EMCUT

C Rother, P. H. S. Torr
Microsoft Research, 7 JJ Thomson Avenue,
Cambridge, CB3 0FB,UK

Abstract

We present a new interactive tool for image segmentation. One of the key goals, in the design of such a system, is to minimize the amount of work that a potential user has to undertake to achieve the desired result. The algorithm must be fast to allow immediate interaction, and repeatable so that the same set of interactions produces the same result each time. With this in mind a new algorithms for segmentation is derived, EMCUT which is a 'one click' solution. The formulation is Bayesian, and the prior is constructed so as to combine region and boundary information in a principaled manner. It can be shown that the algorithm will converge to (at least) a local optimum. Because of this the results are entirely reproducible. In the rare cases that the algorithms converge to local minima usually one or two more user 'clicks' can guarentee convergence to the global optimum.

1 Introduction

Within this paper we address the problem of interactive segmentation of images, e.g. to allow the user to cut out part of an image (such as a face) and paste it into another. This has found a variety of uses in medical imaging, movie special effects and the application we consider here: home editing of digital photographs. The wide variety of content within images and the differing desires of users precludes a fully autonomous system, so the goal is to make the interactive segmentation process as painless as possible by minimizing the amount of time (mouse clicks) required to achieve the desired segmentation. Many exisiting interactive techniques for segmentation require significant user input.

Previous work in this theme has included the Magic Wand of Adobe [1], which computes a region of connected pixels similar to a pixel chosen by a mouse click, however it cannot work well in textured regions. Another class of approaches attempts to segment the image by finding a boundary that coincides with strong edges in the image, ignoring all other pixel information. These include (a) LIVEWIRE [9] in which the user clicks on two points on the boundary and dynamic programming is used to find the minimum cost path between these two points, on a graph formed on the image pixels with the arcs having weight inversely proportional to the edge strength between pixels. (b) JETSTREAM [11] following a more probablistic approach requires the user to click repeatedly on the boundary of the object and from this generates putative boundaries, along high contrast edges, around the object using a particle filter. (c) SNAKES [8] define an active contour around the object to be segmented and then iteteratively minimize the contour's energy functional to achieve the optimal boundary, combining external forces such as gradient energy with internal forces like curvature to achieve an ideal segmentation. There are systems that then allow the user to interactively modify the energy landscape and thus nudge the snake [6].

All of these method require many mouse clicks as in most typical images there are so many edges present that the cost surface is abundent with local minima. In particular they are often confused by T-junctions or other such edges in the background. In order to make the algorithm more effective a richer set of observations must be used. Where might these observations come from? Recently a new class of INTERACTIVE GRAPH CUT (ICG) methods have been proposed [4] that consider the distribution of the colour of the foreground and background pixels in addition to the boundary pixels. Furthermore a key advantage of (ICG) is that it can also apply to $N - D$ images (volumes) opening it up to be applied for segmenting many types of medical volumetric images such as provided by computed tomography (CT), or video. It transpires that the combination of region and boundary information to assist interactive segmentation leads to far less effort for the user. The ICG allows for both terms to combined to form a posterior likelihood by using a Markov Random Field (MRF) formulation. Within this paper we show demonstrate a clear improvement to the (ICG) which has two parts, which we dub EMCUT The first is that the initialization is in the form of a rectangle whose only constraint is that it bounds the object of interest (this is the 'one click' initialization). The second is the application of the EM algorithm to learn a mixture model for the colour of the foreground and background object. This involves exactly specifying the probablistic formulation which was left somewhat hazy in [4]. The algorithm proceeds by conditional maximization holding sets of the parameters constant whilst maximizing the posterior over the remainder. The structure of the paper is as follows Section 2 briefly overviews the theory of Markov Random Fields. Section 3 outlines our likelihood model and how we In Section4 it is shown how to combine boundary information into the posterior. Some results and discussion are given in Section 5.

2 Markov Random Fields

Within this section the Markov Random field MRF [2, 3, 5] model is described. It is composed of two terms, the likelihood $l(\mathbf{y}|\mathbf{x})$ and the prior $p(\mathbf{x})$. It will be shown that the solution of the Markov random field can, in certain circumstances, be formulated as a quadratic or more generally a semidefinite program.

