Image Segmentation by Bilayer Superpixel Grouping

Michael Ying Yang
Institute for Information Processing (TNT)
Leibniz University Hannover
Appelstr. 9A, Hannover, Germany
yang@tnt.uni-hannover.de

***Abstract*—The task of image segmentation is to group image pixels into visually meaningful objects. It has long been a challenging problem in computer vision and image processing. In this paper we address the segmentation as a superpixel grouping problem. We propose a novel graph-based segmentation framework which is able to integrate different cues from bilayer superpixels simultaneously. The key idea is that segmentation is formulated as grouping a subset of superpixels that partitions a bilayer graph over superpixels, with graph edges encoding superpixel similarity. We first construct a bipartite graph incorporating superpixel cue and long-range cue. Furthermore, mid-range cue is also incorporated in a hybrid graph model. Segmentation is solved by spectral clustering. Our approach is fully automatic, bottom-up, and unsupervised. We evaluate our proposed framework by comparing it to other generic segmentation approaches on the state-of-the-art benchmark database.**

***Keywords*-Segmentation, bilayer graph, spectral clustering, superpixel**

I. Introduction

Image segmentation is a fundamental low-level problem in computer vision and image processing. It provides the basis for high-level image understanding such as object recognition, image retrieval, activity recognition,etc.. Despite a variety of segmentation techniques have been proposed, it remains a challenging problem due to the broad diversity and ambiguity in an image. The task of segmentation is to group image pixels into visually meaningful objects, which are useful for further processing such as recognition.

In image segmentation, one has to consider a prohibitive number of possible pixel groupings. Using prior information about object appearance, or other scene content significantly simplifies the problem. For instance, many segmentation techniques are formulated as a Markov random field based energy minimization problems. However, the corresponding energy functions typically include terms that require prior object knowledge in terms of user interaction [1] or knowledge about object appearance. Approaches to image segmentation in the literature include normalized cuts (Ncut) [2], mean shift (MS) [3], graph-based method (FH) [4], and ultrametric contour maps (UCM) [5]. In recent years an increasingly popular way to solve image segmentation problem uses superpixels [6]. This allows features to be computed over a larger spatial support. In most cases, they are used to initialize segmentation. Endres and Hoiem [7] generated multiple proposals by varying the parameters of a conditional random field built over a superpixel graph. We think of segmentation as a bottom-up preprocessing step for high-level computer vision tasks such as indexing and recognition, providing substantial reduction in the computational complexity of these tasks. It is therefore unclear how segmentation methods that use strong prior knowledge are applicable for object recognition from large databases.

In this paper we address the image segmentation as a superpixel grouping problem, based on the observation that object boundaries are often reasonably well approximated by superpixel boundaries. We propose a novel graph-based segmentation framework which is able to integrate cues from bilayer superpixels simultaneously. Our approach is fully automatic, bottom-up, and unsupervised. The key idea is that segmentation is formulated as grouping a subset of superpixels that partitions a bilayer graph over superpixels, with graph edges encoding superpixel similarity. We first construct a bipartite graph incorporating superpixel cue and long-range cue (neighboring superpixels in two layer). Segmentation is solved by spectral clustering. Furthermore, mid-range cue (neighboring superpixels within one layer) is also incorporated in a hybrid graph model. Given an image, we first compute two layer superpixel segmentation of the image. Based on two superpixel images, segmentation is performed as a superpixel grouping problem.

II. Problem Formulation

In this section, we propose a novel graph-based segmentation framework which is able to integrate different cues from bilayer superpixels simultaneously. We formulate segmentation as a superpixel grouping problem, based on the observation that object boundaries are often reasonably well approximated by superpixel boundaries. A bipartite graph is constructed to incorporate superpixel cue and long-range cue. Segmentation is then solved using spectral clustering. Furthermore, we propose a hybrid graph model that incorporates superpixel cue, mid-range cue, and long-range cue.

