost function / energy function. This cost function will depends on both boundary and regional properties of the pixels. A Min-Cut/Max-Flow technique is used in order to segment the graph. Boykov and Kolmogorov (2004) have the complete details as to how the graphs can be constructed.

"Grab-Cut" uses the technique of segmentation described by Boykov and Jolly (2001a), but extends the technique by allowing colour images (the original technique only works for grey-scale images) and implementing incomplete labeling (the user only has to specify hard con-

Table 1: Comparison Between Segmentation Techniques

Segmentation Technique	Information used in Segmentation	Extends to N-D	Method of Segmentation
Thresholding	Colour	No	Simple Comparison Between Pixels And The Threshold
Live Wire	Edge	No	2D Graph Search Using Dynamic Programming
Snakes	Edge	No	Energy Minimization of Internal and External Functions
Graph cut	Colour and Edge	Yes	Max-Flow/Min-Cut Algorithm For Energy Minimization

straints for either the background or foreground). The algorithm that is presented is interactive and allows the user to relabel pixels after the initial segmentation in performed.

"Grab-Cut" also provides a border matting feature for use in areas of the image that are "fuzzy" and don't have clear boundaries. Boykov and Jolly (2001a) use gray-scale histograms in order to model the intensities of the pixels in the image, however the "Grab-Cut" technique uses Gaussian mixture models in order to model the colour statistics of the image.

Table 1 shows a comparison between the various segmentation techniques with regard to the information they use to perform the segmentation, if the method can be extended from 2-D to N-D data, and the method that is employed in order to perform the segmentation. Segmentation techniques are used primarily for 2-D images, but some segmentation techniques can be extended to 3-D data (or volumes) and even N-D data. From table 1 it can be seen that graph cut techniques are able to be extended to 3-D, and so can be used to segment 3-D models. Graph cut techniques also make use of more information from the image compared to the other techniques, giving them a broader range of cost functions to work with.

3 Matting

The boundaries of objects in images are sometimes well defined and can easily be traced with some sort of edge detection mechanism. Natural or organic objects often have very intricate boundaries. Hair is very hard to segment and to trace with standard edge detection methods. Water spray, smoke and leaves are also very difficult to segment using any of the standard techniques. The difficulty arises because of the limited resolution of cameras. If a pixel in the camera receives light from more than one source (as will happen with very fine objects) the pixel will take on a colour that is a mixture of the colour from the two objects. The percentage that a colour contributes to the overall colour of a pixel is referred to as the alpha value of the colour. Matting techniques try to segment images by calculating this alpha value (Ruzon and Tomasi, 2000). The problem of matting is inherently under constrained, since for each pixel that is mixed, there are an infinite number of combinations of what the colour of the object is and its alpha value (Chuang et al., 2001).

Most matting techniques need a tri-map of the image before they can pull a matte. The tri-map specifies which regions of the image that are background, foreground, and an “unknown region” where matting must take place.

3.1 Probabilistic Alpha Estimation Using Colour Statistics

Ruzon and Tomasi (2000) describe a matting technique of using two un-orientated Guassians distributions to model the colours in the foreground and background region. All pixels in the unknown region are then said to fall within the colour space between these two Gaussians. The alpha value of pixels in the unknown region can then calculated, as well as the unmixed colours for each pixel.

The technique proposed by Chuang et al. (2001) is very similar to that of Ruzon and Tomasi (2000). They also build Gaussian probability distributions for the foreground and background, however it differs in that already computed values of alpha are used in calculating un-computed nearby values. This is referred to as a “sliding window” for neighbourhood definitions. They also formulate the problem of calculating the matte values as a Bayesian Framework. This means that the algorithm uses Baye’s theorem to calculate the most likely values for a pixels alpha value, and the foreground and background colours of the pixel.

3.2 Poisson Matting

Ruzon and Tomasi (2000) and Chuang et al. (2001) use colour samples from the background and foreground regions to perform the matting using statistical information about these colour samples. These approaches are not interactive, and once the tri-map is specified no further editing of the matte can be performed. The Poisson matting approach by Sun et al. (2004) works directly on the gradient of the matte. This solves the effects of the wrong classification of colours in complex scenes in the statistical methods. The gradient of the matte is estimated in the image, and then recreated by solving Poisson equations. A smooth gradient is assumed between the foreground and background. This technique solves for a global matte gradient, and then allows the user to edit the gradient with various tools in a local manner. This technique produces better results than Bayesian techniques.