Markov random field models provide a powerful framework for specifying nonlinear spatial interactions between features. Following Besag [3], we shall next define an MRF on an image I being a rectangular arrays of pixels. Each pixel can take one of 2 labels, labeled, $-1, +1$, corresponding to the background (-1), and the foreground (+1). Thus the sample space, for an image with n pixels, is $\mathbf{\Omega} = \{1, 2\}^n$. An arbitrary labeling of the image will be denoted $\mathbf{x} = (x_1 \ldots x_n)$ where x_i is the labeling of pixel i. Let $\underline{\mathbf{x}}$ be the true (but unknown) labeling, which is considered a realization of a random vector $\mathbf{X} = (X_1 \ldots X_n)$ where X_i is the label of pixel i. Let y_i denote the observation at pixel i, which could be a continuous variable, and $\mathbf{y} = (y_1 \ldots y_n)$ which is a realization of a random variable $\mathbf{Y} = (Y_1 \ldots Y_n)$. Assuming that the random variables $Y_1 \ldots Y_n$ are conditionally independent and that each Y_i has the same known conditional density function $f(y_i|x_i)$, dependent only on x_i, then the likelihood can be written

$$l(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{i=n} l(y_i|x_i). \tag{1}$$

Next the MRF prior is defined.

Let $\{p(x)\}$ be a probability distribution over labelings of the image I. Consider the conditional probability, $\Pr(x_i|x_j : j \in I \setminus i)$, of the labeling x_i occurring at pixel i conditioned on the labeling of all the other pixels in the image $I \setminus i$. In most work only local dependencies are considered, thus for each pixel a neighbourhood is defined for that pixel $\mathcal{N}(i)$ e.g. for images this might be the four nearest neighbours. For these sorts of fields it is assumed

$$\Pr(x_i|x_j : j \in I \setminus i) = \Pr(x_i|x_j : j \in \mathcal{N}(i)). \tag{2}$$

One important restriction, given by the Hammersly-Clifford theorem, is that the neighbourhoods must be symmetric i.e. if i is a neighbour of j, j must be a neighbour of i. Within this paper we concentrate on pairwise interactions: $\{p(\mathbf{x})\}$ is called a pairwise interaction MRF's if, for any sample space $\mathbf{\Omega}$

$$p(x) \propto \exp\left\{\sum_{i=1}^{i=n} G_i(x_i) + \sum_{i=1}^{i=n}\sum_{j=1}^{j=n} G_{ij}(x_i, x_j)\right\}, \tag{3}$$

where to ensure the Markov property $G_{ij} = 0$ if pixels i and j are not neighbours. This us just equation (9) of Besag [3].

The simplest useful MRF is first order in which only the four immediate neighbours of the pixel are considered (left, right, above, below); as shown in Fig. **??**. Second order includes the eight nearest neighbours. To obtain the maximum a posteriori MAP estimate of the MRF, $\hat{\mathbf{x}}$ is chosen to have maximum probability given the $\mathbf{y}$ i.e.

$$\max_{\hat{\mathbf{x}}} \Pr(\mathbf{x}|\mathbf{y}) \propto l(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \tag{4}$$

Greig, Porteous and Seheult [7] show that (3) can be solved exactly for the case when each pixel is to be classified into one of two unordered labels or categories. The prior distribution for their model is

$$p(\mathbf{x}) \propto \exp\left\{\frac{1}{2}\sum_{i=1}^{i=n}\sum_{j=1}^{j=n} \beta_{ij}(1 - x_i x_j)\right\}. \tag{5}$$

They use a variation of the Ford-Fulkerson algorithm [10] to solve the MINCUT problem. Next we explain the relation of MINCUT and MRF's. The MINCUT problem involves partitioning the nodes into two sets $\mathcal{V}_1 \subseteq \mathcal{V}$ and $\mathcal{V}_2 = \mathcal{V} \setminus \mathcal{V}_1$ such that the sum of edge weights with one endpoint in $\mathcal{V}_1$ and the other in $\mathcal{V}_2$ is minimized. More formally, define $\mathcal{C}_{12}$ as the set of edges, $ij \subset \mathcal{E}$, with one endpoint in $\mathcal{V}_1$ and the other in $\mathcal{V}_2$; The MINCUT problem is to find the cut minimizing the sum of edge weights in $\mathcal{C}_{12}$, i.e. $\min_{\mathcal{V}_1,\mathcal{V}_2} \textsc{cut}(\mathcal{V}_1, \mathcal{V}_2) = \sum_{ij\in\mathcal{C}_{12}} w_{ij}$. An algebraic formulation can be obtained by introducing cut vectors, $\mathbf{x} \in \{-1, 1\}^n$, with $x_i = 1$ if $i \in \mathcal{V}_1$, and $x_i = -1$ if $i \in \mathcal{V}_2$. The minimum cut can be obtained as

$$\min_{\mathbf{x}\in\{-1,1\}^n} \frac{1}{2}\sum_{i<j} w_{ij}(1 - x_i x_j). \tag{6}$$

Following Boykov [4] We propose using the Ford-Fulkerson algorithm to solve the segmentation problem, the two labels being foreground and background. However we differ from him in that we learn the emission model of the MRF allowing for a much weaker labelling of the data than required by Boykov. In the next section the emission model and its estimation via EM is explained.

