Modeling vs. Segmenting Images Using A Probabilistic Approach

Datong Chen
Computer Science Department, Carnegie Mellon University
email: datong@cs.cmu.edu

Abstract

Image segmentation is conventionally formulated as a pixel-labeling problem, in which "hard" decisions have to be made to partition pixels into regions. As image segmentation is usually used as a preprocessing step in many image analysis applications, the segmentation errors introduced by the "hard" decisions bring difficulties to higher-level image analysis. In this paper, we propose a "soft" image segmentation method to model the object appearance and spatial layouts in an image with an incremental mixture of probabilistic models. The proposed approach extracts "soft" regions incrementally using adaptive apertures without making any hard decisions. We show that "soft" regions not only bring more robustness than conventional "hard" regions but also enable a higher-level region-based analysis.

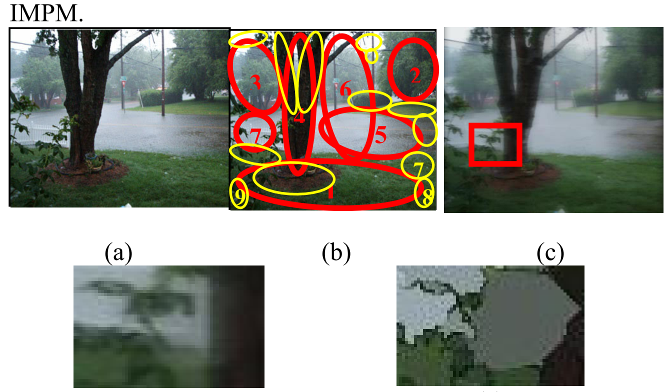
Index Terms— s*oft image segmentation, incremental mixture of probabilistic models*.

1. INTRODUCTION

Image segmentation, a preprocessing vision technique for region-based image analysis, produces regions or edges that can be used for higher-level analysis, e.g., shape extraction, object detection, and object recognition. Segmentation is usually formulated as a pixel-labeling problem, in which hard decisions are made according to some predefined criteria to assign each pixel a class label. A drawback of this formulation is that hard decisions must be made at the beginning of the image analysis. Research shows that perfect segmentation for general higher-level analysis tasks is not usually possible without having knowledge of the target shape or object at the higher-level [4]. Segmentation errors, either merging regions of different objects together or splitting an object into too many regions, cannot be easily corrected by algorithms in higher-level analysis. In this paper, we propose a "soft" segmentation method to avoid making hard decisions in image segmentation.

Prewer and Kitchen presented a non-probabilistic soft image segmentation method [4]. In their work, a weighted linked pyramid is employed to represent the non-unique segmentations of an image instead of making one hard segmentation. Soft image representation can be achieved more smoothly by probabilistic modeling. R. Wilson [3] proposed Multi-resolution Gaussian Mixture Models (MGMM) on the basis of scale space theory. MGMM use a GMMs tree to represent an image in multiple resolutions. Leaf models can be considered as a regression of the image, which is also a soft representation of the image. However, the MGMM method is proposed as an image representation algorithm instead of a segmentation algorithm. The algorithm focuses on fitting the image properly. Apertures are not adaptive to the local areas for better segmentation. Recently, many researchers have treated image segmentation as a mid-level or high-level image analysis process. In these works, an image is modeled as a graph. Segmentation is performed by graph cut algorithms, such the n-cut algorithm [5] and the spectral rounding algorithm [6], to integrate objects' information into the segmentation. Although image segmentation in these works is addressed as a higher-level algorithm, necessary low-level segmentation processes, such as region extraction, are also implicitly employed in the graph cut algorithms.

