Hierarchical Region Mean-Based Image Segmentation

Slawo Wesolkowski and Paul Fieguth
Systems Design Engineering
University of Waterloo
Waterloo, Ontario, Canada, N2L-3G1
s.wesolkowski@ieee.org, pfieguth@uwaterloo.ca

Abstract

Gibbs Random Fields (GRFs), which produce elegant models, but which have very poor computational speed have been widely applied to image segmentation. In contrast to block-based hierarchies usually constructed for GRFs, the irregular region-based approach is a more natural model in segmenting real images. In this paper, we show that the fine-to-coarse region-based hierarchical regions framework for the well-known Potts model can be extended to non-edge based interactions. By deliberately oversegmenting at the finer scale, the method proceeds conservatively by avoiding the construction of regions which straddle a region boundary by computing region mean differences. This demonstrates the hierarchical method is able to model region interactions through new generalizations at higher levels in the hierarchy which represent regions. Promising results are presented.

Keywords: hierarchical models, image segmentation, Markov Random Fields, Gibbs Random Fields, region-based.

1 Introduction

A key problem in computer vision is to distinguish between separate objects in an image scene. A critical step is that of image segmentation, which seeks to separate objects on the basis of distinct appearance. The image segmentation process is dependent on two interactive components: 1) a pixel dissimilarity criterion and 2) a framework for grouping similar pixels and separating dissimilar ones.

The focus of this paper is the pixel grouping algorithm. That is, given a specified dissimilarity criterion, what is an efficient and effective means of constructing groups of pixels or image segments? We consider region-based hierarchical methods based on Markov/Gibbs Random Fields [5] given the ease of constructing such models for segmentation [7]. Many Gibbs Random Fields methods have been introduced in recent years [3, 5, 7, 9], however, most of these methods are computationally slow and, therefore, not practical.

To increase the convergence speed of the algorithm, it is necessary at some point to move away from processing individual pixels to processing image patches or regions, which can be achieved using hierarchical methods. In fine-to-coarse hierarchical methods, the regions are built from the bottom up [11]. This allows regions to keep arbitrarily complex shapes at ever higher scales since no square or other structure is imposed from the top as in a coarse-to-fine hierarchy. Therefore, the resulting regions can naturally fit the structures of the image being analyzed. In this paper, the main question becomes how to properly merge the small regions obtained after the initial classic GRF algorithm has been applied to the image. In our previous work [11], we extended the Potts model from a pixel-level GRF into a hierarchical GRF accelerating the random walk but keeping the same minimization objective. Here we extend the framework to another region merging modality. Namely, instead of comparing regions on the basis of boundary gradients we propose to use the difference of region means.

The hierarchical regions model shares similarities with a few other models in the literature. Zhu's region competition method [10] is similar in that it minimizes an energy function. However, it differs considerably by fostering "competition" between regions (expanding regions from seeds and allowing region splitting) instead of a careful merging strategy adopted here. Angulo's and Serra's ordered mergings algorithm [1] is similar in that it creates a hierarchy of region mergings however it does this in a morphological and not stochastic framework. Their algorithm requires heuristics for merging regions and a stopping criterion for the algorithm. Other hierarchical methods have also been used such as coarse-to-fine block-based approaches [3, 7]; however, resulting images contain many block artefacts and the computational gain is minimal. Finally, Barbu and Zhu [2] propose a Bayesian method which searches the space of image segmentations to find the global optimum. They re-

Proceedings of the 3rd Canadian Conference on Computer and Robot Vision (CRV'06)
0-7695-2542-3/06 $20.00 © 2006 IEEE

formulate the Swendsen-Wang algorithm for graphs by allowing the algorithm to split, merge or re-group a sizeable subgraph (sets of pixels) and thus by achieving fast mixing at low temperatures, it eliminates the slow Gibbs sampling procedure. Although not hierarchical in nature, this algorithm is similar to ours in that it allows groups of pixel labels to be flipped at any one time. The major difference being the splitting of regions/subgraphs in addition to merging them.