3 Formulation for EMCUT Algorithm

The problem now is to segment the image into two sets of labels, one set of pixels $\mathcal{F}$ corresponding to the foreground object, one to the background $\mathcal{B}$, given some partial labeling of the image. We propose to make a MAP estimate, in order to do this we must first define the posterior likelihood.

The observations $\mathbf{y}$ are the RGB values of each pixel in the image, thus instead of there being a single observation at each pixel there are three, which requires a slight adjustment of notation. Let $\mathbf{y}_i$ denote the observation at pixel i, which is a 3×1 vector representing RGB; and $\mathbf{y} = (\mathbf{y}_1 \dots \mathbf{y}_n)$ is the collected set of observations for the whole image.

First we define the emission model or likelihood for each of the two labels foreground and background $x_i \in \{F, B\}$. These are each assumed to be a Gaussian mixture with different parameters, Θ^F for the foreground and Θ^B for the background with $\Theta = \{\Theta^F, \Theta^B\}$. Let the foreground Gaussian parameters by defined as a mixture of m Gaussians with means $\boldsymbol{\mu}_j^F$ and covariances $\boldsymbol{\Gamma}_j^F$, $j = 1 \dots m$, and a m dimensional vector of mixing parameters $\boldsymbol{\eta}^F$ with $\sum \eta_j^F = 1$ e.g.

$$\Theta^F = (\boldsymbol{\mu}_1^F, \dots, \boldsymbol{\mu}_m^F, \boldsymbol{\Gamma}_1^F, \dots \boldsymbol{\Gamma}_m^F, \boldsymbol{\eta}^F). \tag{7}$$

If a pixel i is foregound then the likelihood of the observation $\mathbf{y}_i$ is

$$p(\mathbf{y}_i|\Theta_F, x_i = F) = \sum_{j=1}^{j=m} \eta_j^F \left(\frac{1}{(2\pi)^{3/2} |\boldsymbol{\Gamma}_m^F|^{1/2}} \right) \exp\left(-\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_j^F)^\top \boldsymbol{\Gamma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_j^F) \right). \tag{8}$$

The set of parameters Θ^B and likelihood for the background are similarly defined,

$$p(\mathbf{y}_i|\Theta_B, x_i = B) = \sum_{j=1}^{j=m} \eta_j^B \left(\frac{1}{(2\pi)^{3/2} |\boldsymbol{\Gamma}_m^B|^{1/2}} \right) \exp\left(-\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_j^B)^\top \boldsymbol{\Gamma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_j^B) \right). \tag{9}$$

We are now in a position to specify the full posterior likelihood of the model $(\mathbf{x}, \Theta^F, \Theta^B)$,

$$p(\mathbf{x}, \Theta^F, \Theta^B|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}, \Theta^F, \Theta^B) p(\mathbf{x}, \Theta^F, \Theta^B) \tag{10}$$

it is reasonable to assume independence between the shape (labelling) and texture model of the foreground and background therefore

$$p(\mathbf{x}, \Theta^F, \Theta^B) = p(\mathbf{x}) p(\Theta^F) p(\Theta^B) \tag{11}$$

with a binary markov random field prior as given in (5) for $\mathbf{x}$ with $x_i \in \{F, B\} = \{-1, +1\}$, and some suitable prior for $(\Theta^F), (\Theta^B)$, Dirichilet on $\boldsymbol{\eta}$, **fix what should the others be *** The likelihood $p(\mathbf{y}|\mathbf{x}, \Theta^F, \Theta^B)$ decomposes as

$$p(\mathbf{y}|\mathbf{x}, \Theta^F, \Theta^B) = \prod p(\mathbf{y}_i|x_i, \Theta^F, \Theta^B) \tag{12}$$

which can be evaluated using (8) and (9).

Having defined the posterior likelihood the question is now how to maximize this over the parameters $(\mathbf{x}, \Theta^F, \Theta^B)$? One approach is to hold some of the parameters constant and maximize the posterior for the remainder if some suitable decomposition of the parameters can be found that makes this tractable. Given fixed (Θ^F, Θ^B) we can use the MINCUT method to maximize $\mathbf{x}$, given fixed $\mathbf{x}$ we can use EM to maximize (Θ^F, Θ^B). This amounts to a set of line searches in the space of $(\mathbf{x}, \Theta^F, \Theta^B)$ that can be shown to always increase the full posterior (10) and so should eventually converge at least to a local minimum.

