Normalized Cut Meets MRF

Meng Tang[1], Dmitrii Marin[1(✉)], Ismail Ben Ayed[2], and Yuri Boykov[1]

[1] Computer Science, University of Western Ontario, London, Canada
{mtang73,yuri}@csd.uwo.ca, dmitrii.a.marin@gmail.com
[2] Ecole de Technologie Supérieure, University of Quebec, Montreal, Canada
ismail.benayed@etsmtl.ca

Abstract. We propose a new segmentation or clustering model that combines Markov Random Field (MRF) and Normalized Cut (NC) objectives. Both NC and MRF models are widely used in machine learning and computer vision, but they were not combined before due to significant differences in the corresponding optimization, *e.g.* spectral relaxation and combinatorial max-flow techniques. On the one hand, we show that many common applications for multi-label MRF segmentation energies can benefit from a high-order NC term, *e.g.* enforcing balanced clustering of arbitrary high-dimensional image features combining color, texture, location, depth, motion, etc. On the other hand, standard NC applications benefit from an inclusion of common pairwise or higher-order MRF constraints, *e.g.* edge alignment, bin-consistency, label cost, etc. To address NC+MRF energy, we propose two efficient multi-label combinatorial optimization techniques, *spectral cut* and *kernel cut*, using new unary bounds for different NC formulations.

1 Introduction

Let Ω be a collection of pixels/voxels or any other data points p that may also be referred to as graph nodes. A segmentation can be equivalently represented either as a labeling $S := (S_p | p \in \Omega)$ including integer node labels $1 \leq S_p \leq K$ or as a partitioning $\{S^k\}$ of set Ω into K segments $S^k := \{p \in \Omega | S_p = k\}$. Our general energy formulation combining Normalized Cut clustering and MRF regularization potentials is

$$E(S) \;=\; -\sum_k \frac{assoc(S^k, S^k)}{assoc(\Omega, S^k)} \;+\; \gamma \sum_{c \in \mathcal{F}} E_c(S_c) \tag{1}$$

where the first term is the standard NC energy [1] with *association*

$$assoc(S^i, S^j) := \sum_{p \in S^i, q \in S^j} A_{pq} \;\equiv\; {S^i}' \mathcal{A} S^j \qquad \text{for} \quad 1 \leq i, j \leq K \tag{2}$$

based on *affinity matrix* or *kernel* $\mathcal{A} := [A_{pq}]$ with $A_{pq} := A(f_p, f_q)$ defined by some similarity function $A(\cdot, \cdot)$ for node features f_p. The equivalent matrix expression in (2) represents segments S^k as indicator vectors such that $S^k_p = 1$

B. Leibe et al. (Eds.): ECCV 2016, Part II, LNCS 9906, pp. 748–765, 2016.
DOI: 10.1007/978-3-319-46475-6_46

iff $S_p = k$ and symbol $'$ means a transpose. The second term in (1) is a general formulation of arbitrary MRF *potentials* [2–4]. Constant γ sets a relative weight of this term. Symbol c represent a *factor* or subset of nodes $c \subset \Omega$ and $S_c := (S_p | p \in c)$ is a restriction of labeling S to c. Energy terms or potentials $E_c(S_c)$ for a given set of factors $\mathcal{F}$ represent various forms of second or higher-order constraints. For example, common pair-wise factors represent smoothness and contrast/edge alignment [2,5]. Popular higher-order factors are superpixel/bin consistency [3,6], label cost [4], and many others. Section 2.2 details several standard MRF potentials used in this paper's example.

1.1 Motivation and Related Work

Due to significant differences in applicable optimization algorithms, Normalized Cut (NC) and Markov Random Fields (MRF) techniques are used separately in many applications of vision and learning. They have complementary strengths and weaknesses.