The paper is organized as follows. The second section describes the hierarchical Gibbs Random Fields framework. The third section details the new region mean-based hierarchical approach. Section four presents results while the fifth section concludes the paper.

2 Bayesian Framework

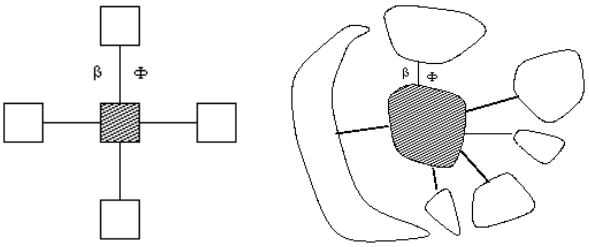
The modelling problems in this paper are addressed from the computational viewpoint by using Gibbs Random Fields to model the image segmentation process. There are two primary concerns: how to *define* an objective function for the optimal solution of the image segmentation, and how to *find* this optimal solution. For the purpose of this paper, the "exact" solution to our segmentation problem will be interpreted as the optimum solution to the optimization objective. In principle, the solution is straightforward: simulated annealing [5] is widely used in solving Gibbs problems; however, it is very slow. The definition of a hierarchical approach produces faster annealing.

Suppose we are given an image $\vec{X}$ with labels l on a pixel lattice $\mathcal{L} = \{i, j\}$ with dissimilarity criterion $\Phi(\cdot)$. We will assume $\mathcal{L}$ has a first order neighborhood structure on a regular grid shown in Figure 1a (a second order neighborhood structure would also be feasible). The energy model is then written as follows:

$$U = \sum_{i,j} \{\Phi(\vec{X}_{i,j}, \vec{X}_{i,j+1})\delta_{l_{i,j},l_{i,j+1}} + \Phi(\vec{X}_{i,j}, \vec{X}_{i+1,j})\delta_{l_{i,j},l_{i+1,j}} + \beta\left[(1-\delta_{l_{i,j},l_{i,j+1}}) + (1-\delta_{l_{i,j},l_{i+1,j}})\right]\} \quad (1)$$

where β controls the relative constraints on the degree of region cohesion and fragmentation, while $\delta_{l_{i,j},l_{i,j+1}}$ is the Kronecker δ. This model operates directly on pixels and is therefore a fine or pixel level model. β is usually determined experimentally. This is essentially a region growing-type model [8, 6] where decisions to integrate a pixel into the region are done with respect to the criterion Φ. The major difference between this local GRF model and region growing methods is that it is non-causal.

Model (1) suffers from a slow random walk of information as shown in Figure 2. This implies that only the slowest of annealing schedules will successfully converge. One way to overcome this limitation would be to merge adjacent regions in successive higher stages after the annealer has converged on the previous finer level. This would occur only if the merging would lower the overall energy.



(a) (b)

Figure 1. Illustration of β and Φ interactions between adjacent pixels/regions: (a) first order neighborhood on a regular grid for the finest or pixel-level model, (b) region neighborhood on an irregular grid for higher level region-based model.

3 Hierarchical GRF Region Grouping

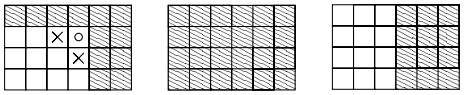
A hierarchical fine-to-coarse region-based approach can be devised where region merging is an integral part of the model. We first reformulate model (1) in order to define interactions between regions:

$$U^{(s)} = \sum_{r,r' \in \mathcal{R}^{(s)}, r \neq r'} \{\Phi^{(s)}_{r,r'}\delta_{l_r,l_{r'}} + \beta^{(s)}_{r,r'}(1-\delta_{l_r,l_{r'}})\} \quad (2)$$

where r is a region indicator, s is the level on the hierarchy, $\mathcal{R}^{(s)}$ is the set of all regions r, $\Phi^{(s)}_{r,r'}$ is the dissimilarity criterion between regions r and r' and $\beta^{(s)}_{r,r'}$ is the region coupling parameter between regions r and r'. When $s = 0$, the formulation corresponds to the special case of model (1) [11] indicating that $\Phi^{(s)}_{r,r'}$ and $\beta^{(s)}_{r,r'}$ define relationships between all pixels (and are non-zero only for adjacent pixels). Furthermore, the neighborhood structure is now defined on an irregular grid as shown in Figure 1b.