So far we have only discussed the regional information, i.e. the log likelihood integrated over the foreground and background. However the boundary itself has a rôle distinct from the likelihoods we have discussed so far. For instance given to segmentations with equal likelihood (as defined above) we would still say that the one that most faithfully adheres to edges in the image is more likely. The next section considers how to encode this information in the posterior.

4 Integrating Regional and Boundary Information

The model in the previous section encodes region information as opposed to boundary information. We would like to introduce into the model the idea that the gradient along the boundary follows a distribution that is distinct from non boundary gradients within the image. This involves considering the joint distribution of neighbouring pixels. Next we establish a new model in which (12) no longer holds, rather the model introduces a new pairwise term to model the covariance between adjacent pixel colours $\mathbf{y}_i, \mathbf{y}_j$

$$p(\mathbf{y}|\mathbf{x}, \Theta) = \frac{1}{Z} \exp \left\{ \sum_{i=1}^{i=n} K_i(y_i, x_i, \Theta) + \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} K_{ij}(y_i, y_j, x_i, x_j, \Theta) \right\}, \quad (13)$$

where $K_i(y_i, x_i, \Theta)$ is given just by the mixture model defined for the likelihood given previously e.g. by the log of the functions given in (8) and (9)

$$K(\mathbf{y}_i, x_i, \Theta,) = \begin{cases} \sum_{j=1}^{j=m} \eta_j^F \left(\frac{1}{(2\pi)^{3/2} |\mathbf{\Gamma}_m^F|^{1/2}} \right) \exp \left(-\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_j^F)^\top \mathbf{\Gamma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_j^F) \right), & x_i = F \\ \sum_{j=1}^{j=m} \eta_j^B \left(\frac{1}{(2\pi)^{3/2} |\mathbf{\Gamma}_m^B|^{1/2}} \right) \exp \left(-\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_j^B)^\top \mathbf{\Gamma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_j^B) \right), & x_i = B \end{cases} \quad (14)$$

So the interesting case is what form $K_{ij}(y_i, y_j, x_i, x_j, \Theta)$ should take? One is to choose $K_{ij}(y_i, y_j, x_i, x_j, \Theta)$ so as to weaken the connectivity links between adjacent pixels with high contrast:

$$K_{ij}(y_i, y_j, x_i, x_j, \Theta) = \frac{1}{2} \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} (1 - x_i x_j) \exp - \left(\frac{||\mathbf{y}_i - \mathbf{y}_j||^2}{2\pi\sigma_e^2} \right), \quad (15)$$

where Θ is now augmented to include an extra parameter σ_e. This is just the Bayesian formulation of the likelihood used by Boykov [4]. Initially it is not clear how to optimize the new posterior arising by incorporating this likehood. This is because the partition function Z in (13) cannot be readily evaluated. Recall there are two parts to the optimization, one is to optimize $\mathbf{x}$ holding Θ constant, inspection reveals that the problem is equivalent to an inhomogeneous Markov Random Field with varying β_{ij} between the sites, this can be rapidly solved by graph cut methods. The second part is to optimize Θ holding $\mathbf{x}$ constant, unfortunately the unknown partition function, Z, means that the EM algorithm proposed in the previous section is no longer exact. The procedure we follow is first to estimate σ_e^2 as the variance of image gradients, $||\mathbf{y}_i - \mathbf{y}_j||$, over the image in question. We then fix σ_e^2. Then set all the

$$\beta_{ij} = \exp - \left(\frac{||\mathbf{y}_i - \mathbf{y}_j||^2}{2\pi\sigma_e^2} \right), \quad (16)$$

these are then fixed as well, the EMCUT algorithm is then run as described in Sections 2 and 3. The algorithm is summarized in Table 1. Of course the ideal would be to optimize σ_e^2, doing this exactly is problematic because of the problem of estimating the partition function, but it may be possible using fourier methods, or pseudo likelihoods and this is the subject of current research. Other choices for the β_{ij} are examined in the results section.

