An Adaptive Segmentation Method Using MRF Model for Suburban Aerial Images

Mansuo Zhao, David Tien

Abstract — An adaptive image segmentation method using Markov Random Field Model for suburban aerial images is presented in this paper. The image is modeled as a collection of regions characterized by slowly moving averages and standard deviation. Decreasing sized windows are used to calculate the moving averages during the iteration process. A function based weighting parameter between the two components in the energy function is also used to improve the performance of unsupervised segmentation. A hierarchical implementation scheme is also introduced to reduce the computation load and increase the segmentation speed.

Index Terms — Arial Suburban Images, Hierarchical Implementation, Markov Random Field Model (MRF), Segmentation.

I. INTRODUCTION

Aerial suburban images have been used increasingly in many areas in recent years, such as land-usage monitoring and suburban development controlling. How to extract manageable information of the objects that users are interested from the images is a challenging problem.

There are usually two kinds of approaches for this problem. One is called a top-down approach which is done by extracting objects with pre-defined features from image. The pre-defined features are either selected or determined by user or from a feature database consisting of the models of targeted objects. Many current commercial software are using this kind of approach. The other is called a bottom-up approach. In this kind of approach the image is segmented into areas which represent different objects. This is a much harder approach in its nature because of the unknown features of the objects. There are not many successful software known in today’s market. In this paper we are trying to find a good method which gives a satisfying segmentation for suburban images.

Many different kinds of segmentation methods have been presented in literatures, such as multilevel thresholding [12], class clustering [12], edge based segmentation [12], region growing methods [4], etc. Markov Random Field Model has been widely applied in image segmentation as well [5]-[7]. In this paper we proposed a method based on MRF Model with a moving average and hierarchical implementation.

MRF model was first introduced into image processing by German and German [8]. It has attracted much attention for its powerful performance in many applications, such as edge detection [9], image restoration [10], stereovision, long range motion and image classification [11]. Image segmentation using MRF model is posed as an optimization process measured by a cost function. The goal is to minimize the cost function. The advantage of using MRF model in image segmentation is that not only different kinds of image features can be used in the cost function, but also the spatial relationship is integrated into the function as well.

The cost function in MRF model consists of two components: a feature modeling component and a region labeling component which are joined by a constant weighting parameter. One constrains the region intensity to be close to the image data and the other imposes spatial continuity so that pixel with the neighboring pixels in this region is more likely to belong to the same region. This could avoid fractural segments in the result. Eight connected Gibbs random field model is used for the spatial constraints in the literature and achieves good results for the testing images. The four connected is also tested in our experiment and the results are similar to eight connected.

The parameters in the two components need to be estimated from the training data in order to have a stable segmentation result. However, this model is not able to work consistently without supervision. A variable weighting parameter was introduced to achieve a balance between the spatial information and feature information during its optimization process [12]. The introduction of a weighting parameter is initialized by the observation that the simple MRF based segmentation model is easily trapped in local maxima due to the imposed spatial constraint. A detailed analysis of the relationship between the two terms in the MRF model can be found in [16].

The image is modeled as a collection of regions with features having a distribution of a Gaussian function. Uniformed feature mean values for each class are used throughout the whole image in traditional MRF model which works well with images consisting of uniformed regions. It does not have good performance with images consisting of smooth surfaced objects. Moving feature averages instead of constant feature averages are used in our method for suburban aerial images because of the varying illumination in the images. The advantage of using moving averages is region class averages can adjust to local image features without increasing the number of feature classes.

Mansuo Zhao was with Charles Sturt University, Australia. (e-mail: mansuo@gmail.com)

David Tien is with Charles Sturt University, Australia. (e-mail: dtien@csu.edu.au)

The work has been supported by the Cooperative Research Centre for Spatial Information, whose activities are funded by the Australian Commonwealth's Cooperative Research Centres Programme.

Implementation of a MRF model is usually computation intensive. A fast simulated annealing scheme is commonly used to speed up the optimization process. Its shortcoming is that the convergence may be trapped in the local optima. In this paper we use a hierarchical structure to reduce the computation by working with the generated image pyramid. An image pyramid is first constructed by convolving the image with a low-pass filter and down-sampling. The segmentation starts from the coarsest image with an initial segmentation label obtained from K-Mean clustering and then moves to the next level when the pre-defined minimum window size is reached.

The detailed discussion of the model is presented in Section II. The experiment results are shown in Section III and conclusion and discussion can be found in Section IV.

II. PROPOSED METHOD

A. Simple Markov Random Field Model

First, let's review a few concepts.