A. Bipartite graph construction

We construct a bipartite graph over two layer superpixels of one image I, as shown in Fig. 1. Superpixels are

978-1-4799-2190-4/13 $26.00 © 2013 IEEE
DOI 10.1109/ACPR.2013.62

generated by some unsupervised segmentation algorithms, such as NCut [2], UCM [5], SLIC [6], etc.. Formally, let

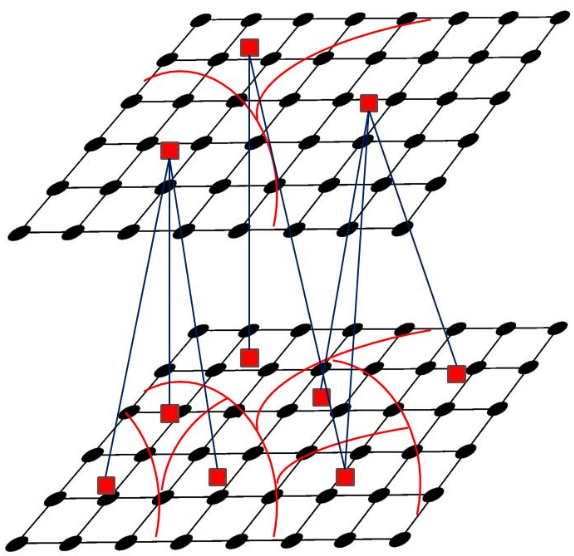

Figure 1. The proposed bipartite graph model of segmentations of an image. A black dot denotes a pixel while a red square denotes a superpixel.

$G_b = (U, V, E_{UV})$ be a bipartite graph with node set $U \cup V$ corresponding to two layers of superpixels and E_{UV} corresponding to graph edges between two layers, where $U = \{u_i\}_{i=1}^n$ and $V = \{v_j\}_{j=1}^m$. Given the above bipartite graph $G_b = (U, V, E_{UV})$, the task is to partition it into k groups. We further define an edge weight w_{ij} to encode the similarity between two superpixels u_i and v_j in two layers that are connected by an edge. The weight matrix $W = (w_{ij})_{n \times m}$ is constructed as follows, which could also be seen as an across-affinity matrix between U and V,

$$w_{ij} = \begin{cases} \alpha & \text{if } |u_i \cap v_j| = \min(|u_i|, |v_j|) \\ e^{-\beta d_{ij}} & \text{if } u_i \sim v_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $|u_i \cap v_j|$ is the number of pixels in the intersection between superpixels u_i, v_j, d_{ij} denotes the distance [1] between the features of superpixels u_i and v_j, $\sim$ denotes a certain neighborhood between superpixels [2], and α, β are free parameters controlling the balance between the superpixel cue and the long-range cue, respectively. By this construction, two neighboring superpixels are more likely to be grouped together if they are closer in feature space.

In [8], the easiness and difficulty of describing one superpixel u_i is evaluated by its description length in terms of visual codewords. Inspired by [8], we define the distance as the Kullback-Leibler divergence between two superpixels u_i and v_j. Specifically, given a dictionary of visual codewords, and the histogram of occurrence of the codewords in u_i, we define

$$d_{ij} = -\log \mathrm{KL}(u_i, v_j) \quad (2)$$

where KL denotes the Kullback-Leibler divergence. Below, we explain how to extract the dictionary of codewords. First, SIFT descriptors [9] are extracted for each pixel of the superpixel at a fixed scale and orientation, using the fast SIFT framework in [10]. The pixel descriptors are then clustered using K-means (with $K = 100$). All pixels grouped within one cluster are labeled with a unique codeword id of that cluster. Then, the histogram of their occurrence in every superpixel is estimated.