To address the soft segmentation problem, we use an incremental mixture of probabilistic models (IMPM) to model the appearance and spatial layouts of objects in an image from coarse apertures to fine apertures. Instead of assigning each pixel a class label, the IMPM estimates the probabilities of each pixel belonging to models, and produces "soft" regions in the form of probabilistic distributions. Fig. 1(b) demonstrates the spatial centers and variances of some models learned from the image (Fig. 1(a)) using an unsupervised algorithm. Red ellipses indicate the models extracted in a coarse aperture, which correspond well to the spatial layout of the objects in the image. Yellow ellipses are models in a finer aperture for objects' details. More image details can be incrementally modeled at smaller and smaller apertures, which are not displayed in the figure. The resulting "soft" regions have comparable capabilities to conventional "hard" regions supporting region-based image analysis. Fig 1 (c) illustrates the image reconstructed by using "soft" regions (models), in which each pixel is rendered as the expectation value in the IMPM.



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

Fig. 1 A soft image segmentation using the IMPM method. Fig. 1(d) zooms in the red rectangle region in Fig. 1(c). In comparison, Fig. 1(e) shows the hard segmentation result of the same region using the ML method. The "soft" regions retain more information on detail structures of the objects than the hard segmentation result. The IMPM algorithm has many advantages in a mid-level image analysis.

2. INCREMENTAL MIXTURE OF PROBABILISTIC MODELS

Intuitively, an IMPM method estimates the layout of the objects in an image incrementally with multiple models from coarse apertures to fine apertures. As illustrated in Fig. 1(b), models 1-6, estimated from the entire image (the largest aperture), represent the layout of large objects, such as sky, grassland, and trees. The key contribution of the IMPM is that more detail structures of the objects can be represented by incrementally estimating more and more new models within smaller and smaller apertures without destroying the structures learned at the coarser levels. Therefore, we can transform the "hard" image segmentation task into a "soft" coarse-to-fine modeling process.

Formally, let us define a color image $I = \{v_{xy}\}$ by a set of pixels v_{xy}, where each pixel is denoted as a feature vector $v_{xy} = (f, x, y)$ with feature vector f (color or texture features) and a spatial location. We model the image I using a set of ordered models $G = \{G_i\}, G_i = (\alpha_i, M_i, C_i)$. The likelihood of the image I over the mixture models G is computed as:

$$P(I \mid G) = \sum_{G_i \in G} \alpha_i P(I \mid M_i, C_i), \tag{1}$$

where

$$P(I \mid M_i, C_i) = \prod_{V_{xy} \in C_i} P(v_x \mid M_i). \tag{2}$$

Each probabilistic model M_i can be a predefined object shape or just an unsupervised learned Gaussian model. An aperture constraint C_i specifies a subset of pixels in the image, from which the model M_i is computed. The algorithm of the IMPM method is described in the following pseudo code.

Algorithm of the proposed IMPM method

1. Initialize the aperture set C with the entire image C_0. Initialize the model set G with predefined models $G = \left\{G_1, ..., G_{k_0}\right\}$ if available.
2. For each aperture C_i in C, incrementally estimate new models M_i and weights α_i with the existing model set G.
3. Add new models in to the model set G.
4. Compute a new set of apertures C using the model set G.
5. Add newly estimated models from every aperture in C into the model set G.
6. Loop the step 2 to 5, until the size of all new apertures in C are smaller than a stop threshold.

1.1. Initialize the model set

Initial IMPM models are learned from the largest aperture constrained as the entire image. This learning process can be totally unsupervised as in the example in Fig. 1. It can also be initialized by predefined or manually specified shapes. To combine these two cases together, we assume that the model set is initialized by a set of predefined models $\left\{G_1, ..., G_{k_0}\right\}$, where $k_0 = 0$ if no models are predefined.