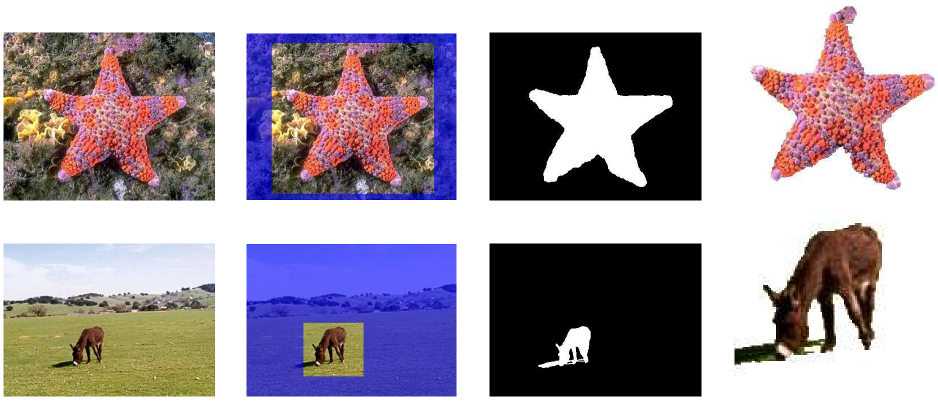
5 Results

Our approach is demonstrated on the 'starfish' and 'donkey' images shown in Figure 1 column (a). In each case the user drags and clicks a rectangle, shown in column (b), over the object of interest (a one click operation). As can be seen the results, column (d) are close to the ground truth, column (c), which was obtained by careful hand labelling of the pixels. The starfish example took seven iterations to converge (an iteration being one maximization

1. N is the number of pixels in the image.
2. Fix the $\beta_{ij} = \exp - \left(\frac{||\mathbf{y}_i - \mathbf{y}_j||^2}{2\pi\sigma_e^2} \right)$, for all ij.
3. Repeat an Iteration of EMCUT:
   (a) Optimize $p(\mathbf{y}|\mathbf{x}, \Theta^F, \Theta^B)$, using MINCUT, holding Θ^F, Θ^B constant.
   (b) Optimize $p(\mathbf{y}|\mathbf{x}, \Theta^F, \Theta^B)$, using EM, holding $\mathbf{x}$ constant.

Table 1: *A brief summary of all the stages of the* EMCUT *algorithm.*

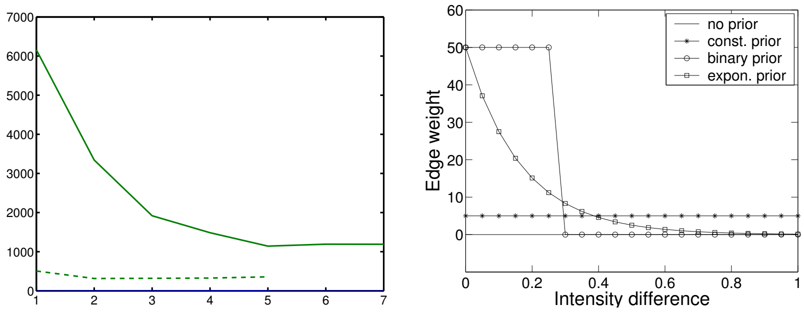
of $\mathbf{x}$ via graph cut followed by one maximization of Θ via EM) starting from a rectangle with 49159 mislabelled pixels. After the first iteration this was dramatically reduced to 6161 pixels, and after seven iterations convergence was achieved with 1190 mislabelled pixels. It should be noted that most of the mislabelled pixels are on the boundary, characterized by 'mixed pixels' where the ground truth supplied by hand is somewhat suspect. In each case it was found that three components for foreground and background distributions was used.



(a) (b) (c) (d)

Figure 1: *(a) The 'starfish' and 'donkey' image. (b) shows the rectangle selected by the user, drawn so as to loosely bound the object of interest. (c) The ground truth, generated by hand, there are 49159 mislabelled pixels in the starfish case, 6455 in the donkey. (d) The resulting cut out obtained after iterating to convergence, with 1190 and 357 mislabelled pixels respectively. In each case it was found that three components for foreground and background distributions was used.*

Next we explore the effects of using different functions to define β_{ij}. The frog example, figure 3 shows the effect of using the functions of β_{ij} shown in figure 2 (b). It can be seen that using no or even a homogenuous markov random field prior produce inferior results, as with $\beta_{ij} = 0$ the segmentation lacks spatial cohesion, but with β_{ij} set to a constant, the image is cohesive, but the segmentation does not adhere to image edges. A binary function of β_{ij} which shows much improved results, but some artefacts are introduced by making a hard decision. Over all the examples the best results were obtained by using a soft boundary exponential function.