For example, NC can find a balanced partitioning of data points from pairwise affinities for high-dimensional features [1,7,8]. In contrast, discrete MRF as well as continuous regularization methods commonly use *model fitting* to partition image features [4,9–11]. Such *probabilistic K-means* clustering [12] is well justified when data supports low complexity models, e.g. Gaussians [9] or geometric lines/planes [4]. However, data clustering by fitting complex models like GMM or histograms [10,11] is highly sensitive to local minima and over-fitting even for low dimensional color features [6]. A similar point is made in [13] comparing [11] to a binary energy combining the Potts model and *average association*. Our multi-label energy (1) allows a general MRF framework to benefit from widely-known NC balanced clustering of high-dimensional image features. We show potent results for basic formulations of NC+MRF segmentation with features like RGBXY, RGBD, RGBM where standard MRF methods fail.

On the other hand, standard NC applications can also benefit from an inclusion of additional constraints [14–16]. We show how to add a wide class of standard MRF potentials. For example, standard NC segmentation has weak alignment to contrast edges [8]. While this can be addressed by post-processing, inclusion of the standard pair-wise Potts term [2,5] offers a principled solution. We show benefits from combining NC with lower and higher-order constraints, such as sparsity or label costs [4]. In the context of a general graph clustering, higher-order consistency terms based on a P^n-Potts model [3] also give significant improvements.

The synergy of the general NC+MRF segmentation energy (1) can be illustrated by juxtaposing the use of the pixel location information (XY) in standard NC and MRF techniques. The basic pairwise MRF Potts model for images typically works on the nearest-neighbor grids $\mathcal{N}_4$ or $\mathcal{N}_8$ where XY information allows accurate contrast edge alignment and enforces "smooth" segment boundaries. Wider connectivity Potts leads to denser graphs with slower optimization and poorer edge localization. In contrast, common NC methods [1] augment pixel features, *e.g.* color, with XY information using relatively wide

kernels for the XY dimension. This encourages segments with spatially "compact" regions. Narrower XY kernels may improve edge alignment [8], but weaken regional color/feature consistency. On the other hand, an extremely large XY kernels ignore spatial information producing color-only clustering with incoherent segments. Combining regional color consistency with spatial coherence in a single NC energy requires a compromise XY kernel width. Our general energy (1) can separate the regional consistency (*e.g.* balanced NC clustering term) from the boundary smoothness or edge alignment (*e.g.* Potts potential). Interestingly, it may still be useful to augment colors with XY in the NC term in (1) since given width XY kernel can separate similar appearance objects at larger distances, see Sect. 3.2.1.

Our experiments (Fig. 2) also show that the standard pairwise edge-alignment MRF (Potts) term may significantly improve the energy of the NC term compared to its independent optimization via spectral relaxation. This suggests that powerful combinatorial graph cut methods may reduce sensitivity of NC to local minima.

1.2 Summary of Contributions

This paper proposes a new joint NC+MRF model (1) for multi-label image segmentation and general clustering, efficient move-making bound optimization algorithms, and demonstrates many useful applications. Our main contributions are outlined below:

- We propose a general multi-label segmentation or clustering energy combining Normalized Cut (NC) objective with standard second or higher-order MRF regularization potentials. NC term can enforce balanced partitioning of observed image features and MRF terms can enforce many standard regularization constraints.
- We obtain *kernel* (exact) and *spectral* (approximate) bounds for NC providing two *auxiliary functions* for our joint multi-label NC+MRF energy (1). In the context of standard MRF potentials (*e.g.* Potts, robust P^n-Potts, label cost) we propose move-making algorithms exploring new generalizations of α-expansions and $\alpha\beta$-swap designed for multi-label bound optimization.[1]
- Our experiments demonstrate that typical NC applications benefit from extra MRF constraints, as well as, MRF segmentation benefit from the high-order NC term encouraging balanced partitioning of image features. In particular, NC+MRF framework works for higher-dimensional image features (*e.g.* RGBXY, RGBD, RGBM) where standard model-fitting clustering [4,10,11] fails.