B. Superpixel spectral clustering

To make spectral clustering method applicable to our problem, we can simply denote an expanded similarity matrix

$$S = \begin{bmatrix} O & W \\ W^T & O \end{bmatrix} \quad (3)$$

where W is the across-affinity matrix of the bipartite graph G_b. Note that this similarity matrix is sparse and symmetric. We denote by

$$L = I_{n+m} - H^{-1/2} S H^{-1/2} \quad (4)$$

the Laplacian matrix, where I_{n+m} is identity matrix and H the diagonal matrix composed of the row sums of S [2]. It can be easily shown that for any S with nonnegative elements, the Laplacian matrix is symmetric positive semi-definite. Spectral clustering captures essential cluster structure of a graph using the spectrum of graph Laplacian matrix. Mathematically, it solves the generalized eigen-problem [2]:

$$L\nu = \gamma H \nu \quad (5)$$

where γ and ν are corresponding eigen-values and eigen-vectors. The first k generalized eigenvectors $r_1, \cdots, r_k$ of the generalized eigen-problem Eq. (5) are computed by Lanczos method [11], where k is the cluster number. Let $R \in \Re^{(n+m)\times k}$ be the matrix containing the vectors $r_1, \cdots, r_k$ as columns. The $n + m$ rows of R can thus be easily clustered by k-means [12] or certain discretization technique [13].

C. Hybrid graph model

In the above graph construction, our graph model incorporates both long-range grouping cues by bilayer graph construction and short-range superpixel cues by superpixel representation. In addition, mid-range smoothing cues could naturally be incorporated in this graph model, which we call hybrid graph model. let $G = (U, V, E_{UV}, E_U, E_V)$ be a expanded general graph from the bipartite graph G_b with

[1] For example, if color space is used as the feature space, and a superpixel u_i (v_j) is represented by the average color c_i (c_j) of the pixels within it, then $d_{ij} = \|c_i - c_j\|_2$.

[2] For example, $u_i \sim v_j$, if u_i is spatially adjacent to v_j or most similar to v_j in terms of (average) color. This neighborhood relationship is adopted in this paper.

E_U corresponding to graph edges within one layer of U and E_V corresponding to graph edges within the layer of V, as shown in Fig. 2.

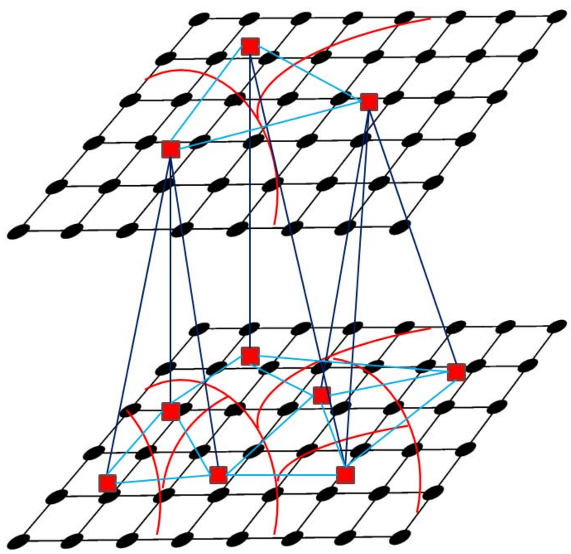

Figure 2. The proposed hybrid graph model of segmentations. A black dot denotes a pixel while a red square denotes a superpixel.

We define an edge weight $p_{ii'}$ ($q_{jj'}$) to encode the similarity between two spatially adjacent superpixels u_i (v_j) and $u_{i'}$ ($v_{j'}$) that are connected by an edge as follows