This model is non-local in that it operates on regions rather than pixels however it is still Markovian as the information is preserved between levels in the hierarchy. Indeed, the region-to-region interactions are cumulative local interactions between the pixels. This model still performs a random walk; however, the operation is now sped-up since the label comparisons now happen on a regional, multi-pixel level rather than the single pixel interactions of model (1) thus speeding the convergence process considerably.

Proceedings of the 3rd Canadian Conference on Computer and Robot Vision (CRV'06)
0-7695-2542-3/06 $20.00 © 2006 IEEE



(a) (b) (c)

Figure 2. Slow random walk of annealing illustration in 2-D between two regions in a homogenous image (i.e., no energy gradient) for a two-label assignment (Ising model) at $T = 0$. (a) Within the domain of flat energies, the annealer performs a random walk eventually finding one of the optimal endpoints (in this case a shaded or unshaded region). Considering that the probability of flipping each vertex from +1 to -1 is $p_o = 1/2$, vertex 'o' has a 50% probability of becoming $+1$. If it does then pixels 'x' will be in the same situation otherwise nothing changes. (b) To transform this image into a homogeneous patch, a random walk of labels needs to take place. (c) Note, that the random walk can result in local minimum which can then be circumvented at the next level in the hierarchy.

We are able to keep the information from level s to level $s+1$ equivalent by constructing transition equations between levels which transfer the interaction information, as well as by choosing a conservative value for β. The transition for the dissimilarity criterion between two regions r and r' at level $s+1$ is described by a sum of all the individual distances between the regions in $G_r^{(s)}$ and $G_{r'}^{(s)}$:

$$\Phi_{r,r'}^{(s+1)} = \sum_{t \in G_r^{(s)}} \sum_{t' \in G_{r'}^{(s)}} \Phi_{t,t'}^{(s)} \tag{3}$$

where t and t' are region indicators. The coupling parameter between two neighboring regions r and r' at level $s+1$ is written in an analogous fashion:

$$\beta_{r,r'}^{(s+1)} = \sum_{t \in G_r^{(s)}} \sum_{t' \in G_{r'}^{(s)}} \beta_{t,t'}^{(s)} \tag{4}$$

Therefore, we now have model (2) which governs how the labelling is done at each level s together with between level transition equations (3) and (4). Some more details are available in [11].

Given an *appropriate* distance metric Φ and region coupling parameter β, the algorithm performs an accurate oversegmentation of the image at the first level by creating a multitude of small, compact regions. However, the edge-based region formation model still fails in scenarios where region growing algorithms fail, namely, region spilling from one region into the other through a small gradient connection in an otherwise high gradient boundary. This will be illustrated in the results section.

4 Region Mean-Based Model

One way to mitigate the effects of edge-based region comparison is to compare regions according to their means. Computing the mean of a region averages all the pixel values within that region including the edge pixels. This means that parts of a slowly varying gradient should be merged into the corresponding adjacent regions or form another third region. The mean-based model is exactly the same in appearance as model (2). However, the distance computation between regions are now carried out between means. This introduces a subtle shift in the way larger regions form. At this level of pixel organization edges do not matter as much as they did when forming small region patches or blobs. Here we would like to aggregate regions based on their features such as their color. We think that the region mean (in this case first principal component or principal direction of all pixels as opposed to vector mean of pixels) will characterize the regions in a better fashion and facilitate the merging of the regions that should be merged.