Perez et al. (2003) discuss techniques as to how images can be seamlessly blended together as well as how portions of an image can be deformed and seamlessly integrated using Poisson equations.

Table 2: Comparison Between Matting Techniques

Matting Technique	Method Used	Interactive
Probabilistic Alpha Estimation	Colour Statistics are used to create Gaussians	No
Poisson Matting	The Alpha matte is modeled using Poisson equations	Yes

Table 2 show a comparison between the different matting techniques with regards to the method used to perform the matting, and if the technique is interactive. "Grab-Cut" makes use of a probabilistic approach to matting. A smooth alpha profile is assumed, and dynamic programming is used to minimize an energy function in order to calculate a smoothly varying profile. Gaussian parameters for the foreground and background are estimated by using near-by known values. "Grab-Cut" makes use of the approach by Chuang et al. (2001) to estimate the foreground colour of pixels.

4 Conclusion

There are many different techniques that perform segmentation. The main categories being Thresholding, Snakes, Live-Wire Boundaries and Graph cut. Thresholding is a very simple method and not suitable for complex segmentation. Snakes and Live-Wire boundaries use edge information in the image to perform segmentation. They are both interactive, but do not work well for objects which don't have well defined boundaries. Graph cut techniques use colour and edge information to perform segmentation. With the new Max-Flow/Min-Cut algorithm the energy minimization process is fast enough to provide an interactive method of segmentation using graph cuts. "Grab-Cut" appears to be an efficient and powerful graph cut technique, and includes a border matting feature for badly defined object boundaries.

References

Ballard, D. H. and Brown, C. M. (1982). *Computer Vision*. Prentice-Hall Inc., New Jersey.

Blake, A., Rother, C., Brown, M., Perez, P., and Torr, P. (2004). Interactive image segmentation using an adaptive GMMRF model. In *Computer Vision - ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic*, pages 428 – 441.

Boykov, Y. and Jolly, M.-P. (2001a). Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In *International Conference on Computer Vision (ICCV)*.

Boykov, Y. and Jolly, M.-P. (2001b). Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume I, pages 105–112.

Boykov, Y. and Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137.

Chuang, Y.-Y., Curless, B., Salesin, D. H., and Szeliski, R. (2001). A Bayesian approach to digital matting. In *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR 2001)*, volume II, pages 264–271.

Davison, N. E., Eviatarc, H., and Somorjaic, R. L. (2000). Snakes simplified. *Pattern Recognition*, 33(10):1651–1664.

INCORP, A. S. (2002). *Adobe Photoshop User Guide*.

Kass, M., Witkin, A., and Terzopoulos, D. (1987). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331.

Kolmogorov, V. and Zabih, R. (2004). What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26 NO. 2.

Li, Y., Sun, J., Tang, C.-K., and Shum, H.-Y. (2004). Lazy snapping. *ACM Transactions on Graphics*, 23(3):303–308.

Mortensen, E. N. and Barrett, W. A. (1995). Intelligent scissors for image composition. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 191–198. ACM Press.

Mortensen, E. N. and Barrett, W. A. (1999). Toboggan-based intelligent scissors with a four parameter edge model. In *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR '99)*, volume II, pages 452–458, Fort Collinsw CO.

Perez, P., Gangnet, M., and Blake, A. (2003). Poisson image editing. *ACM Transactions on Graphics*, 22(3):313–318.

Rother, C., Kolmogorov, V., and Blake, A. (2004). "GrabCut": interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23(3):309–314.

Ruzon, M. A. and Tomasi, C. (2000). Alpha estimation in natural images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 18–25, Hilton Head Island, SC.

Sun, J., Jia, J., Tang, C.-K., and Shum, H.-Y. (2004). Poisson matting. *ACM Transactions on Graphics*, 23(3):315–321.