- **Definition of neighborhood system and clique**
  Let $S$ be a set of lattice points, $s$ be a lattice point which belongs to $S$, $s\in S$, $f_s$ be the value of $F$ at $s$ and $\partial_s$ be the neighboring points of $s$, $\partial_s$ is a neighborhood system which is symmetric $r\in\partial_s \iff s\in\partial_r$ and $s\notin\partial_r$.

  A clique is a set of points which are neighbors of each other as shown in Fig. 1.

![Fig. 1. 8-connected cliques](image)

- **Markov Random Field**
  A random field $X = \{x_1, x_2, \ldots, x_s\}$ on the lattice $S$ with neighborhood system $\partial_x$ is defined as a Markov Random Field if for all $s\in S$ if
  \[
P(x_s | x_r, r \neq s) = P(x_s | x_{\partial_s}) .
  \]

- **Gibson distribution**
  A Gibson distribution relative to a neighborhood system is defined as
  \[
p(x) = \frac{1}{Z} \exp\{-\sum_{c\in C} V_c(x_s)\} ,
  \]
  where $C$ is the set of all cliques, $Z$ is the normalizing constant for the density, $V_c(x_s)$ is a potential function for $x_s$.
  \[
  V_c(x_s) = \begin{cases}
  -\beta, & \text{if } x_s = x_q \text{ and } s, q \in C \\
  +\beta, & \text{if } x_s \neq x_q \text{ and } s, q \in C.
  \end{cases}
  \]
  (2)

  The parameter $\beta$ is a positive number so that two neighboring pixels are more likely to belong to the same region than to different regions. The bigger $\beta$ is, the larger a region could be and the smoother the region border.

  The segmentation of an image can be modeled as below using a MRF model.

  Given a observed image data $F = \{f_s, s\in S\}$, where $f_s$ is image feature value at pixel $s$, we need to find the labeling $\hat{Y}$ which maximizes $P(Y | F)$. By Bayes’ theorem
  \[
P(Y | F) \propto P(F | Y) \cdot P(Y),
  \]
  where $P(Y)$ is a priori density of the region labeling process and $P(F | Y)$ is the conditional density of the observed image given the distribution of regions.

  According to the Hammersley-Clifford theorem the density of $Y$ is given by a Gibbs density function which is
  \[
P(Y) = \frac{1}{Z} \exp\{-\sum_{c\in C} V_c(Y_s)\} .
  \]
  (4)

  Assuming all the $K$ feature components $f^1, f^2, \ldots, f^K$ are independent to each other, that derives
  \[
P(F | Y) = \prod_{k=1}^K P(f^K | Y = m).
  \]

  The distribution of all feature data is assumed as a Gaussian function with different means $\mu^k_m$ and standard deviation $\sigma^k_m$.

  Then
  \[
P(f^k_s | Y = m) = \frac{1}{\sqrt{2\pi (\sigma^k_m)^2}} \exp\left\{-\frac{(f^k_s - \mu^k_m)^2}{2(\sigma^k_m)^2}\right\}
  \]
  (5)

  where $\mu^k_m$ and $\sigma^k_m$ are the mean and standard deviation for the $m$ th class in the $k$ th feature component.

  The two energy components in a MRF model are label distribution energy
  \[
  E_L(x) = \sum_{c\in C} V_c(x_s)
  \]
  (6)

  and feature modeling energy
  \[
  E_F = \sum_{s, m=1}^{K} \sum_{k=1}^{K} \left(\frac{(f^k_s - \mu^k_m)^2}{2(\sigma^k_m)^2} + \log(\sqrt{2\pi \sigma^k_m})\right)
  \]
  \[
  \sum_{s, m=1}^{K} \sum_{k=1}^{K} \left(\frac{(f^k_s - \mu^k_m)^2}{2(\sigma^k_m)^2} + \log(\sqrt{2\pi \sigma^k_m})\right)
  \]
  (7)

  The total energy is defined as
  \[
  E = E_L + \alpha E_F
  \]
  (8)
where $\alpha$ is a weighting parameter to determine how much $E_R$ and $E_F$ contribute to the entire energy $E$, individually.

Thus, the problem to maximize $P(Y \mid F)$ becomes one to minimize the energy $E$.

