Multi-task Low-rank Affinity Pursuit for Image Segmentation

Bin Cheng[1], Guangcan Liu[1], Jingdong Wang[2], Zhongyang Huang[3], Shuicheng Yan[1]
[1] Department of Electrical and Computer Engineering, National University of Singapore, Singapore;
[2] Microsoft Research Asia, China; [3] Panasonic Singapore Laboratories, Singapore;
{chengbin, eleliug, eleyans}@nus.edu.sg, jingdw@microsoft.com, zhongyang.huang@sg.panasonic.com

Abstract

This paper investigates how to boost region-based image segmentation by pursuing a new solution to fuse multiple types of image features. A collaborative image segmentation framework, called multi-task low-rank affinity pursuit, is presented for such a purpose. Given an image described with multiple types of features, we aim at inferring a unified affinity matrix that implicitly encodes the segmentation of the image. This is achieved by seeking the sparsity-consistent low-rank affinities from the joint decompositions of multiple feature matrices into pairs of sparse and low-rank matrices, the latter of which is expressed as the production of the image feature matrix and its corresponding image affinity matrix. The inference process is formulated as a constrained nuclear norm and $\ell_{2,1}$-norm minimization problem, which is convex and can be solved efficiently with the Augmented Lagrange Multiplier method. Compared to previous methods, which are usually based on a single type of features, the proposed method seamlessly integrates multiple types of features to jointly produce the affinity matrix within a single inference step, and produces more accurate and reliable segmentation results. Experiments on the MSRC dataset and Berkeley segmentation dataset well validate the superiority of using multiple features over single feature and also the superiority of our method over conventional methods for feature fusion. Moreover, our method is shown to be very competitive while comparing to other state-of-the-art methods.

1. Introduction

The task of *image segmentation* [36] is widely accepted as a crucial function for high-level image understanding. As pointed out by [14, 25, 31], a successful image segmentation algorithm can significantly reduce the complexity of object segmentation and recognition, which form the core of high-level vision. Hence, it has been widely studied in computer vision [1, 3, 6, 7, 10, 12, 29, 30, 34, 35]. Generally, image segmentation is a comprehensive task which is related with several cues, $e.g.$ regions, contours and textons [21]. In particular, in this paper we are interested in region-based methods [11], which aim at partitioning an image into homogenous regions by grouping together the basic image elements (e.g., superpixels) with similar appearances.

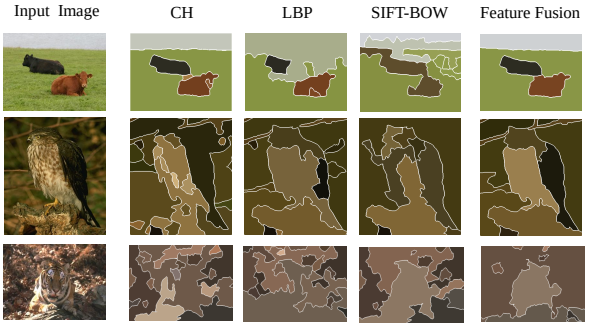

Figure 1. Illustration of the necessity and superiority of fusing multiple types of features. From left to right: the input images; the segmentation results produced by CH; the results produced by LBP; the results produced by SIFT based bag-of-words (SIFT-BOW); the results produced by integrating CH, BLP and SIFT-BOW. These examples are from our experiments.

Many efforts have been devoted to this topic (e.g., [8, 19, 31, 32]). However, some critical problems remain unsolved. Most existing region-based methods focused on exploring the criteria, such as the widely used normalized cut (NCut) [31] and the recently established minimum description length (MDL) [19], for seeking the optimal segmentation. The feature space is however usually predetermined by mildly choosing a feature descriptor such as the color histograms (CH). However, as a data clustering problem, image segmentation performance heavily depends on the choice of the feature space. What is more, it is hard to find a single feature descriptor that can generally work well for various images with diverse properties, since each feature descriptor generally has its own advantages and limitations: the CH descriptor is very informative for describing color, but inappropriate for describing other visual information; the local binary pattern (LBP) [18] descriptor

can defend the change of light conditions, but may cause some loss of information; the scale invariant feature transform (SIFT) [17] descriptor can be invariant to some image transformations, but some useful information of the original image may also be lost. Hence, it is crucial to establish a good solution that can integrate multiple types of image features for more accurate and reliable segmentation. Figure 1 further illustrates the necessity and superiority of fusing multiple types of features.

Although image segmentation may intuitively benefit from the integration of multiple features, to the best of our knowledge, there is no previous work that intensively explores the fusion of multiple features in region-based segmentation. This is mainly due to the fact that it is actually not easy to well handle the multiple features of various properties. In machine learning community, the methods towards this issue are also quite limited. Possibly, the multi-view spectral clustering technique established by Zhou et al. [38] is an optional choice. Namely, one could first construct an undirected (or directed) graph by inferring an affinity matrix from each type of image features, resulting in a multi-view graph (there are multiple affinities between each pair of nodes), and then obtain the segmentation results by combing those multiple affinity matrices [38]. However, this option may not fully capture the advantages of multiple features, because the affinity matrices are still computed from different features *individually*, and thus the cross-feature information is not well considered during the inference process.