(a) (b)

Figure 2: *(a) A graph of the number of mislabelled pixels, comparing the result after each iteration (one maximization of* $\mathbf{x}$ *followed by one of* Θ*), with the ground truth. Note that the for the sake of resolution the initial number of mislabelled pixels 49159 and 6455 are not shown on this graph. (b) Different functions for the* β_{ij} *used in Figure 3, shown against normalized intensity difference,* $\delta_{ij} = \frac{||\mathbf{y}_i - \mathbf{y}_j||}{256^3}$*. Solid line* $\beta_{ij} = 0$*, starred line* $\beta_{ij} = 5$*, circles line has* $\beta_{ij} = 0$*,* $\delta_{ij} < 0,3$*-*$\beta_{ij} = 50$ *otherwise. Finally the squared line sets* σ_e *as the variance of the edge gradients.*

The next example demonstrates the power of our interactive cut-out system. The head of Vincen VanGogh has been seperated nicley from his portrait with two user interactions (see fig. 4).

We represent a distribution of fore- and background measurements with a Gausian mixture model in colourspace. Obviously, more mixtures approximate a given distribution better. However, in order not to overfit the data, a limited number of mixtures should be choosen. Since our aim is too build an interactive system which short user interaction cycles, the number of mixtures should be small and we do not consider the issue of determine automatically a good number of mixtures. In pratice we choose 3 mixtures. However, a minimum number of Gaussian is necessary to approximate the (probably) muli-modal fore- and background distribution sufficently. The negative effect of too few mixtures is illustrated in Figure 5.

6 Conclusion

Within this paper we have presented a Bayesian formulation of the interactive image cut out problem. The emphasis has been on ease of use, and we aim for a one click system, although we have also demonstrated that if the initial result is unsatisfactory the result can be rapidly touched up. Some interesting future research areas are (1) to enrich the representation of the boundary distribution i.e. to go beyond just gradient information; (2) to improve the learning of the parameters of this distribution; (3) at present we do not consider alpha blending around the boundary, improved results would be obtained by modeling the process of image composition. Overall the system provides a fast and optimal way of interactive segmentation. The convergence is extremely fast and the algorithm is real time providing immediate feed back to the user.

References

[1] *Adobe PhotoShop Version 5.5 (Chapter 7: Seclecting).* Adobe Systems Inc., 1999.

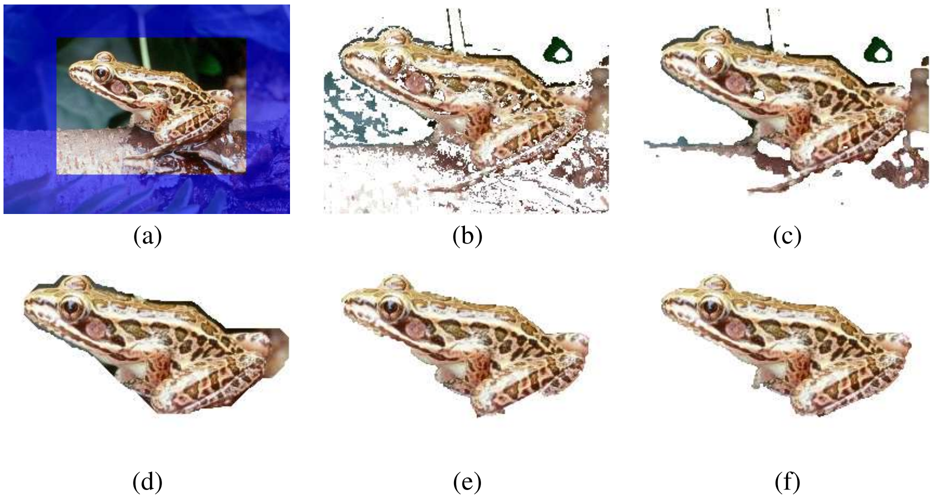


Figure 3: *Frog Image: (a) Initial region selected by the user. (b) No* MRF*, without which the segmentation lacks spatial cohesion, (c)* $\beta_{ij} = 5$ *is a constant, here the image is cohesive, but the boundary does not adhere to image edges. (d)* $\beta_{ij} = 50$ *here the cost function is dominated by minimizing the length of the boundary excessively rounding off the object. (e) A binary function of* β_{ij} *which shows much improved results, but some artefacts are introduced by making a hard decision. (f) Over all the examples the best results were obtained by using a soft boundary exponential function.*

[2] J. Besag. Spatial interaction and statistical analysis of lattice systems. *J. Roy. Stat. Soc. Lond. B.*, 36:192–225, 1974.