1.2. Incremental model estimation

The estimation of models given an aperture constraint C_i and the existing model set G can be implemented as a semi-supervised clustering problem. Our goal is to find k models $\left\{G_{k_0+1}, ..., G_{k_0+k}\right\}$ together with the existing models $\left\{G_1, ..., G_{k_0}\right\}$ that maximize the likelihood of the image I:

$$\max_{\{G_{k_0+1}, ..., G_{k_0+k}\}} P(f \mid G_1, ..., G_{k_0+k}), \tag{3}$$

where k is an unknown number. The optimal parameters can be estimated using a semi-supervised Expectation-Maximization (EM) algorithm [1]. The value of k can be determined by using the Bayesian Information Criterion. However, the EM algorithm, which is a time consuming algorithm, must be performed many times. In practice, we propose a semi-supervised spatial co-EM algorithm to perform EM only once.

In the semi-supervised spatial co-EM algorithm, the input space of pixels v_{xy} is split into two spaces: the feature space (f), which can be any color or texture features, and the spatial space (x,y). For each model $G_i = (\alpha_i, M_i, C_i)$, the statistical model M_i is also split into two models $M_i^{'}$ corresponding to the feature space and $M_i^{''}$ corresponding to the spatial space:

$$P(\cdot \mid M_i) = P(\cdot \mid M_i^{'}) P(\cdot \mid M_i^{''}). \tag{4}$$

The co-EM process uses the expectation of the feature model as hidden data to update spatial model while using the expectation of the spatial model to update the feature model. Notice that the proposed spatial constraint co-EM algorithm does not satisfy the sufficient requirements of conventional co-EM algorithms [2]. The spatial space does not contain enough information leading to a good clustering by itself. In fact, the clusters resulting from the expectation of the feature models may exist in separated regions in the spatial space. Fortunately, in our case, it is not reasonable to force the color space and the spatial space to converge to the same clusters. Therefore, we allow the two spaces to have different number of clusters in the co-EM process, and employ a split-merge procedure to connect them together. The algorithm is implemented as the following:

E-step:

1. Compute the temporary feature expectation of each pixel in the aperture using the current $k_0 + k_c$ color models $M_i^{'}(i = 1, ..., k_0 + k_c)$;

   Split: Label the pixels in the aperture as their feature expectations. Then find connected components that are sufficiently large in the expectation image. The number of the resulting regions is $k_0 + k_r^{t+1}$. We ensure that all existing models share the same cluster labels between the color space and the spatial space.

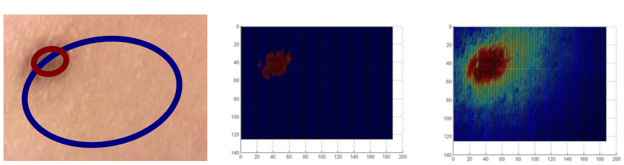
   Using the region label at each pixel as its current spatial expectation, record the map from spatial clusters to color clusters $A^{t+1} : k_r^{t+1} \to k_c$.
2. Compute the temporary spatial expectation of each pixel in the aperture using the current $k_0 + k_r^t$ spatial models $M_i^{"}(i = 1, ..., k_0 + k_r^t)$;

   Merge: convert the temporary spatial expectation of each pixel to feature expectations using the map A^t.

M-Step:

1. Update feature models $M_i^{'}(i = k_0 + 1, ..., k_r^t)$ using the feature expectations.
2. Compute the new spatial models $M_i^{"}(i = k_0 + 1, ..., k_r^t)$ using the spatial expectations.

k_c is optimized by using Bayesian Information Criterion (BIC). The weights α_i of each model is proportional to area size of its aperture.



(a) (b) (c)

Fig. 2 Illustration of estimated models, expectation image and a model pdf of a skin cancer image. (a) 2 estimated IMPM models. (b) expectation image. (c) pdf of the small red model.

This algorithm can achieve a reasonable estimation in no more than 10 iterations. Fig. 2 shows an example of estimated models, expectation image and model pdf of a skin cancer image. In the image, there are no clear boundary between cancer and normal skin. Using the model pdf, we can avoid make hard segmentation decisions on the boundary of the cancer regions.

The number of resulting initial models $k_0 + k$ is adaptively determined by the eventual spatial models. The predefined models are treated as a set of "existing" models, which are not updated in the algorithm. If no model is predefined, the algorithm is completely unsupervised, which still performs "soft" segmentation well.