In order to make sure that the comparisons between the $\beta_{r,r'}^{(s)}$ and $\Phi_{r,r'}^{(s)}$ values are similar to those in the classic Potts model the new mean-based distance is multiplied by the number of common edge pixels. In this way, the previous edge-based interaction is now somewhat smoothed. In this model, the transitions take on a different form for $\Phi_{r,r'}^{(s)}$ (they remain identical to the edge-based model for $\beta_{r,r'}^{(s)}$). Given that the model no longer cares about individual pixels but groups of pixels, the transitions need to account for changes in the number of pixels in a region, as well as the change in region mean values.

$$x_r^{(s)} = \sum_{t \in G_r^{(s)}} x_t^{(s-1)} \frac{n_t^{(s-1)}}{n_r^{(s)}} \tag{5}$$

$$n_r^{(s)} = \sum_{t \in G_r^{(s)}} n_t^{(s-1)} \tag{6}$$

$$\Phi_{r,r'}^{(s)} = \Phi(x_r^{(s)}, x_{r'}^{(s)}) \tag{7}$$

where at any level s, $n_r^{(s)}$ are the number of pixels in a given region r and $x_r^{(s)}$ represent the region means. Finally, the results of the finest-scale segmentation are first median filtered to remove one-pixel regions which tend to permeate through the hierarchies under the mean model.

The image segmentation algorithm is divided into two parts: a trivial image splitting part in the first step and a region merging part in subsequent iterations.

Proceedings of the 3rd Canadian Conference on Computer and Robot Vision (CRV'06)
0-7695-2542-3/06 $20.00 © 2006 IEEE

Algorithm:

- Assign labels $\{l_{i,j}\}$ randomly to pixels $\{X_{i,j}\}$
- Make each $\{X_{i,j}\}$ its own region
- Loop over levels: from finest (pixel) to coarsest:
  - Anneal until convergence:
    - Apply Gibbs sampling using model (2)
    - Update each region's label based on Gibbs sampling
  - Apply transition equations (4) and (7)

It is difficult to ascertain in theory whether this algorithm converges to the global minimum just like the Gibbs sampler does [5]. Our current difficulty lies in the evaluation of the convergence at intermediate levels in the hierarchy. Each pair of nodes can converge to one of four possible configurations shown in Figure 3. The possibilities Figure 3a-c are acceptable if we are to reach a good minimum point (if not a global minimum) after the last level in the hierarchy converged. However, should Figure 3d occur then a global minimum will never be reached. Therefore, convergence to a good local minimum hinges on a delicate balance of $\Phi^{(s)}_{r,r'}$ and $\beta^{(s)}_{r,r'}$ at each level s. This balance seems to be achieved in our preliminary experiments. Furthermore, we are currently investigating whether convergence can be proven formally.

Other more creative models can also be used where the relationships between the regions are more complex combinations of mean-based and edge-based calculations.

5 Results and Discussion

Results are presented on a color image. The pixel dissimilarity criterion Φ was chosen to be the vector angle measure following [4] as the image has some intensity differences (e.g. shading). Furthermore, the region mean is replaced by its vector angle analogue, namely, the first principal component of the covariance matrix of the pixel vectors. This value represents the most prevalent vector direction in the region. Results for the original edge-based model from [11] are shown in Figure 4. Model (2) encodes only distances between individual pixels and not for example distances between region prototypes [4, 10]. Therefore, regions connected by a slowly varying gradient will be merged as can be clearly seen. However, when the mean-based model is used region spilling does not readily occur.

The results in in Figure 5 show image segmentation using the mean model. Region spilling is less prevalent than in the classic Potts model case however there is still some occurring. The distance calculation is now based on non-local values (i.e., means) and, therefore, the random field is no longer exactly model (2).

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

Figure 3. There are four possible outcomes when combining two nodes A and B which represent pixels or regions: (a) nodes that are supposed to be merged are merged and the energy is lowered, (b) nodes that are supposed to be separated are kept separated and the energy stays the same, (c) nodes that are supposed to be merged are not merged and the energy stays the same, and (d) nodes that are supposed to be separated are merged and the energy is increased. Outcomes (a)-(c) are desirable in that (a) and (b) lead us closer to a global minimum while option (c) hopefully delays the inevitable and will lead to nodes merging at a higher level in the hierarchy. Option (d) should not occur as the energy formulation does not permit it unless there are too few labels to distribute between the node and its neighbors.