To make effective use of multiple features, in this paper we introduce the so-called multi-task low-rank affinity pursuit (MLAP) method, which aims at inferring a unified affinity matrix from multiple feature spaces, and thus producing accurate and reliable segmentation results. Like the traditional methods such as NCut [31], we also treat image segmentation as a graph partitioning problem. That is, an image is represented as an undirected graph with each node corresponding to a *superpixel* [24]. Then the segmentation can be done by partitioning the nodes of the graph into groups. Unlike existing methods, which usually adopt a single feature space, each node (superpixel) in our method is described by multiple features with different properties. To integrate those multiple features and make effective use of the cross-feature information, our MLAP method infers a unified affinity matrix by seeking the sparsity-consistent low-rank affinities from the joint decompositions of multiple feature matrices into pairs of sparse and low-rank matrices, the latter of which is the production of image feature matrix and its corresponding low-rank affinity matrix. The inferring process is formulated as a constrained nuclear norm and $\ell_{2,1}$-norm minimization problem, which is convex and can be solved efficiently with augmented Lagrange multiplier (ALM) [13] method. Provided with the affinity matrix encoding the similarities among superpixels, the final segmentation result can be simply obtained by applying the NCut algorithm to the inferred affinities. Compared with existing methods, the contributions of this work mainly include:

- We propose a method for learning a unified affinity matrix from multiple feature spaces and so performing image segmentation collaboratively. Since the cross-feature information has been well considered, such a joint inference scheme can produce more accurate and reliable results than those methods directly combining multiple affinity matrices, each of which is learnt individually.
- We introduce a simple yet effective new image segmentation algorithm that achieves comparable performance with the state-of-the-art methods, as demonstrated on the MSRC database [33] and the Berkeley dataset [3].

2. Image Segmentation by Multi-task Low-rank Affinity Pursuit

2.1. Problem Formulation

For efficiency, superpixels other than image pixels are used as basic image elements. Using the over-segmentation algorithm in [24], a given image is partitioned into subregions, each of which is called a superpixel. In this way, the problem of segmenting the image is cast into clustering the superpixels into groups. By choosing an appropriate feature descriptor to describe each superpixel, the image segmentation problem can be formulated as follows.

Problem 2.1 *Let $X = [x_1, x_2, \cdots, x_N]$ be a feature matrix, each column of which is a feature vector x_i corresponding to a superpixel P_i. Then the task is to segment the superpixels into groups according to their features represented by X.*

A weak point of the above definition is that only one type of features is considered. To boost the performance, the problem definition based on multiple types of features can be formulated as below.

Problem 2.2 *Let $X_1, X_2, \cdots, X_K$ be K feature matrices for K types of features, where the columns in different matrices with the same index correspond to the same superpixel. Then the task is to segment the superpixels into groups by integrating the feature matrices $X_1, \cdots, X_K$.*

2.2. Multi-task Low-rank Affinity Pursuit

For easy of understanding, Problem 2.1 is explored first. Accordingly, the case for the multiple types of features targeting on Problem 2.2 will be further examined with a well-established solution.

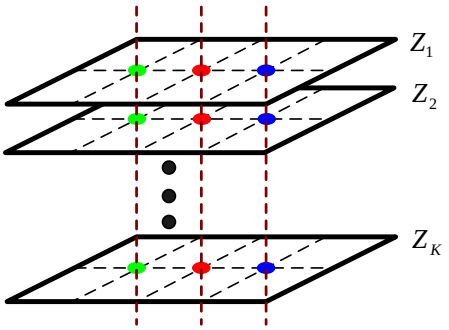

Figure 2. Illustration of the $\ell_{2,1}$-norm regularization defined on Z. Generally, this technique is to enforce the matrices $Z_i, i = 1, 2, \cdots, K$, to have sparsity-consistent entries.

2.2.1 Single-Feature Case (Problem 2.1)

According to the explorations in [20], the superpixel data from natural image usually has a structure of low-rank subspace, i.e., the task of segmenting the superpixels into homogeneous regions could be cast as segmenting the feature vectors into their respective subspaces. Hence, the task stated in Problem 2.1 may be handled by the subspace segmentation algorithms, e.g., the Sparse Subspace Clustering (SSC) algorithm presented in [5] and the Low-Rank Representation (LRR) algorithm presented in [16]. LRR based modeling is chosen for single-feature case owing to its effectiveness and robustness. Namely, for a matrix $X = [x_1, x_2, \cdots, x_N]$ with each x_i representing the i-th superpixel, the affinities among superpixels are computed by solving the following LRR problem:

$$\begin{aligned} \min_{Z_0, E_0} \quad & \|Z_0\|_* + \lambda \|E_0\|_{2,1}, \\ \text{s.t.} \quad & X = XZ_0 + E_0, \end{aligned} \tag{1}$$

where $\|\cdot\|_*$ denotes the nuclear norm, also known as the trace norm or Ky Fan norm (sum of the singular values), $\|\cdot\|_{2,1}$ is the $\ell_{2,1}$-norm [16, 15] for characterizing noise and the parameter $\lambda > 0$ is used to balance the effects of the two parts.

