ABSTRACT

In this paper, we present a regularization approach on discrete graph spaces for perceptual image segmentation via semi-supervised learning. In this approach, first, a spectral clustering method is embedded and extended into regularization on discrete graph spaces. In consequence, the spectral graph clustering is optimized and smoothed by integrating top-down and bottom-up processes via semi-supervised learning. Second, a designed nonlinear diffusion filter is used to maintain semi-supervised learning, labeling and differences between foreground or background regions. Furthermore, the spectral segmentation is penalized and adjusted using labeling prior and optimal window-based affinity functions in a regularization framework on discrete graph spaces. Experiments show that the algorithm achieves perceptual and optimal image segmentation. The algorithm is robust in that it can handle images that are formed in variational environments.

1. INTRODUCTION

Image segmentation is an important but large topic in image and video processing. A meaningful segmentation needs to be integrated with a specific task, i.e., perceptual image identification and segmentation for partially-blurred images. Many current automatic segmentation methods can be considered as directly supervised or indirectly supervised, and towards the target of perceptual image segmentation.

Regularization theory has been recognized as a unified framework for further statistical and mathematical modeling, e.g., variational regularization [1], kernel based regularization [2]. On discrete graph spaces, many methods have been developed, e.g., Markov random field based segmentation methods, graph based energy minimization methods [3]. Based on spectral graph theory, several optimized clustering criteria [4] are used to measure the disassociation between two groups by efficient eigenvector calculations. However, most of these methods are difficult to integrate the “top-down” and “bottom-up” principle for perceptual image segmentation.

To motivate the algorithm, we propose to solve two basic questions. Firstly, a “bottom-up” question is how to get a global convergence of multi-levels of local distributions (local pixel gray level distributions, randomly distributed local blurry regions and unblurred regions, and different color, intensity, texture regions), shown in Fig. 1. Secondly, a “top-down” question is how to achieve perceptual image segmentation using semi-supervised learning and labeling prior.

In this paper, we introduce a novel discrete regularization approach on graph spaces for perceptual image segmentation via semi-supervised learning and optimal control. Conceptually, this approach reveals the roles of optimization algorithms. Algorithms such as graph-cuts [3], variational regularization [1] can be viewed as either discrete regularization with energy on discrete binary spaces or on continuous bounded variation spaces. Compared to Markov random fields based stochastic optimization approaches [5], this paradigm is optimized in a regularized deterministic manner. In contrast to these related work, e.g., discrete regularization [6], [2], spectral clustering and matting [4], [7], background cut [8] and supervised perceptual image segmentation [9], our algorithm has some important improvements and effects: firstly, different from traditional regularization methods, regularization on graph spaces has more flexibilities for achieving optimal and perceptual image segmentation. Secondly, semi-supervised learning and labeling can replace traditional supervised labeling methods for perceptual image segmentation. Furthermore, nonlinear diffusion filter and adaptive window-based affinity functions can improve semi-supervised learning, labeling and image segmentation. Finally, the approach is robust to noise and it can handle different images and video data in real environments.

The paper is organized as follows. In Section 2, we discuss regularization on graph spaces and its segmentation model.
In Section 3, semi-supervised learning and optimal control are presented. Numerical experiments are shown in Section 4. Conclusions are summarized in Section 5.

2. DISCRETE REGULARIZATION ON GRAPHS

A general weighted undirected graph $G = (V, E)$ consists of two sets $A$ and $B$ with edges $E$, i.e., edges with one endpoint in $A$ and the other in $B$, where $V = \{v_i\}_{i=1}^n$ are the vertices and $E \subseteq \{(v_i, v_j)\}$ are the edges between these vertices.