We have presented hierarchical regions based on a region mean model. In contrast to the hierarchical region-based Potts model [11] the new model exhibits less region spilling due to the smoothing effects of using means to represent regions instead of edge pixels. However, some region merging persists. We are currently investigating whether the merging effects are caused by the choice of distance metric, a deficiency in the Potts model (or the hierarchical Potts model) or both. This modelling flexibility demonstrates an inherent benefit to the hierarchical method beyond the computational acceleration, namely, that of being able to model region interactions through new generalizations at the higher hierarchies. Finally, we are investigating whether other models might improve segmentation results and hierarchical Potts can be shown (or not) to converge to a global minimum.

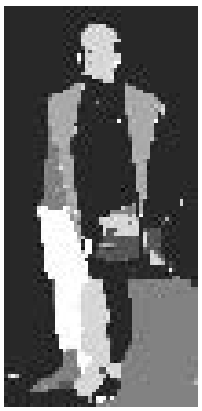
References

[1] J. Angulo, and J. Serra, "Color segmentation by ordered mergings," *IEEE ICIP*, Vol. 2, pp. 125-128, Barcelona: September 2003.

[2] A. Barbu and S.C. Zhu, "Graph Partition by Swendsen-Wang Cut," *IEEE Trans. on Pattern Analy-*

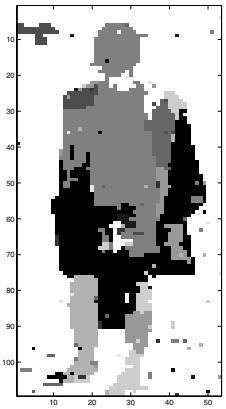
Proceedings of the 3rd Canadian Conference on Computer and Robot Vision (CRV'06)
0-7695-2542-3/06 $20.00 © 2006 IEEE

(a) Original



(b) Final segmentation (12 Levels)

Figure 4. Color image segmentation results with model (2) using $\beta = 0.015$ (image pixels were first normalized to unit length). Upon closer examination (b) shows region spilling from the sea to the shirt (upper central part of image) and from the sand to the coat and leg (lower right part).



(a) $\beta = 0.020$

(b) $\beta = 0.040$

Figure 5. Color image segmentation results with the mean model using different β values.

sis and Machine Intelligence, vol. 27, no. 8, pp. 1239-1253, 2005.

[3] Z. Kato, M. Berthod, and J. Zeroubia, "A Hierarchical Markov Random Field Model and Multitemperature Annealing for Parallel Image Classification," *Graphical Models and Image Processing*, vol. 58, no. 1, 1996, pp. 18-37.

[4] P. Fieguth and S. Wesolkowski, "Highlight and Shading Invariant Color Image Segmentation Using Simulated Annealing," *Energy Minimization Methods in Computer Vision and Pattern Recognition III*, Sophia-Antipolis, France, September 2001, pp. 314-327.

[5] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans-PAMI*, Vol. 6, No. 6, 1984.

[6] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Vol. 1, Addison-Welsey, 1992.

[7] S. Z. Li, *Markov Random Field Modelling in Image Analysis*, Springer: Tokyo, Japan, 2001.

[8] A. Tremeau, and N. Borel, "A Region Growing and Merging Algorithm to Color Segmentation," *Pattern Recognition*, vol. 30, no. 7, pp. 1191-1203, 1997.

[9] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer-Verlag, Berlin, Germany, 1995.

[10] S. C. Zhu and A. Yuille, "Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 9, pp. 884 -900, Sept. 1996.

[11] S. Wesolkowski, and P. Fieguth, "Hierarchical Region-Based Gibbs Random Field Image Segmentation," *International Conference on Image Analysis and Recognition*, Porto, Portugal, Sept 2004, pp. 9-17.

Proceedings of the 3rd Canadian Conference on Computer and Robot Vision (CRV'06)
0-7695-2542-3/06 $20.00 © 2006 IEEE