According to [15, 16], the optimal solution Z_0^* (with respect to the variable Z_0) to problem (1) naturally forms an affinity matrix that represents the pairwise similarities among superpixels. Namely, the affinity $\mathbb{S}_{ij}$ between two superpixels P_i and P_j could be computed by $\mathbb{S}_{ij} = |(Z_0^*)_{ij}| + |(Z_0^*)_{ji}|$, where $(\cdot)_{ij}$ denotes the (i,j)-th element of a matrix. Provided with such symmetric affinities, the NCut method in [31] can be applied for producing the final image segmentation results.

2.2.2 Multi-feature Case (Problem 2.2)

The above LRR can only be used to a certain type of visual features and not directly applicable for multi-feature cases. For multiple feature integration, an intuitive approach is to directly combine the affinity matrices individually inferred by LRR. The combination can be done by simply adding together multiple affinities or utilizing the multi-view spectral clustering technique presented in [38] to produce the final segmentation results. However, the inference of the individual affinity matrix does not well utilize the cross-feature information, which is crucial to produce accurate and reliable results.

For effectively fusing multiple features, we propose a new solution of multi-task low-rank affinity pursuit (MLAP) that aims at *jointly* inferring a collection of affinity matrices $Z_1, Z_2, \cdots, Z_K$, where each $N \times N$ matrix Z_i corresponds to the i-th feature matrix X_i. Here, our consideration for formulating the inference process is two-side: to inherit the advantages of LRR, the affinity matrices should be encouraged to be of low-rank; to make effective use of the cross-feature information, the affinity matrices may be enforced to be sparsity-consistent. By considering both sides, the affinity matrices $Z_1, Z_2, \cdots, Z_K$ in MLAP are inferred by solving the following convex optimization problem:

$$\begin{aligned} \min_{\substack{Z_1,\cdots,Z_K \\ E_1,\cdots,E_K}} \quad & \sum_{i=1}^{K} (\|Z_i\|_* + \lambda \|E_i\|_{2,1}) + \alpha \|Z\|_{2,1}, \\ \text{s.t.} \quad & X_i = X_i Z_i + E_i, i = 1, \cdots, K, \end{aligned} \tag{2}$$

where $\alpha > 0$ is a parameter and the $K \times N^2$ matrix Z is formed by concatenating $Z_1, Z_2, \cdots, Z_K$ together as the following:

$$Z = \begin{bmatrix} (Z_1)_{11} & (Z_1)_{12} & \cdots & (Z_1)_{NN} \\ (Z_2)_{11} & (Z_2)_{12} & \cdots & (Z_2)_{NN} \\ \vdots & \vdots & \ddots & \vdots \\ (Z_K)_{11} & (Z_K)_{12} & \cdots & (Z_K)_{NN} \end{bmatrix}.$$

The $\ell_{2,1}$-norm regularization defined on Z plays a key role in our MLAP method: it is the minimization of $\|Z\|_{2,1}$ that enforces the affinities $(Z_l)_{ij}, l = 1, 2, \cdots, K$, to have consistent magnitudes, all either large or small, as shown in Figure 2. That is, the fusion of multiple features is "seamlessly" performed by minimizing the $\ell_{2,1}$-norm of Z. Without this regularization term, the formulation (2) will reduce to a "trivial" method that is equal to applying LRR to each feature matrix X_i individually.

Let $(Z_1^*, Z_2^*, \cdots, Z_k^*)$ be the optimal solution to problem (2). To obtain a unified affinity matrix, we only need a simple step to quantify the columns of the matrix Z:

$$\mathbb{S}_{ij} = \frac{1}{2}\left(\sqrt{\sum_{l=1}^{K} (Z_l)_{ij}^2} + \sqrt{\sum_{l=1}^{K} (Z_l)_{ji}^2}\right). \tag{3}$$

Algorithm 1 Image Segmentation by MLAP

Input: An image and the required parameters.

1. Separate the image into superpixels by using the algorithm in [24].
2. Compute K feature matrices by extracting K types of features to describe each superpixel.
3. Obtain the sparsity-consistent low-rank affinity matrices $Z_1, \cdots, Z_K$ by solving problem (2), and define the edge weights of an undirected graph according to (3).
4. Use NCut to segment the nodes of the graph into a pre-specified number of groups.

Output: A map that encodes the segmentation result.

Note here that $\sqrt{\sum_{l=1}^{K}(Z_l)_{ij}^2}$ is right the ℓ_2-norm of the $((i-1)n+j)$-th column of Z used in (2) and thus (3) should not be considered as late fusion of Z_i's. Same as single feature case, the NCut method can be applied on such affinity matrix to produce the final image segmentation results. Algorithm 1 summarizes the entire image segmentation algorithm of MLAP.