The bisection problem can be formulated as the minimization of a quadratic objective function by means of the Laplacian matrix $L = L(G)$ of the graph $G$. Let $d(i)$ denote the degree of a vertex $i$, i.e., the number of vertices adjacent to $i$. The Laplacian matrix $L$ can be expressed in terms of two matrices associated with a graph as $L = D - W$ in positive semi-definite, $W = \{w_{ij}\}$ is the adjacency matrix of a graph, and $D$ is the $n \times n$ diagonal matrix of the degrees of the vertices in $G$. Let $x$ be an $n$-vector with component $x_i = 1$ if $i \in A$ and $x_i = -1$ if $x \in E$, then $x^T L x = x^T D x - x^T W x = \sum (x_i - x_j)^2$ where $(i,j) \in E, i \in A, j \in B$. The bisection problem is thus equivalent to the problem of maximizing similarity of the objects within each cluster, or, find a cut edge through the graph $G$ with minimal weight in the formulation $\max(x^T W x) \iff \min(x^T L x)$.

Since the bisection-partitioning problem is NP-complete, we need to approximate this intractable problem by some relaxing constraints. Likewise, to avoid unnatural bias of partitioning, a general strategy is to scale the cut weight. Currently, several spectral clustering criteria [4], [9], have been proposed based on pure "bottom-up" processing.

2.1. Discrete Regularization for Bisection Partition

Our goal is to design a discrete regularization approach that is adopted to the structure of graphs and prior information (labeling) for perceptual and optimal image segmentation. We consider the image as the composite of foreground regions $F$ and background regions $B$ in linear blending of radiance values based on $i$-th pixel-level information. The sum of foreground $\alpha F$ and background $(1 - \alpha) B$ is equal to the entire image $I$,

$$I = \alpha F + (1 - \alpha) B$$ (1)

where $\alpha$ is the opacity of foreground regions or objects. $(1 - \alpha)B$ is the rest part of background regions or objects. This equation brings us several meaningful interpretations, e.g., reducing the computation complexity and avoid over segmentation, etc. The computation of $\alpha$ is crucial to segment foreground regions. We use a transform to simplify the formulas by allowing $u = 1/(F - B), v = -B/(F - B)$, the Eq. 1 becomes $\alpha = uI + v$, where $I$ is the input image, and output parameters $\alpha, u$ and $v$.

For the entire image, the cost function on discrete image spaces with respect to $i$-th pixel-vertex becomes,

$$J(\alpha, u, v) = \arg \min_k \{ \sum_{i \in w_k} ||I_i u_k + v_k - \alpha_i||^2 + \varepsilon u_k^2 \}$$

where $\varepsilon u^2$ is a penalty smoothing term with parameter $\varepsilon$, and $w_k$ is a small window around the pixel $k$.

For each small window $w_k$ in the image, the solution can be written in a least squares form,

$$J(\alpha, u, v) = \sum_k \left\| \begin{array}{c} u_k \\ v_k \end{array} \right\| \Psi_k - \bar{\alpha}_k \right\|^2 \quad (2)$$

where $k$ is a pixel vertex. $\Psi_k$ is defined as a matrix $([w_k] + 1) \times 2$ and contains a row of the form $[1, 1]$ for each window $i \in w_k$ and the last row of $\Psi_k$ is $[0, 0]$. The partition region $\bar{\alpha}_k$ is a $([w_k] + 1)$ vector with elements $\alpha_i, i \in w_k$ and the last element is $0, |w_k|$ is the number of pixels in this window. To solve the segmented regions $\alpha_k$, the optimal $\hat{u}_k, \hat{v}_k$ is the solution to the minimization of least squares (LSQ) problem,

$$(\hat{u}_k, \hat{v}_k) = \min \left\| \begin{array}{c} u_k \\ v_k \end{array} \right\| \Psi_k - \bar{\alpha}_k \right\|^2 = (\Psi_k^T \Psi_k)^{-1} \Psi_k^T \bar{\alpha}_k$$

Substituting this solution into the energy minimization in Eq. 2, we get a quadratic cost function with unknown $\alpha$.