[1] Our kernel and spectral bounds for NC can be also integrated into auxiliary functions with other standard regularization potentials (truncated, cardinality, TV) addressed by discrete (*e.g.* message passing, relaxations) or continuous (*e.g.* convex, primal-dual) algorithms.

The rest of the paper is organized as follows. Section 2 presents our spectral and kernel bounds for (1) and details combinatorial move making graph cut algorithms for its optimization. Section 3 presents many proof-of-the-concept experiments where NC benefits from the additional MRF constraints, Sect. 3.1, and common MRF formulations benefit from an additional balanced NC clustering term for high-dimensional features, Sect. 3.2.

2 Our Algorithms

In this section we propose *bound optimization* and *move-making* algorithms for our high-order NC+MRF functional (1). In particular, we derive bounds or *auxiliary functions* for our problem[2]. Bound optimization have recently led to competitive algorithms for different high-order binary segmentation functionals, *e.g.* distribution-matching constraints [17], entropy, or non-submodular pairwise energies [18]. These greedy procedures iteratively minimize a sequence of *auxiliary functions* for a given energy $E(S)$ assuming that they are easier to optimize. An auxiliary function $a_t(S)$ at iteration t is an upper bound for $E(S)$ touching the original energy at the current solution S_t

$$a_t(S) \geq E(S), \qquad E(S_t) = a_t(S_t).$$

To decrease $E(S)$ one can minimize the auxiliary function a_t giving the next solution

$$S_{t+1} = \arg\min_S a_t(S).$$

This iterative process guarantees the original energy decrease: $E(S_{t+1}) \leq a_t(S_{t+1}) \leq a_t(S_t) = E(S_t)$. Note that the bound does not have to be optimized globally. As long as $a_t(S_{t+1}) \leq a_t(S_t)$, the original energy is guaranteed to decrease $E(S_{t+1}) \leq E(S_t)$.

2.1 Unary Bounds for Normalized Cut

Below we derive two bounds for (1) and propose move-making algorithms such as *expansion* and *swap* [2] to optimize such multi-label auxiliary functions for our NC+MRF energy. To the best of our knowledge, this is the first use of move-making algorithms in the context of bound optimization. The existing high-order bound-optimization graph cut techniques apply to binary segmentation [17,18]. The computation aspects of evaluating our bounds for large-scale problems are discussed in Sec.2.1.

Our first bound, called *kernel bound*, is an exact auxiliary function for the high-order energy (1). It is expressed as a function of the pairwise affinities or kernels. Our second bound, called *spectral bound*, can be viewed as an auxiliary function for the K-means discretization step in standard spectral relaxation

[2] The MRF terms in (1) preclude the direct use of standard spectral methods for NC [1].

methods [1,19]. Unlike our kernel bound, this approximate bound requires eigen vector computations.

The next lemma helps to derive our kernel bound for energy (1) in Proposition 1.

Lemma 1 (Concavity). *Equivalently rewrite the first (NC) term in* (1) *as*

$$-\sum_k \frac{assoc(S^k, S^k)}{assoc(\Omega, S^k)} \equiv \sum_k e(S^k) \qquad for \quad e(X) := -\frac{X'\mathcal{A}X}{d'X} \tag{3}$$

using the matrix notation in (2) *with affinity matrix* $\mathcal{A} := [A_{pq}]$ *and vector* $d := \mathcal{A}\mathbf{1}$ *of node degrees* $d_p = \sum_q A_{pq}$. *Function* $e : \mathbf{R}^{|\Omega|} \to \mathbf{R}$ *in* (3) *is concave over region* $d'X > 0$ *assuming that affinity matrix* $\mathcal{A}$ *is positive semi-definite.*

Proof. It follows from negative definiteness of the Hessian $\nabla\nabla e$, see [20, Lemma 1]. □

The first-order Taylor expansion $T_t(X) := e(X_t) + \nabla e(X_t)'\,(X - X_t)$ at a current solution X_t is an obvious bound for the concave function $e(X)$ in (3). Its gradient $\nabla e(X_t) = d\,\frac{X_t'\mathcal{A}X_t}{(d'X_t)^2} - \mathcal{A}X_t\frac{2}{d'X_t}$ implies bound $T_t(X) \equiv \nabla e(X_t)'\,X$ and Proposition 1.