2.3. Optimization Procedure

Problem (2) is convex and can be optimized in polynomial time. We first convert it into the following equivalent problem:

$$\begin{aligned} \min_{\substack{J_1,\cdots,J_K,\\ S_1,\cdots,S_K\\ Z_1,\cdots,Z_K\\ E_1,\cdots,E_K}} \quad & \sum_{i=1}^{K}(\|J_i\|_* + \lambda\|E_i\|_{2,1}) + \alpha\|Z\|_{2,1}, \quad (4)\\ \text{s.t.} \quad & X_i = X_iS_i + E_i,\\ & Z_i = J_i,\\ & Z_i = S_i, i = 1, \cdots, K. \end{aligned}$$

This problem can be solved with the augmented Lagrange multiplier (ALM) method [13], which minimizes the following augmented Lagrange function:

$$\begin{aligned} \alpha\|Z\|_{2,1} + \sum_{i=1}^{K}(\|J_i\|_* + \lambda\|E_i\|_{2,1}) + \sum_{i=1}^{K}(\langle W_i, Z_i - J_i\rangle + \\ \langle Y_i, X_i - X_iS_i - E_i\rangle + \langle V_i, Z_i - S_i\rangle + \\ \frac{\mu}{2}\|X_i - X_iS_i - E_i\|_F^2 + \frac{\mu}{2}\|Z_i - J_i\|_F^2 + \frac{\mu}{2}\|Z_i - S_i\|_F^2), \end{aligned}$$

where $Y_1, \cdots, Y_K$, $W_1, \cdots, W_K$ and $V_1, \cdots, V_K$ are Lagrange multipliers, and $\mu > 0$ is a penalty parameter. The inexact ALM method [13], also called alternating direction method (ADM) [13], is outlined in Algorithm 2. Note that the sub-problems of the algorithm are convex and they all have closed-form solutions. Step 1 is solved via the singular value thresholding operator [4], while Steps 3 and 4 are solved via Lemma 3.2 of [16].

Algorithm 2 Solving Problem (2) by ADM

Inputs: Data matrices $\{X_i\}$, parameters λ and α.

while not converged **do**

1. Fix the others and update $J_1, \cdots, J_K$ by

$$J_i = \arg\min_{J_i} \frac{1}{\mu}||J_i||_* + \frac{1}{2}||J_i - (Z_i + \frac{W_i}{\mu})||_F^2.$$

2. Fix the others and update $S_1, \cdots, S_K$ by

$$\begin{aligned} S_i = (\mathtt{I} + X_i^TX_i)^{-1}(X_i^T(X_i - E_i) + Z_i \\ + \frac{X_i^TY_i + V_i - W_i}{\mu}). \end{aligned}$$

3. Fix the others and update Z by

$$Z = \arg\min_{Z} \frac{\alpha}{2\mu}\|Z\|_{2,1} + \frac{1}{2}\sum_{i=1}^{K}\|Z - M\|_F^2,$$

where M is a $K \times N^2$ matrix formed as follows:

$$M = \begin{bmatrix} (F_1)_{11} & (F_1)_{12} & \cdots & (F_1)_{nn} \\ (F_2)_{11} & (F_2)_{12} & \cdots & (F_2)_{nn} \\ \vdots & \vdots & \ddots & \vdots \\ (F_K)_{11} & (F_K)_{12} & \cdots & (F_K)_{nn} \end{bmatrix},$$

where $F_i = (J_i + S_i - (W_i + V_i)\mu)/2, i = 1, \cdots, K$.

4. Fix the others and update $E_1, \cdots, E_K$ by

$$E_i = \arg\min_{E_i} \frac{\lambda}{\mu}||E_i||_{2,1} + ||E_i - (X_i - X_iS_i + \frac{Y_i}{\mu})||_F^2.$$

5. Update the multipliers

$$\begin{aligned} Y_i &= Y_i + \mu(X_i - X_iS_i - E_i),\\ W_i &= W_i + \mu(Z_i - J_i),\\ V_i &= V_i + \mu(Z_i - S_i). \end{aligned}$$

6. Update the parameter μ by $\mu = \min(\rho\mu, 10^{10})$ ($\rho = 1.1$ in all experiments).

7. Check the convergence condition: $X_i - X_iS_i - E_i \to 0$, $Z_i - J_i \to 0$ and $Z_i - S_i \to 0$, $i = 1, \cdots, K$.

end while

Output: Z.

2.4. Discussions

2.4.1 On the Optimization Algorithm

Since ADM is a variation of the exact ALM method whose convergence properties have been generally proven, Algorithm 1 should converge well in practice, although proving the convergence properties of ADM in theory is still an open issue [37]. Supposing the number of superpixels is N,

then the computation complexity of Algorithm 1 is $O(N^3)$, which is practical (note that the complexity is actually the same as computing the SVD of an $N \times N$ matrix). In our experiments, since the number of superpixels N is small ($N \approx 100$), the computational cost of Algorithm 1 is actually low. On an Intel Xeon $X5450$ workstation with 3.0 GHz CPU and 16GB memory, for example, it takes about 20 seconds to finish the computation for an image.