1.3. Estimating finer apertures

Finer apertures are adaptively extracted using the existing model set *G*. Given the existing models $\{G_1, ..., G_n\}$, the centers of the new apertures are located around the pixels that are close to the boundary of the existing models. The candidate pixels of these new aperture centers can be extracted using the following model-crossing algorithm.

We first compute expectation image *E*, in which each pixel of the image *I* is labeled using the model that gives the highest probability:

$$E(v_{xy}) = G_m = \arg\max_{i=1}^{n} P(v_{xy} \mid G_i) \cdot \quad (5)$$

The pixels in the expectation image *E* that cross at least two models are extracted as candidate pixels. We then randomly sample 5% of the candidate pixels as the centers of the new apertures. Each new aperture has a square shape (only the aperture of an initial model may have a rectangle shape). If the center of a new aperture is located at pixel v_{xy}, the length of each side of the aperture is determined by the average distance from the pixel v_{xy} to the centers of apertures of all the models it crossed. A threshold can be set on the size of newly generated apertures to stop the expansion of the model set *G* if pixel level precision is not necessary.

3. EXTRACTING "SOFT" REGIONS

The IMPM method enables regions or shapes to be extracted in a "soft" way. A resulting "soft" region is a probability distribution (a pdf) over the entire image, which models the probability of each pixel belonging to the region.

Each IMPM model can be associated with a region. Let us denote a region as R_{G_i} associated with a model G_i. If the goal is modeling the image, simply computing the region pdf $P(v_{xy} \mid R_{G_i})$ as the model pdf $P(v_{xy} \mid G_i)$ is good enough. If the goal is segmentation, the model pdf does not provide reasonable probabilities for the pixels that are far away from the model. To address this problem, we first compute the expectation image *E* using the model set *G*. Then for each pixel v_{xy} with $E(v_{xy}) = G_j$, we define its probability $P(v_{xy} \mid R_{G_i})$ as:

$$P(v_{xy} \mid R_{G_i}) = \alpha_{R_{G_i}} P(v_{xy} \mid G_j) P(G_j \mid G_i), \quad (6)$$

where the model merging probability $P(G_j \mid G_i)$ is computed as $P(G_j \mid G_i) = 1$ if $G_j = G_i$; Otherwise, let S_j

denote the set of pixels that are labeled as G_j in the expectation image *E*. The merging probability is defined as:

$$P(G_j \mid G_i) = \frac{1}{|S_j|} \sum_{v_{xy} \in S_j} P(v_{xy} \mid G_i) \cdot \tag{7}$$

The normalizer $\alpha_{R_{G_i}}$ can be obtained by:

$$\alpha_{R_{G_i}} = \frac{1}{\sum_{v_{xy} \in f} P(v_{xy} \mid G_j) P(G_j \mid G_i)} \cdot \tag{8}$$

1.4. "Soft" region properties

"Soft" regions extracted by IMPM can still be manipulated as "hard" regions in a probabilistic way, for example region rendering, areas, orientations, bounding box and so one. In Fig. 1 (c), we render each pixel by the mixture of the mean colors of the extracted "soft" regions:

$$color(v_{xy}) = \frac{\sum_{i=1}^{n} P(v_{xy} \mid R_{G_i}) RGB(R_{G_i})}{\sum_{i=1}^{n} P(v_{xy} \mid R_{G_i})}, \tag{9}$$

where the mean color of each region $RGB(R_{G_i})$ is computed using the region model $P(\cdot \mid R_{G_i})$ and the RGB value of every pixel $I(xy)$ in the image *I*:

$$RGB(R_{G_i}) = \frac{\sum_{v_{xy} \in f} P(v_{xy} \mid R_{G_i}) I(xy)}{\sum_{v_{xy} \in f} P(v_{xy} \mid R_{G_i})} \cdot \tag{10}$$

The area of the region R_{G_i} as used in Eq. (10) can be calculated using the mean value:

$$area(R_{G_i}) = \sum_{v_{xy} \in f} P(v_{xy} \mid R_{G_i}) \cdot \tag{11}$$

Region orientation can be easily defined similarly. Some properties, for example the bounding box, can be plotted in varying probabilities. We will not go into the details of these definitions in this paper.