[3] J. Besag. On the statistical analysis of dirty pictures. *J. Roy. Stat. Soc. Lond. B.*, 48:259–302, 1986.

[4] Y.Y. Boykov and M.P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In *ICCV01*, pages I: 105–112, 2001.

[5] N. Cressie. *Statistics for Spatial Data*. Wiley, 1993.

[6] Daneels *et al*. Interactive outlining: An improved approach using active contours. storage and retrieval for image and video databases. In *SPIE93*, pages 226–233, 1993.

[7] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B.*, 51:271–279, 1989.

[8] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *Proc. 1st Int. Conf. on Computer Vision*, pages 259–268, 1987.

[9] E.N. Mortensen and W.A. Barrett. Interactive segmentation with intelligent scissors. *GMIP*, 60(5):349–384, September 1998.

[10] C. Papadimitriou. *Combinatorial Optimization*. Prentice Hall, 1998.

[11] P. Perez, A. Blake, and M. Gangnet. Jetstream: Probabilistic contour extraction with particles. In *ICCV01*, pages II: 524–531, 2001.

A Thought

we can turn this into a path planning problem if we only want one boundary.

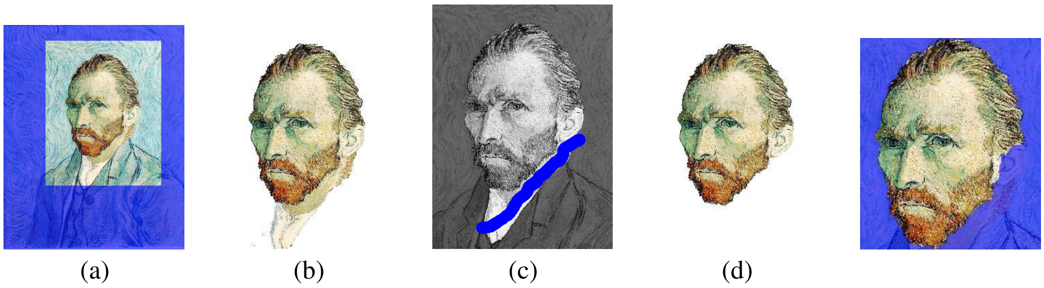


Figure 4: *The first user interaction is to specify the enlighted rectangle in (a). The result of our* EMCUT *method is illustrated in (b). The neck and parts of the shirt were segmented as foreground. The user edits this segmentation with a foreground brush (black line). The result of the subsequent* EMCUT *loop is shown in (d). A final edit removes the ear (e)*

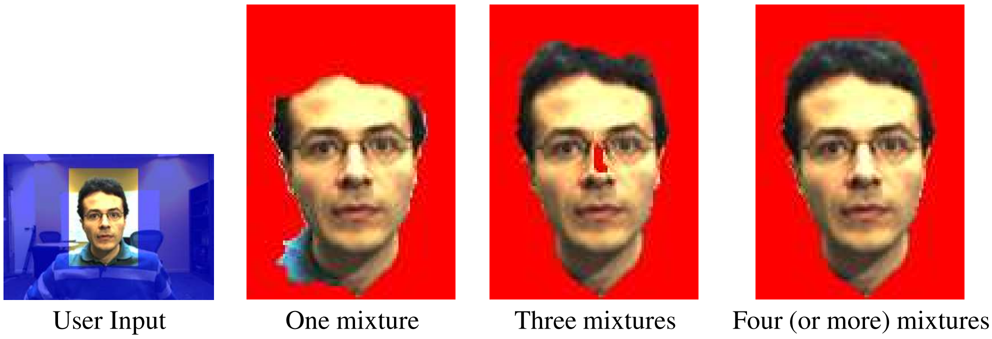


Figure 5: *One mixture for the fore- and background distribution was not enough to represent the complete color distribution of the head. For three mixtures, the nose (bright color) was labeled as background (also bright). In case of four mixtures the face was segmented succesufully.*

B Experimental results

In this section we presents results of our interactive EMCUT method.

B.1 The performance of the EMCUT method

The following examples illustrate the results of our system with only one user interaction. The segmentation after each iteration of our EMCUT method is illustrated. In general, convergence is achieved after a few iteration. In all examples it is visible that the amount of mislabeld data decreases in each iteration of the EMCUT. This means that the foreground and background distributions get better "seperated" in each iteration. Due to our boundary term, the changing of labeling effects more larger connected areas in contrast to individual pixles.