$$p_{ii'} = \mathrm{TD}(u_i, u_{i'}) \tag{6}$$

where $\mathrm{TD}(u_i, u_{i'}) = \|t_i - t_{i'}\|_1$. t_i is the histogram of texton occurrence of superpixel u_i. The weight matrix for the layer of U is $P = (p_{ii'})_{n\times n}$. In the same way, the weight matrix for the layer of V is $Q = (q_{jj'})_{m\times m}$. The histogram of texton occurrence is computed as follows. We first convert I to grayscale and convolves it with the set of 17 Gaussian derivative and center-surround filters [5], as shown in Fig. 3. We use 8 oriented even- and odd-symmetric Gaussian derivative filters and a center-surround (difference of Gaussians) filter. Each pixel is associated with a 17-d vector of responses, containing one entry for each filter. These vectors are then clustered using K-means (with $K = 64$). The cluster centers define a set of image-specific textons and each pixel is assigned the integer id of the closest cluster center. Then, the histogram of their occurrence (t_i) in every superpixel (u_i) is estimated.

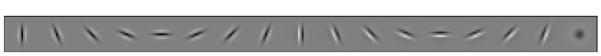

Figure 3. Filters for creating textons [5].

Based on the across-affinity matrix W, and similarity matrices P and Q, we can denote an expanded similarity matrix

$$\tilde{S} = \begin{bmatrix} P & W \\ W^T & Q \end{bmatrix} \tag{7}$$

Then image segmentation using this hybrid graph model can be solved in a similar manner by spectral clustering in Sec. II-B. Simply replace S by $\tilde{S}$ in Eq. (4) to compute the Laplacian matrix. The overall superpixel segmentation algorithm is summarized in Algorithm 1.

Algorithm 1 Image segmentation by bilayer superpixel grouping

Input: Image I, k: number of clusters

1. Partition I into superpixels U and V by segmentation algorithm
2. Construct the graph $G = (U, V, E_{UV}, E_U, E_V)$
3. Compute across-affinity matrix W based on Eq. (1)
4. Compute affinity matrix P (Q) based on Eq. (6)
5. Build similarity matrix $\tilde{S}$ according to Eq. (7)
6. Compute the Laplacian matrix L
7. Compute the first k generalized eigenvectors $r_1, \cdots, r_k$ of Eq. (5)
8. Let $R \in \Re^{(n+m)\times k}$ be the matrix containing the vectors $r_1, \cdots, r_k$ as columns, use k-means algorithm to cluster $(n + m)$ rows of R into k groups

Output: k clusters

III. Experimental results

We evaluate the proposed image segmentation algorithm on some images from Berkeley Segmentation Data Set (BSDS), and compare it with state-of-the-art methods.

Implementation details Our framework builds a graph on superpixel nodes, which are generated by SLIC [6], though other choices are also possible. The main reason of choosing SLIC is that it is currently state-of-the-art superpixel segmentation algorithm and practically efficient. The SLIC parameters are the region size and the regularizer. For our experiments, we set region size proportional to the image size to make around 200 and 100 superpixels for two layers for every image. The regularizer is set as 0.15 for all the images. The parameters in the bipartite graph construction are set as follows $\alpha = 0.9$, and $\beta = 0.35$. The number of clustering k is set to 6 for all the experiments.

Results Fig. 4 shows the segmentation results for an example image. The red boundary overlays with the superpixel segmentation image for visualization. By comparing with mean shift, normalized cut and UCM segmentation methods, our proposed bipartite and hybrid segmentation methods produce more reasonable results with respect to object boundaries and small objects.

Some more segmentation examples of BSDS images can be visualized in Fig. 5. The top 4 rows are perceptually