$$J(\alpha, u, v) = \sum_k \left\| \Psi_k (\Psi_k^T \Psi_k)^{-1} \Psi_k^T \bar{\alpha}_k - \bar{\alpha}_k \right\|^2 \quad (3)$$

where we denote $\bar{\Psi}_k = I - \Psi_k (\Psi_k^T \Psi_k)^{-1} \Psi_k^T$, then we get

$$J(\alpha) = \sum_k \left\| \bar{\Psi}_k \bar{\alpha}_k \right\|^2 = \sum_k \bar{\alpha}_k^T \bar{\Psi}_k^T \bar{\Psi}_k \bar{\alpha}_k = \alpha^T L \alpha$$

where $L = \bar{\Psi}_k^T \bar{\Psi}_k$ can be expressed in the following,

$$K_{ij} = \frac{1}{|w_k|} \left( 1 + (I_i - \delta_k)(I_j - \delta_k)(\sigma_k^2 + \varepsilon |w_k|)^{-1} \right) \quad (4)$$

where $K_{ij}$ is the kronecker delta, $\delta_k$ and $\sigma_k^2$ are the mean and variation of the intensities in the window $w_k$ around $k$. We refer the semidefinite matrix $L$ to an affinity Laplacian matrix. The matrix $L$ can also be explained in $L = D - W$ in spectral graph theory, with $D(i,i) = \sum_j W(i,j)$ is a diagonal matrix. The $W$ is a symmetric matrix and its off-diagonal matrix are defined by the definition of weights, shown in Fig. 2.

2.2. Regularized Spectral Graph Clustering

Partitioning foreground and background regions is poorly conditioned if these two regions have similar cluttered conditions (blurring, gray value, intensities, discontinuities, structure, and contrasts). To avoid such ill-posed conditions, we embed and extend the spectral graph clustering into regularization so
that the regularization parameters and the penalty term can adjust and optimize the segmentation. To extract the optimal opacity of segmented regions $\alpha$, we construct a regularization energy function which can use prior and labeling information,

$$J(\alpha) = \arg \min \{ \alpha^T L_\alpha + \xi (\alpha^T - d_i^T) D_\alpha (\alpha - d_i) \}$$

where $\xi$ is a strength number of prior labeling, $d_i$ is the vector containing the labeling or prior information and 0 for all other pixels. $D_\alpha$ is a diagonal matrix with diagonal value 1 for detected feature labeling and 0 for all other pixels. Since this energy function is quadratic regularization, we can differentiate the equation and set the derivatives to 0. This equation can be well adapted to use labeling information (e.g., detected feature corners or patches) and difference between foreground and background regions or objects. It also allows a globally optimal partition of $\alpha$ using these sparsely distributed priors or labels to merge unlabeled data via this equation.

### 3. SEMI-SUPERVISED LEARNING AND LABELING

In active learning [10] situation, the learner itself is responsible for acquiring the training set. Active learning minimize the variance component of the estimated generalization error. Here we propose a semi-supervised learning and labeling method in that the labeling is trained and controlled using a designed nonlinear diffusion filter. To achieve high-quality partition, an idea is to enhance the difference between foreground and background regions. In other words, we attenuate the contrast inside one of regions so that the boundary between these two regions can be enhanced. Simultaneously, other inner discontinuities inside each region are eliminated.

Inspired by Perona-Malik (P-M) nonlinear diffusion filter in scale spaces [11], we extend a P-M like filter [8] to attenuate the difference inside the background region while preserving the contrast $Z_{ij}$ across the boundaries between foreground and background regions,

$$Z_{ij} = ||I_i - I_j||^2 \cdot (1 + |\nabla I_B|^2 / K^2 \cdot \exp(d_{ij}^2 / \sigma_d))^{-1}$$

where $K$ is a contrast parameter. Compare to the original P-M filter, this filter includes an exponential Hausdorff distance measure to optimally control the diffusion process.