2.4.2 On the Extension to Multiple Visual Cues

Since our current MLAP method requires the features to be represented by vectors, it can only model image regions, which however only form one aspect of the visual cues. Generally, as pointed out by Malik et al. [3, 21], the other cues such as contour and spatial information should also be taken into account. Fortunately, it is actually feasible for our method to handle multiple cues. Namely, the affinity matrix can be learnt by jointly utilizing the information supplied by other cues. For example, when two superpixels P_i and P_j are separated by a strong contour, the edge between them may be removed (i.e., set $\mathbb{S}_{ij} = 0$) or give high penalty to Z_{ij}'s. In a similar way, our method may also model the spatial information. We leave these as our future work.

3. Experiments

3.1. Experiment Setting

3.1.1 Datasets

We use two publicly available databases, the MSRC [33] dataset and the Berkeley [3] segmentation dataset, in our experiments. The MSRC dataset consists of 591 images from 23 categories. Here we use the cleaned up ground-truth object instance labeling [22], which is cleaner and more precise than the original data. The Berkeley segmentation dataset is comprised of 500 natural images, which cover a variety of nature scene categories, such as portraits, animals, landscape, beaches and so on. It also provides ground-truth segmentation results of all the images obtained by several human subjects. On average, five segmentation maps are available per image.

3.1.2 Superpixel and Features

As aforementioned, we need to partition each image into superpixels. There are several methods that can be used to obtain a superpixel initialization, such as those from Mori et al. [24], Felzenszwalb et al. [9] and Ren et al. [28]. Here we use the method in [24] and the number of superpixels for each image is set to be around 100 in our experiments.

After the superpixel initialization, three different types of features, color histogram (CH), local binary pattern (LBP) and bag-of-visual-words (BOW), are used to describe the appearance of each superpixel. Color histogram (CH) represents the number of pixels that have colors in each of a fixed list of color ranges. It can be built for any kind of color space such as RGB or HSV. In the experiments we use the RGB histogram. The local binary pattern (LBP) operator describes each pixel by the relative gray-levels of its neighbor pixels. In our experiments, the LBP codes are computed using 8 sampling points on a circle of radius 1. Bag-of-visual-words (BOW) is also applied to describe the appearance of the superpixels. In the experiments we compute the SIFT [17] features for each pixel and then perform the vector quantization to construct the visual vocabulary by K-means clustering approach. The number of clustering centers is set to be 50 for MSRC and 100 for Berkeley.

3.1.3 Baselines and Evaluation

To demonstrate the advantages of fusing multiple types of features, we consider the individual performance of applying LRR to a certain single feature, resulting in three benchmark baselines: LRR (CH), LRR (LBP) and LRR (BOW). Moreover, we also report the performance of our MLAP solution when integrating two types of features, which results in three competing methods: MLAP (CH+LBP), MLAP (CH+BOW) and MLAP (LBP+BOW). To show the superiority of our formulation in (2), we also consider some "naive" methods for fusing multiple types of features, including using the multi-view technique to combine multiple affinity matrices (denoted as "Multi-view"), an approach of stacking together multiple types of features to form a unified long vector (denoted as "Vector-stack"), a widely used approach of utilizing NCut to combine multiple affinity matrices, and a benchmark approach that performs the Mean-shift [6] operator on multiple features (denoted as "Mean-shift"). For presentation convenience, we refer to our method of integrating all three features as "MLAP(CH+LBP+BOW)".

To quantitatively evaluate the performance of various methods, three metrics for comparing pairs of image segmentations are used: the variation of information (VOI) [23], the probabilistic rand index (PRI) [26] and the segmentation covering rate (CR) [2].

3.2. Experiment Results

3.2.1 Main Results

We evaluate and compare all the algorithms under a unified setting, as discussed in Section 3.1. In summary, the results shown in Table 1 well verify the advantage of fusing multiple types of features and the superiority of our MLAP method. Namely, while all three features are combined together for segmentation, our method distinctly outperforms the other methods (which also use multiple types of features), including Multi-view, Vector-stack, NCut and Mean-

Table 1. Evaluation results on the MSRC dataset and the Berkeley 500 segmentation dataset. The details of all the algorithms are presented in Section 3.1.3. The results are obtained over the best tuned parameters for each dataset (the parameters are uniform for an entire dataset). For comparison, we also include the results reported in [27], but note that, for the Berkeley dataset, [27] used Berkeley 300 instead.