B. Varied Weight in The Energy Function

The weighting parameter $\alpha$ plays an important role in the energy function and can have a big impact on the segmentation result. If $\alpha$ is set too small, which makes the region labeling component dominant, we will have a segmentation with large regions but not complying well with the image feature data. On the other hand, if it is set too large the spatial information may get ignored and results in many fragmented regions. A good balance can be achieved if we set the value carefully by supervised data training. This method is not able to work consistently without human supervision and data training. To achieve a better performance under the unsupervised circumstances a new implementation scheme is proposed by H. Deng and D. Clausi [12] for unsupervised image segmentation. It introduced a function-based weighting parameter between the two components in a MRF model. This method can produce good segmentation results without parameter training in traditional MRF model. The function for the weighting parameter is formulated as below.

$$\alpha(t) = c_1 0.9^t + c_2$$

where $\alpha(t)$ is a time varying weighting parameter for the two energy components. $c_1$ and $c_2$ are constants given by user. Usually $c_1 = 80$ and $c_2 = 1/K$ (where $K$ is the number of features) give a good result. And $t$ is the run time in the simulated annealing process.

At the start of the process $\alpha(t)$ is larger which means feature modeling component is dominant. The pixels with similar feature values are more likely to belong to the same class. As the procedure processes $\alpha(t)$ gets smaller which means the spatial information out-weights the feature information in the energy function and the pixels neighboring with each other are more likely to belong to one class.

C. Moving Averages And Multigrid Computation

In 1992 T. Pappas proposed Adaptive Clustering Algorithm [14] for segmenting the images with smooth surfaced objects. It can be seen as a generalization of K-mean clustering algorithm to include spatial constraints and to account for local intensity variation. Thus, it avoids the problem of over segmentation in regions with varied illumination. In this method, each region is assumed to have a slowly varying intensity with noise and characterized by an intensity function. The intensity function can be calculated by averaging over a size decreasing sliding window.

In suburban aerial images we found that a moving feature average describes the region feature better than a constant feature class average because of the unevenness of the illumination. So a moving feature class average $\mu_m^*(x)$ is used in the feature modeling component.

Multi-grid interpolation is used to calculate the moving averages because the computation is enormous if the moving averages are calculated pixel by pixel, especially when window size is large. The space between grids is set to be half of the window size. Fig. 2 shows the grid points in a window. We calculate the estimates $\mu_m^*(x)$ at each grid point. Then the values of the rest pixels can be obtained by bilinear interpolation shown in (10).

$$\begin{align*}
\mu_{x,y} & \approx \frac{\mu_{x_1,y_1}}{(x_2 - x_1)(y_2 - y_1)}(x_2 - x)(y_2 - y) \\
& + \frac{\mu_{x_2,y_1}}{(x_2 - x_1)(y_2 - y_1)}(x - x_1)(y_2 - y) \\
& + \frac{\mu_{x_1,y_2}}{(x_2 - x_1)(y_2 - y_1)}(x_2 - x)(y - y_1) \\
& + \frac{\mu_{x_2,y_2}}{(x_2 - x_1)(y_2 - y_1)}(x - x_1)(y - y_1) \\
\end{align*}$$

where $\mu_{x,y}$ is the average value at point $(x, y)$, $\mu_{x_1,y_1}$, $\mu_{x_1,y_2}$, $\mu_{x_2,y_1}$ and $\mu_{x_2,y_2}$ are the average values at its four surrounding grid points.

During the iteration process the window size is decreasing. The reason behind this is the segmentation is crude at the early stage of segmentation and a large window produces more robust estimate of the moving averages. As the algorithm progresses a smaller window helps to give a more accurate and reliable estimation.

Fig. 2. Window for calculating moving averages and the grid points position.
To avoid segmentation small regions, a region size threshold \( T_{\text{min}} \) is introduced. If the number of pixels that belong to class \( i \) in the processed window centered at \((x, y)\) is less than the threshold, the pixel \( P(x, y) \) will not be assigned to the class \( i \). In our method \( T_{\text{min}} \) is a changing throughout the hierarchical process comparing to a constant in ACA [14]. \( T_{\text{min}} \) is small in the coarse level image and increases when the resolution of the image increases. In our experiment \( T_{\text{min}}(l) = 0.3(2W_l + 1)^3 \). And \( l \) is the level of the pyramid image being processed, \( W_l \) is the minimum window size in different pyramid image level. This is to help control the smoothness of the boundary. The smaller the value is, the smoother the boundary. However, if the minimum window size is set too small it not only increases the iteration time but also can result in over-segmentation. So we need to consider a trade-off between the smoothness and robustness. \( W_l \) is 1, 3, 5 and 7 are used in our experiment for the level 1, 2, 3, 4, respectively. Level 4 is the original image.

D. Hierarchical Implementation

In this part we present the hierarchical implementation scheme.