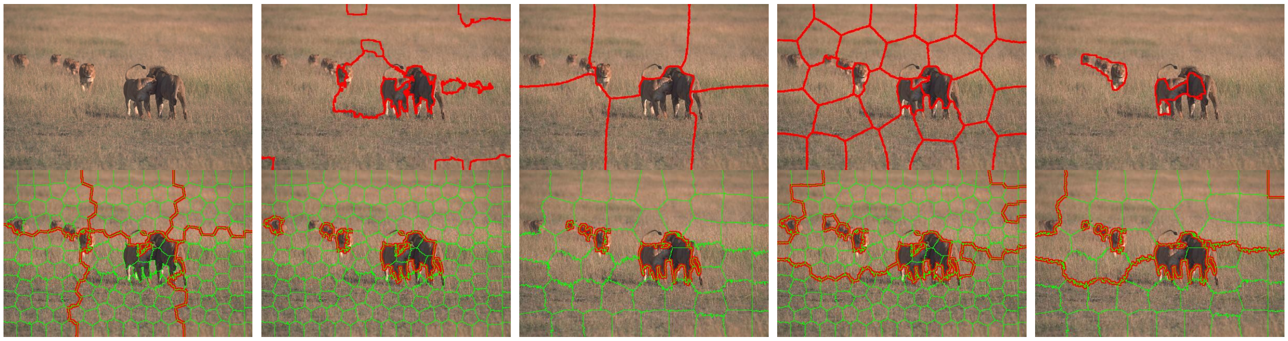

Figure 4. Segmentation example of lion image. Top row (from left to right): original image, mean shift [3] segmentation result, normalized cut [2] result with 6 region partition, normalized cut result with 30 region partition, and UCM [5] result; Bottom row: flat clustering result with local neighborhood information, bipartite segmentation result (bottom layer), bipartite segmentation result (top layer), hybrid segmentation result (bottom layer), and hybrid segmentation result (top layer).

satisfactory results, and the bottom 2 rows show the typical failure cases. We report the results from the flat clustering with only local neighborhood information, bipartite segmentation results, and hybrid segmentation results. These results demonstrate the high performance of our methods on this dataset. Note that it is usually difficult for many segmentation algorithms, even the ones incorporating high-level shape priors, to segment highly textured objects from textured background. Our methods provides perceptually satisfactory results in the bear and lion images. For the typical failure cases, these images usually contain complex object appearance and texture background.

IV. Conclusion

We have presented a novel graph-based framework for image segmentation, which is formulated as grouping a subset of superpixels that partitions a bilayer graph over superpixels, with graph edges encoding superpixel similarity. A bipartite graph is constructed to incorporate superpixel cue and long-range cue. Segmentation is then solved using spectral clustering. Furthermore, we propose a hybrid graph model that incorporates superpixel cue, mid-range cue, and long-range cue. The scheme is fully automatic, bottom-up, and unsupervised. The experiments demonstrate the high performance of our approach on the challenging dataset. Future work should study the incorporation of high-level cues.

Acknowledgment

The work is funded by ERC-Starting Grant ('DYNAMIC MINVIP'). The author gratefully acknowledges the support.

References

[1] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics*, vol. 23, pp. 309–314, 2004.

[2] J. Shi and J. Malik, "Normalized cuts and image segmentation," *PAMI*, vol. 22, no. 8, pp. 888–905, 2000.

[3] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *PAMI*, vol. 24, no. 5, pp. 603–619, 2002.

[4] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *IJCV*, vol. 59, no. 2, pp. 167–181, 2004.

[5] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *PAMI*, vol. 33, no. 5, pp. 898–916, 2011.

[6] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *PAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.

[7] I. Endres and D. Hoiem, "Category independent object proposals," in *ECCV (5)*, 2010, pp. 575–588.

[8] S. Bagon, O. Boiman, and M. Irani, "What is a good image segment? a unified approach to segment extraction," in *ECCV (4)*, 2008, pp. 30–44.

[9] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[10] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," *http://www.vlfeat.org/*, 2008.

[11] G. Golub and C. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, 3rd edition, 1996.

[12] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *NIPS*, 2001, pp. 849–856.

[13] S. Yu and J. Shi, "Multiclass spectral clustering," in *ICCV*, 2003, pp. 313–319.