	MSRC			Berkeley		
	VOI	PRI	CR	VOI	PRI	CR
MLAP(CH+LBP+BOW)	**1.1656**	**0.8306**	**0.7556**	**1.5311**	**0.8538**	**0.6411**
MLAP(CH+LBP)	1.1931	0.8020	0.7121	1.6573	0.8401	0.6227
MLAP(CH+BOW)	1.2505	0.7967	0.7032	1.7262	0.8320	0.6109
MLAP(LBP+BOW)	1.4245	0.7560	0.6541	1.7626	0.8305	0.5947
LRR(CH)	1.3002	0.7912	0.6932	1.7475	0.8295	0.5905
LRR(LBP)	1.4449	0.7490	0.6415	1.7875	0.8261	0.5734
LRR(BOW)	1.4880	0.7343	0.6275	1.8585	0.8045	0.5670
Multi-View	1.2511	0.8116	0.7194	1.6664	0.8441	0.6272
Vector-stack	1.4107	0.7728	0.6668	1.8993	0.7815	0.5516
NCut	1.2516	0.8052	0.7075	1.7235	0.8283	0.6054
Mean-shift	1.7472	0.7307	0.5983	2.0872	0.7196	0.5272
Ma et al. [27]*	1.49	0.76	–	1.76	0.80	–

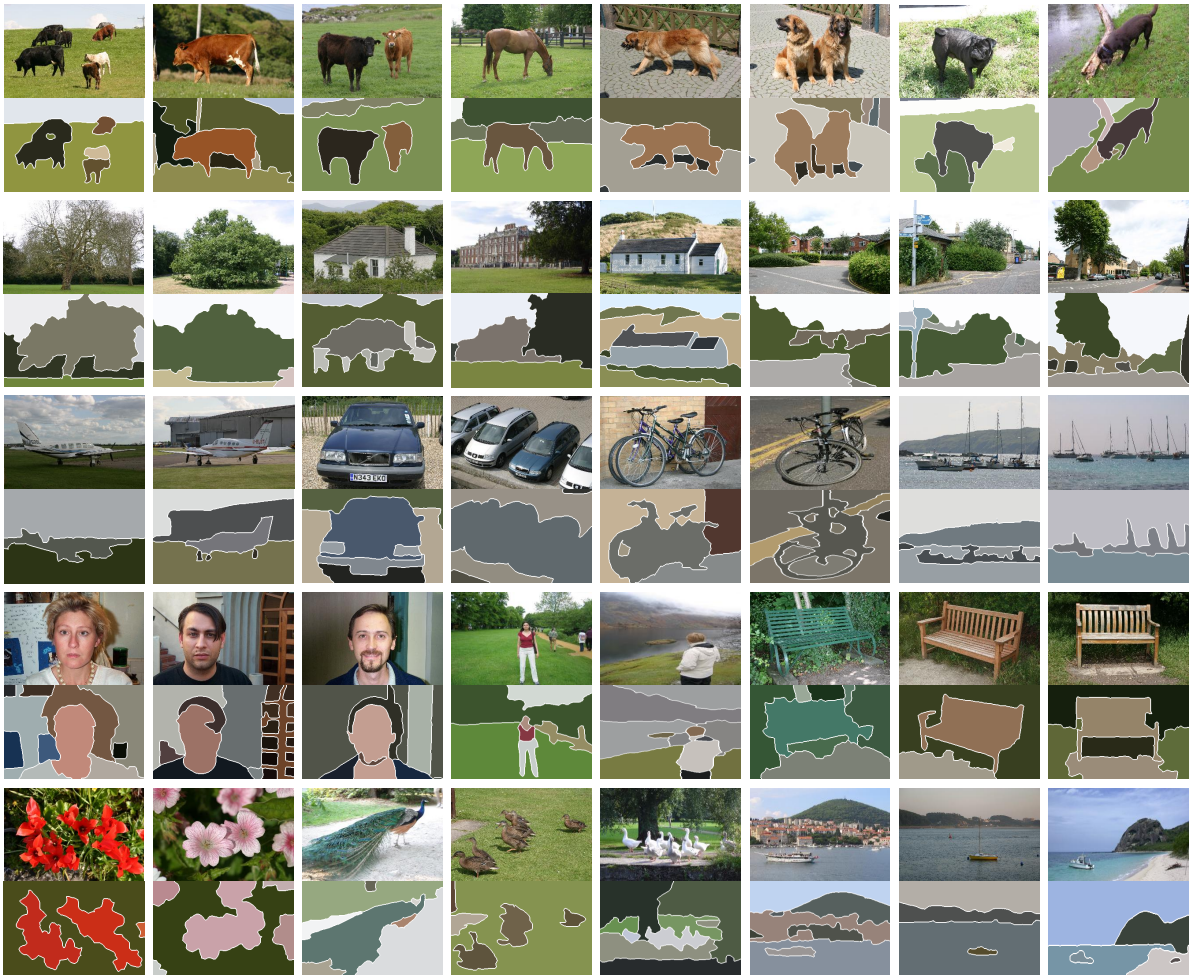

Figure 3. Some examples of the segmentation results on the MSRC database, produced by our MLAP method.