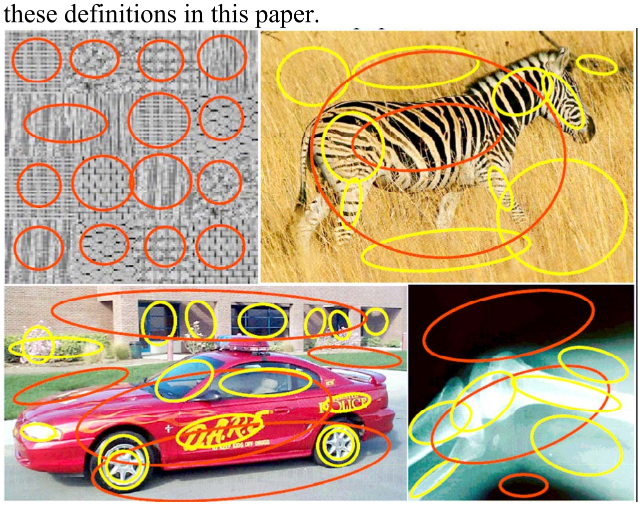
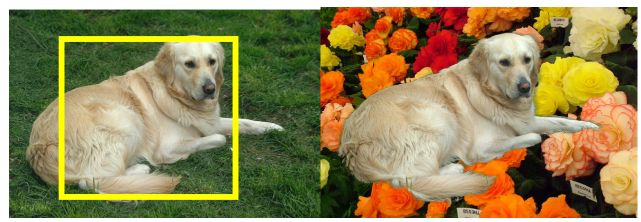


Fig. 2 IMPM models estimated at large apertures.

4. EXPERIMENTS

More experiments were performed on images with different shapes and texture. We employ features (*f*) such as grayscale values, color values, or Gabor features. The number of feature cluster k_c is optimized by using BIC. Fig. 2 shows some the IMPM models estimated at large apertures. We can see that the estimated models can roughly depict the object layout in the images.

Fig 3 illustrates the extraction of a "soft" region using the IMPM algorithm. We extract the largest "soft" region in the yellow rectangle (the dog) and mat the region pdf to a different background.



(a) (b)

Fig. 3 A "soft" region extracted using the IMPM method.

5. CONCLUSIONS

In this paper, we propose a "soft" image segmentation method based on a probabilistic approach to avoid making hard decisions in the low-level image segmentation process. We have presented the IMPM algorithm to incrementally estimate the appearances and the layouts of the objects in images using adaptive apertures. We have shown that the resulting "soft" regions contain comparable capabilities and more robustness to conventional "hard" regions in supporting higher-level region-based image analysis. We can conclude that modeling images using probabilistic approaches has very promising benefits in comparison to conventional image segmentation.

REFERENCES

[1] A. Dong and B. Bhanu, "A New Semi-Supervised EM Algorithm for Image Retrieval", CVPR 2003.

[2] K. Nigam and R. Ghani, "Analyzing the effectiveness and applicability of cotraining", Workshop on Information and Knowledge Management, 2000.

[3] R. Wilson, "MGMM: Multiresolution Gaussian Mixture Models for Computer Vision", ICPR 2000.

[4] D. Prewer and L. Kitchen, "Soft image segmentation by weighted linked pyramid", PRL 2001.

[5] J. Shi and J. Malik, "Normalized cuts and image segmentation", PAMI, 2000.

[6] D. A. Tolliver and G. L. Miller, "Graph partitioning by spectral rounding: applications in image segmentation and clustering", CVPR 2006.