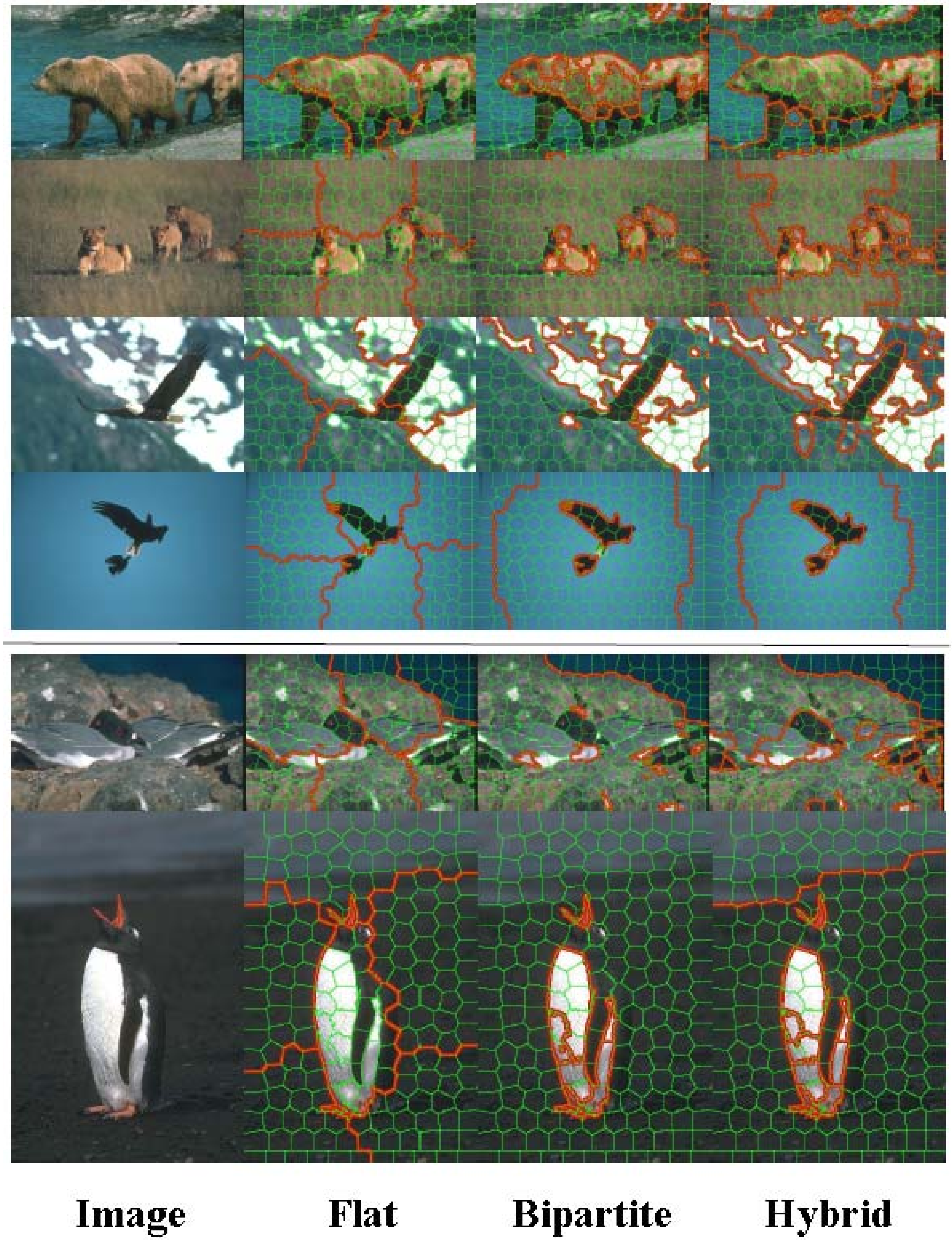

Figure 5. Segmentation example of BSDS images. Top 4 rows (from left to right): original image, flat clustering result with local neighborhood information, bipartite segmentation result, and hybrid segmentation result (the red boundary overlays with the superpixel segmentation image for visualization). Bottom 2 rows: typical failure cases.