First an image pyramid with different resolutions is constructed. The original image is filtered with a low-pass filter and then decimated by a factor of two. If the low-pass filter is chosen to have a sharp cut-off the sharpness of the edges in the original image can be well preserved in the filtered and decimated low resolution image. This obtained low resolution image is then filtered and decimated the same way to acquire the next level image. After the image pyramid is finished we will start the segmentation from the coarsest/lowest resolution image.

The coarsest image is first initialized by K-mean clustering. K-mean clustering is usually computation intensive for large amount of data. Without image pyramid it is time consuming to initialize label with K-mean method. Because the coarsest image is usually in a small size it is easier to use K-mean to get the initial label. The value of \( K \) is not very critical because using the moving average to characterize the region class can adjust adaptively to the local features within the same region class. That means same region class can have different class averages in different places across the image. \( K = 4 \) or 5 is what we normally used in our tests and it gives a good results.

The whole procedure is described as follow.

1. Generate image pyramid with a low-pass filter.
2. Initialization. K-mean clustering is used to get the initial labeling of the coarsest image.
3. Set the moving average window size equal to the whole image.
4. Calculate the moving averages of the feature values using multi-grid interpolation.
5. For all the pixels in the image calculate energy \( E = E_f + \alpha(t) E_p \), update label if necessary.
6. Go to step 4 using new segmentation label until convergence or the number of label changes is less than a threshold.
7. Reduce window size and go to Step 4 until current window size is less than a pre-set minimum window size.
8. Map the current segmentation label to the higher resolution image and set it as the initial label. Go to step 3 using the higher resolution image until the original image is processed.

The flowchart of the whole procedure is shown in Fig. 3.

III. EXPERIMENT RESULTS

In our experiment three methods are used to test their performance in the segmentation of suburban images. The first one is our method. The second one is Markov random Field model with varying weighting parameter and the third one is the adaptive clustering algorithm. Fig. 4(a) is a part of a suburban area with roofs of different colors. Some of the roofs have a color of high contrast against the surrounding area, which are easy to distinguish with most of the segmentation methods. Others have colors which have low contrast against its surrounding. These kinds of roofs are more difficult to separate. The roof information is usually what we require to get the positions of the houses. So we would like the roof area to be segmented from the other part of the image. Fig. 4(b), (c) and (d) are the segmentation results by using our proposed method, MRF method with a varying weighting parameter and
the adaptive clustering algorithm, respectively. We can see that our method has extracted all roofs (see Fig. 4(b)). The gray roofs in the lower right corner of the image got mixed with its surroundings by using MRF model method as shown in Fig. 4(c). And the boundary of these gray roofs didn't get extracted properly in Fig 4(d) by using the adaptive clustering algorithm.

Fig. 5 shows another example. Fig. 5(a) shows a part of a suburban image with long roads and houses. Similarly, Fig. 5(b), 5(c) and 5(d) show the segmentation results using the three methods mentioned above, respectively. We can see from the image that the proposed method well captured the boundary of the road as shown in Fig. 5(b). MRF model method didn't separate the road from the pedestrian area next to it as Fig. 5(c) shows. And the adaptive clustering algorithm over-segmented the road (see Fig. 5(d)).

From the experiment results we can see that the proposed method has improved the segmentation performance. It gives better results than those using the MRF model method and the adaptive clustering algorithm.

Fig. 4. (a) A part of a suburban image with different colors of roofs; (b) Segmentation result using the proposed method; (c) Segmentation result using the MRF method with varying weighting parameter; (d) Segmentation result using ACA.
IV. CONCLUSION AND DISCUSSION

In our method, only color images are considered. The feature we used is color features although other features like texture and shape can also be used with this method. Color features can be represented in different color spaces. Choosing an appropriate color space to work with is very important. RGB is a commonly used color system based on tri-chromatic theory. It is easy to implement but non-linear with visual perception. LUV is more perceptually linear than RGB. From our experiments, LUV gives a better result. Other color spaces such as LAB and HSV are also tested and worked well with our method.

In this paper we presented an unsupervised segmentation method based on Markov Random Field model by using the moving averages to characterize the feature classes in different regions. Using moving averages enables the characteristics of the same region class to adjust adaptively according to the local image features. This gives a better estimation of the region class than using a uniformed class average. A hierarchical implementation scheme working with an image pyramid is also presented to reduce the computation load and speed up the optimization process. We can see from the experiment results that our method gives a reasonable segmentation result for suburban images and can be refined further for practical usage in real applications.

ACKNOWLEDGMENT

The authors would like to express their appreciation for the providing of the test aerial images from the Department of Land, Bathurst, NSW.

REFERENCES