Proposition 1 (Kernel Bound). *For positive semi-definite affinity matrix* $\mathcal{A}$ *and any current solution* S_t *the following is an auxiliary function for NC+MRF energy* (1)

$$a_t(S) \;=\; \sum_k \nabla e(S_t^k)'\,S^k \;+\; \gamma \sum_{c\in\mathcal{F}} E_c(S_c). \tag{4}$$

Our kernel bound (4) for energy (1) combines a unary term and standard MRF regularization potentials. This allows to develop efficient move-making graph cut techniques for the general multi-label NC+MRF energy (1), see Sect. 2.2. Interestingly, for the degenerate case of (1) with no MRF potentials ($\gamma = 0$) iterative optimization of the first unary term in bound (4) can be shown to be equivalent to the *weighted kernel K-means* approach to NC in [21,22], see Sect.1.3.1 and Appendix A in [20]. Note that NC with arbitrary affinity $\mathcal{A}$ can be converted to an equivalent NC objective with p.s.d. affinity $\mathcal{A} + \delta D$ where $D := diag(d)$ is a *degree matrix*, see [20, Sect.1.3.1]. Such *diagonal shift* tricks were proposed by [23,24].

We also develop an approximate *spectral bound* for our NC+MRF energy. Note that standard spectral methods [1,19] optimize the normalized cut objective by relaxing the original integer problem to a (generalized) eigenvalue problem of the form:

$$(D - \mathcal{A})\mathbf{u} = \lambda D\mathbf{u}. \tag{5}$$

Let U denote the matrix whose rows are the (unit) eigenvectors of the eigen system (5) and U^K be a matrix whose rows are the K top (unit) eigenvectors in U. To extract integer labeling from the relaxed solutions produced by (5), spectral methods often apply the basic K-means to some *ad hoc* data embedding

$\phi_p \equiv \phi(f_p)$ based on U^K. For instance, [1,19] use the columns of U^K, *i.e.* $\phi_p = U^K_p$, while [7,25] use a weighted version $\phi_p = [\Sigma^{-\frac{1}{2}} U]^K_p$ where Σ is a diagonal matrix of eigenvalues in (5). We did not observe much difference in practice.

Similar embeddings can also be derived in a principled fashion. Consider kernel $\mathcal{K} = D^{-1}\mathcal{A}D^{-1}$ for weighted ($w = d$) *kernel K-means* [22] equivalent to NC with affinity $\mathcal{A}$. Similarly to *multi-dimensional scaling* (MDS) [26], we obtain rank-m approximate kernel $\tilde{\mathcal{K}}$ minimizing weighted Frobenius error $\sum_{pq} w_p w_q (\mathcal{K}_{pq} - \tilde{\mathcal{K}}_{pq})^2$ and deduce embedding $\tilde{\phi}_p \in \mathcal{R}^m$ satisfying isometry $\tilde{\phi}'_p \tilde{\phi}_q = \tilde{\mathcal{K}}_{pq}$, see details in [20, Sect.3]

$$\tilde{\phi}_p \;=\; \sqrt{\Lambda^m / d_p} V^m_p \qquad \text{for eigen decomposition } V'\Lambda V = D^{-\frac{1}{2}}\mathcal{A}D^{-\frac{1}{2}}. \tag{6}$$

Weighted K-means over $\{\tilde{\phi}_p\}$ corresponds to *kernel K-means* for kernel $\tilde{\mathcal{K}} \approx \mathcal{K}$ and, therefore, approximates NC. We can combine weighted K-means over $\{\tilde{\phi}_p\}$ with MRF regularization via minimizing the following *spectral approximation* of (1)

$$\tilde{E}(S) \;=\; F^w(S, \mu_S) + \;\gamma\; \sum_{c \in \mathcal{F}} E_c(S_c) \tag{7}$$

where $F^w(S,m) := \sum_k \sum_{p \in S^k} w_p \|\tilde{\phi}_p - m_k\|^2$ includes variable $m = \{m_k\}_{k=1}^K$ representing weighted segment means $m_k = \mu^w_{S^k} := \frac{\sum_{p \in S^k} w_p \tilde{\phi}_p}{w' S^k}$ in (7).