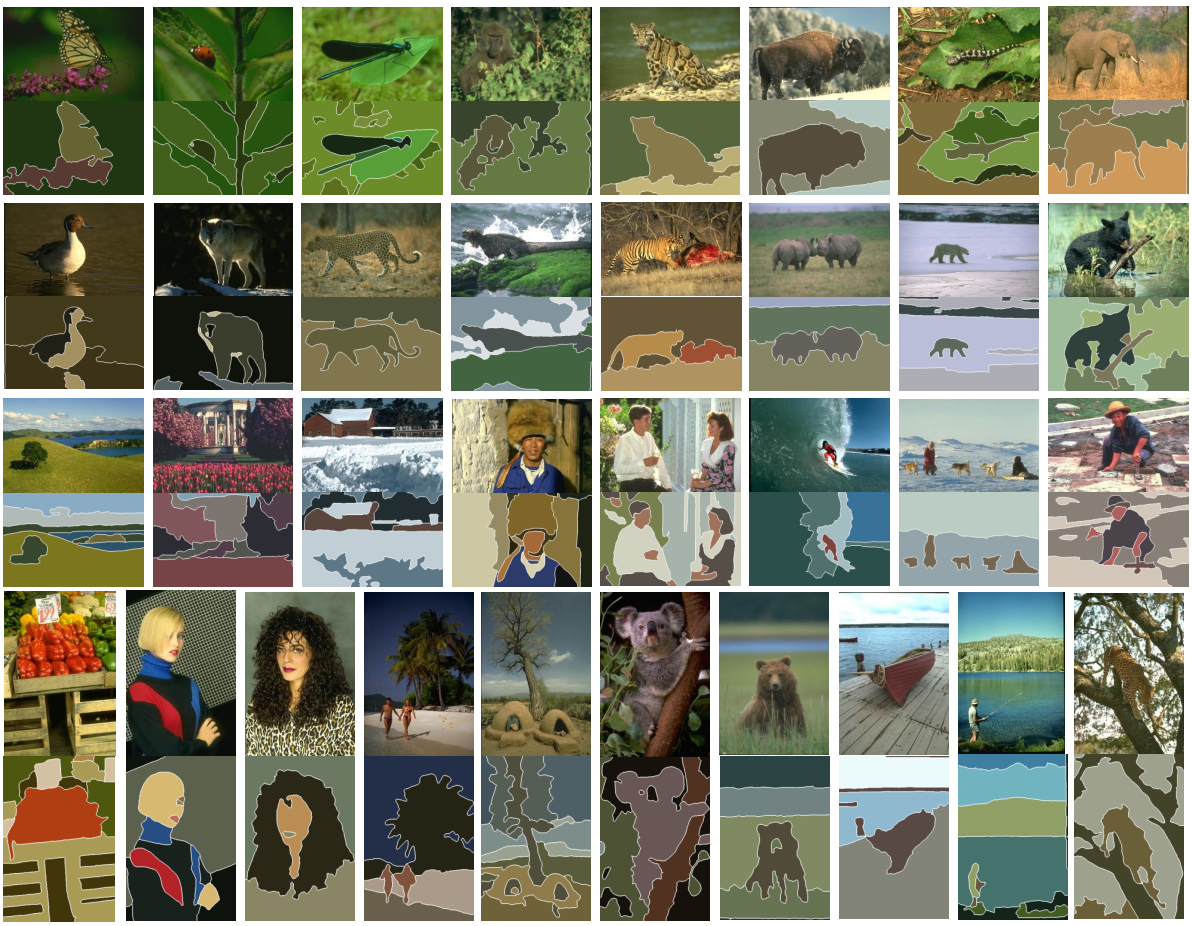

Figure 4. Some examples of the segmentation results on the Berkeley dataset, produced by our MLAP method.

shift. These results illustrate the effectiveness of our formulation (2), which generally learns a unified affinity matrix from multiple feature spaces. Compared with the approach of applying LRR to a certain single feature, again, the results in Table 1 clearly show the advantage of fusing multiple types of features. Figure 3 and Figure 4 show some examples of image segmentation results. It can be seen that the segmentation results produced by MLAP are quite promising.

3.2.2 Comparison to State-of-the-art Methods

To evaluate the competitiveness of the proposed solution, we also compare the results with other state-of-the-art methods, mainly including Rao et al. [27] and Arbelaez et al. [3]. Our method distinctly outperforms the results reported in [27], which achieves a PRI (higher is better) of 0.8 and a VOI (lower is better) of 1.76 on the Berkeley dataset. Whereas, as shown in Table 1, our MLAP method can obtain a PRI of 0.8538 and a VOI of 1.5311. On the Berkeley dataset, our results are better than the results reported from Arbelaez et al. [3] under the optimal dataset scale (VOI=1.69,PRI=0.83 and CR=0.59), and are close to their results under the optimal image scale (VOI=1.48,PRI=0.86 and CR=0.65). On the MSRC database, our results are better than their results obtained under the optimal dataset scale (CR=0.66), and are also slightly better than their optimal image scale results (CR=0.75). These results illustrate that our solution is competitive for image segmentation. It is also worth noting that our current method may be further boosted by integrating other visual cues, e.g., contour and spatial information, as discussed in Section 2.4.2.

4. Conclusions and Future Work

This paper presented a novel image segmentation framework called multi-task low-rank affinity pursuit (MLAP). In contrast with existing single-feature based methods, MLAP integrates the information of multiple types of features into a unified inference procedure, which can be efficiently per-

formed by solving a convex optimization problem. The proposed method seamlessly integrates multiple types of features to collaboratively produce the affinity matrix within a single inference step, and thus produces more accurate and reliable results.