*** Choose pick 2 out of 4 examples ****

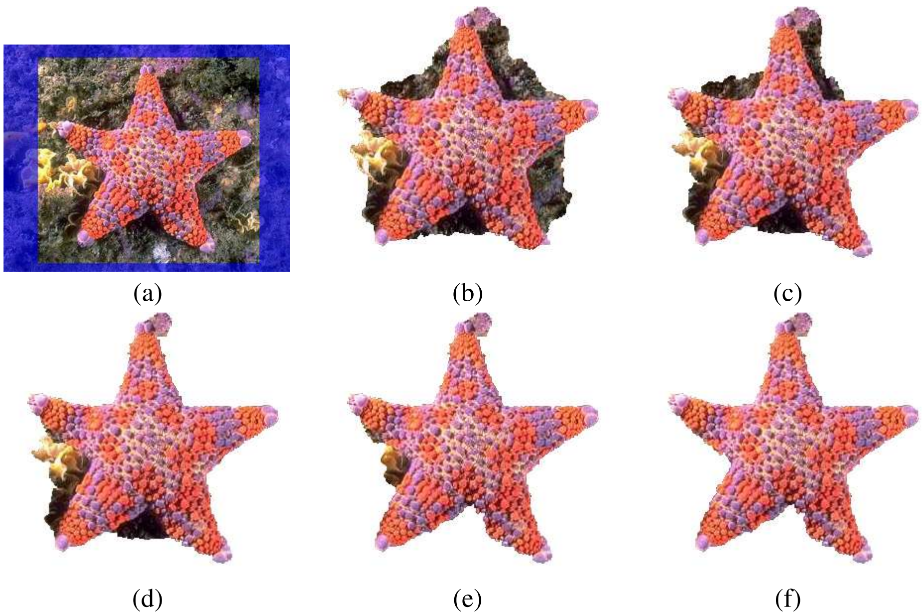


Figure 6: *The enlighted rectangle is the specified by the user (a). The five iterations, until convergnce are shown in (b-f)*

B.2 Example for User Interaction

B.3 Patches instead of pixels as measurement unit

In the following we will demonstrate that patches do not improve the results. **** I am not sure if we should put this in ****************

B.4 Prior function

********** Do more experiments here ***************** As motivated in chapter x, we would like to change the label of being foreground and background when there is an edge in the image. In the following, we demonstrate that using the different prior function has a big influence on the quality of the resulting segmentation.

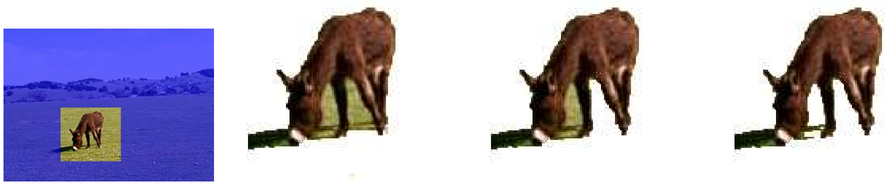
B.5 Size of Neighbourhood

In a last experiment we investigate the influence of the size of neighbourhood in the MRF. The main comlusion is that a four neighborhood produces “blocky” results. A eight nirghborhood works good and as well fast. A larger neighboorhood (larger than 8) does not considerably improve the “blockyness” of the contour. Furthermore, a larger neighboorhood smoothes the segmentation considerably. This has the negative effect that smaller areas (correctly labeled) disappear.

Consequently, the following function has

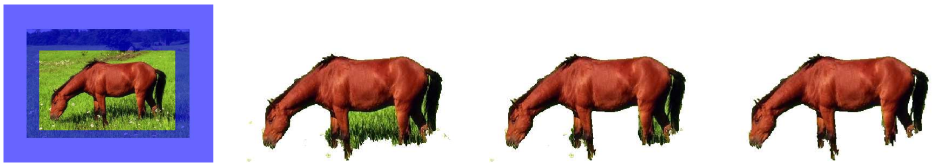
(a) (b) (c)

Figure 7: *The enlighted rectangle is the specified by the user (a). The two iterations, until convergnce are shown in (b-c)*



(a) (b) (c) (d)

Figure 8: *The enlighted rectangle is the specified by the user (a). The three iterations, until convergnce are shown in (b-d)*



(a) (b) (c) (d)

Figure 9: *The enlighted rectangle is the specified by the user (a). The three iterations, until convergnce are shown in (b-d)*

4-neigh. 8-neigh. 212-neigh.

Figure 10: *The "blockyness" of a four-neighborhood is visible at the boundary of the hair. This improves with an eight-neighborhood. A larger neighboorhood (here 212) does not improve the contour considerably.*