The $d_{ij}$ is $\max \{ ||I_i - I_j^B||, ||I_j - I_j^B|| \}$ is a Hausdorff distance like definition. $d_{ij}$ measures the dissimilarity between pair $(I_i, I_j)$ in the image $I$ and $\nabla I_B = I_i^B - I_j^B$ measures pixel difference inside background $I_B$. If $d_{ij}$ is small, the attenuation strength should be large ($\exp(-d_{ij}^2 / \sigma_d) \rightarrow 1$), and the pixel pair $(I_i, I_j)$ might belong to the same region (blurred or unblurred). Otherwise, $d_{ij}$ is large ($\exp(-d_{ij}^2 / \sigma_d) \rightarrow 0$), the attenuation strength is small, it probably belongs to the boundary contrast between foreground and background regions. In our experiments, we take $K = [5, 10]$ and $\sigma_d = [10, 50]$, shown in Fig. 3. This P-M like filter has some distinguished role that small scales and noises are smoothed, so if the method is stopped at a suitable final time and given a suitable $K$, noise is smoothed while learning and sparse labeling are enhanced to represent foreground and background.

### 4. EXPERIMENTS AND DISCUSSION

Experiments are carried out to demonstrate the effectiveness of our algorithm. We use a traditional Guassian RBF affinity weight and the optimal window-based affinity weight in Eq. 4 with labeling prior for image segmentation.

First, we use a global Gaussian density based affinity function for image segmentation without using labeling information, $w_{(i,j)} = \exp(-\frac{||Q(i) - Q(j)||^2}{\sigma_i^2}) \cdot \exp(-\frac{||X(i) - X(j)||^2}{\sigma_X^2})$. The results are shown in Fig. 4. Some error happens on the cut edges. The reason is that the $\sigma_i$ and $\sigma_X$ are global image constant. The affinity is so rough that those nearby pixels with relatively similar intensity values are misclassified.

Second, the segmentation results can be optimized using an improved window-based affinity function using Eq. 4 with semi-supervised learning and labeling prior. The window-based affinity functions use local estimates of means of variances values instead of the global derivation. The strength of affinity between two pixels of the gray value decreases with distance, while the affinity between pixels of different gray value is zero. Neighbor pixels with similar gray values have high affinity, otherwise, the affinity is small.

Here, we use Förstner operator find most intensity cross corners as prior labeling via PM-like filter control, shown in Fig. 5(a). The Laplacian extracts partitions tend to be piece-

---

Fig. 2. a|b|c|d. Segmentation using the second generalized eigenvector with normalized cut criterion without prior knowledge. (a) clustering case. (b) Affinity matrix. (c) Corresponding graph weight matrix $W$. (d) Clustering result.

Fig. 3. a|b|c|d columns. Feature corners and labeling in un-blurred regions are controlled via the P-M like filter. (a) Test images. (b) Feature corners on original partially-blurred images. (c) Feature corners is controlled via P-M filter, $K = 10$.

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 24, 2008 at 08:54 from IEEE Xplore. Restrictions apply.
wise constant in the same region where the smallest eigenvectors are piecewise constant. If the values inside a partition in the eigenvector image are coherent, a simple seeds or patch labeling within such a partition is sufficient to attenuate the difference and find the right cut edges to the entire segment based on sparse labeled corners and labeled gray values. In the experiment, we note that these sparse feature labels are sufficient to segment foreground and background regions, shown in Fig. 5(b)(c) and Fig. 1. The color images are separated into RGB color channels and each channel is processed accordingly. However, there are still some parts are not well segmented in that the intensity of gray values are too similar for a small sized window, e.g., $3 \times 3$. Therefore, an optimal control between the size of windows, and reasonable learning and labeling are important for image segmentation results.

The proposed method is also robust in some noise environment due to optimally controlled diffusion and affinity functions. This method uses the advantages from the spectral graph clustering theory and regularization theory and it is robust for real partially-blurred images.

5. CONCLUSION

In this paper, we have proposed a discrete regularization approach to achieve optimal and perceptual image segmentation. This approach integrates the advantages from both spectral graph theory and regularization theory using learning prior and control. Different from existing off-line and supervised labeling methods, this approach allows on-line semi-supervised learning and diffusion adjustment so that we can achieve a meaningful and knowledge-driven image segmentation. The quality of segmentation has been achieved via optimal scale control. This approach has robust performance on natural images and can be easily extended to other tasks.

6. REFERENCES