In our current model, we only consider the cue of region, which is only one aspect of visual cues. It is flexible to extend our solution to include other cues, such as contours and spatial information for further boosting performance. Moreover, we shall further consider utilizing object-specific information and extending our current solution to address the problem of object segmentation. We leave these as our future work.

References

[1] P. Arbelaez. Boundary extraction in natural images using ultrametric contour maps. In *CVPR Workshop*, 2006. 1

[2] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. From contour to regions: An empirical evaluation. In *CVPR*, 2009. 5

[3] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *TPAMI, to appear*, 2010. 1, 2, 5, 7

[4] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 2010. 4

[5] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. Huang. Learning with ℓ^1-graph for image analysis. *TIP*, 2010. 3

[6] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *TPAMI*, 2002. 1, 5

[7] T. Cour, F. Benezit, and J. Shi. Spectral segmentation with multiscale graph decomposition. In *CVPR*, 2005. 1

[8] A. Delaunoy, K. Fundana, E. Prados, and A. Heyden. Convex multi-region segmentation on manifolds. In *ICCV*, 2009. 1

[9] P. Felzenszwalb and D. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 2004. 5

[10] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 2004. 1

[11] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí. Yet another survey on image segmentation: Region and boundary information integration. In *ECCV*, 2002. 1

[12] S. Geman and D. Geman. Readings in computer vision: issues, problems, principles, and paradigms. 1987. 1

[13] Z. Lin, M. Chen, L. Wu, and Y. Ma. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. Technical report, UILU-ENG-09-2215, 2009. 2, 4

[14] G. Liu, Z. Lin, X. Tang, and Y. Yu. Unsupervised object segmentation with a hybrid graph model (HGM). *TPAMI*, 2010. 1

[15] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *Preprint*, 2010. 3

[16] G. Liu, Z. Lin, and Y. Yu. Robust subspace segmentation by low-rank representation. In *ICML*, 2010. 3, 4

[17] D. Lowe. Object recognition from local scale-invariant features. In *ICCV*, 1999. 2, 5

[18] P. M. and O. T. Texture analysis in industrial applications. *Image Technology*, 1996. 1

[19] Y. Ma, H. Derksen, W. Hong, and J. Wright. Segmentation of multivariate mixed data via lossy data coding and compression. *TPAMI*, 2007. 1

[20] Y. Ma, A. Yang, H. Derksen, and R. Fossum. Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM Review*, 2008. 3

[21] J. Malik, S. Belongie, J. Shi, and T. Leung. Textons, contours and regions: Cue integration in image segmentation. In *iccv*, 1999. 1, 5

[22] T. Malisiewicz and A. Efros. Improving spatial support for objects via multiple segmentations. In *BMVC*, 2007. 5

[23] M. Meila. Comparing clustering: An axiomatic view. In *ICML*, 2005. 5

[24] G. Mori, X. Ren, A. Efros, and J. Malik. Recovering human body configurations: combining segmentation and recognition. In *CVPR*, 2004. 2, 4, 5

[25] C. Pantofaru, C. Schmid, and M. Hebert. Object recognition by integrating multiple image segmentations. In *eccv*, 2008. 1

[26] W. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 1971. 5

[27] S. Rao, H. Mobahi, A. Y. Yang, S. Sastry, and Y. Ma. Natural image segmentation with adaptive texture and boundary encoding. In *ACCV*, pages 135–146, 2009. 6, 7

[28] X. Ren, C. Fowlkes, and C. Malik. Scale-invariant contour completion using condition random fields. In *ICCV*, 2005. 5

[29] T. Schoenemann, F. Kahl, and D. Cremers. Curvature regularity for region-based image segmentation and inpainting: A linear programming relaxation. In *ICCV*, 2009. 1

[30] G. Sfikas, C. Nikou, and N. P. Galatsanos. Edge preserving spatially varying mixtures for image segmentation. In *CVPR*, 2008. 1

[31] J. Shi and J. Malik. Normalized cuts and image segmentation. *TPAMI*, 2000. 1, 2, 3

[32] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *CVPR*, 2008. 1

[33] J. Shotton, J. Winn, C. Rother, and A. Criminisi. TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *ECCV*, 2006. 2, 5

[34] Z. Tu and S.-C. Zhu. Image segmentation by data-driven markov chain monte carlo. *TPAMI*, 2002. 1

[35] J. Wang, Y. Jia, X.-S. Hua, C. Zhang, and L. Quan. Normalized tree partitioning for image segmentation. In *CVPR*, 2008. 1

[36] M. Wertheimer. Laws of organization in perceptual forms. *A Sourcebook of Gestalt Psycychology*, 1938. 1

[37] Y. Zhang. Recent advances in alternating direction methods: Practice and theory. *Tutorial*, 2010. 4

[38] D. Zhou and C. Burges. Spectral clustering and transductive learning with multiple views. In *ICML*, 2007. 2, 3