Proposition 2 (Spectral Bound). *The following is an auxiliary function for* (7) *at current segmentation* S_t *with the corresponding means* $\mu^w_t := \{\mu^w_{S^k_t}\}_{k=1}^K$

$$\tilde{a}_t(S) \;=\; F^w(S, \mu^w_t) + \;\gamma\; \sum_{c \in \mathcal{F}} E_c(S_c). \tag{8}$$

Proof. This bound follows from a simple fact that the standard block-coordinate descent (weighted) K-means procedure is a bound optimizer, *e.g.* see [20, Theorem 1]. □

Bound Evaluation for Large-Scale Problems. Our kernel bound does not require eigen decomposition. The time complexity of evaluating this bound is linear with respect to the number of non-zero entries in affinity (kernel) matrix $\mathcal{A}$. Sampling heuristics [27] can be used to derive an approximate bound efficiently for large scale problems.

Obtaining the spectral bound in Proposition 2 for given ϕ_p has complexity $\mathcal{O}(|\Omega| K m)$. The eigen decomposition is also computationally expensive. Standard methods like [28] can substantially accelerate it allowing applications to large scale problems.

2.2 Move-Making NC+MRF Bound Optimization

As mentioned in the introduction, second and higher-order MRF potentials are widely used for regularization in computer vision. We demonstrate a combined

NC+MRF energy (1) in the context of several common MRF potentials outlined below and propose combinatorial bound optimization algorithms using unary bounds for NC in the previous section. We observe that many standard discrete optimization methods [2,29–31][3] can be developed to work with such unary/linear bounds for NC. For simplicity, we focus on three standard MRF potentials below allowing efficient move-making graph cut techniques for multi-label NC+MRF bound optimization.

Input : Affinity Matrix $\mathcal{A}$ of size $|\Omega| \times |\Omega|$; initial labeling $S_0^1, ..., S_0^K$
Output: $S^1, ..., S^K$: partition of the set Ω
Set $t := 0$;
while *not converged* **do**
  Set $a_t(S)$ to be kernel bound (4) for NC at current partition S_t;
  for *each label* $\alpha \in \mathcal{L} = \{1, ..., K\}$ **do**
    Find $S_t := \arg\min a_t(S)$ within one α expansion of S_t;
  Set $t := t + 1$;

Algorithm 1. α-Expansion for Kernel Cut

Input : Affinity Matrix $\mathcal{A}$ of size $|\Omega| \times |\Omega|$; initial labeling $S_0^1, ..., S_0^K$
Output: $S^1, ..., S^K$: partition of the set Ω
Find top m eigen values/vectors Λ^m, V^m for matrix $D^{-\frac{1}{2}}\mathcal{A}D^{-\frac{1}{2}}$;
Compute embedding $\tilde{\phi}_p$ in (6) and set $t := 0$;
while *not converged* **do**
  Set $\tilde{a}_t(S)$ to be the spectral bound (8) at current partition S_t;
  for *each label* $\alpha \in \mathcal{L} = \{1, ..., K\}$ **do**
    Find $S_t := \arg\min \tilde{a}_t(S)$ within one α expansion of S_t;
  Set $t := t + 1$;

Algorithm 2. α-Expansion for Spectral Cut

Probably the most common MRF regularization potential corresponds to the second-order Potts model [2] used for **edge alignment**

$$\sum_{c \in \mathcal{F}} E_c(S_c) \quad = \quad \sum_{pq \in \mathcal{N}} w_{pq} \cdot [S_p \neq S_q] \tag{9}$$

where a set of pairwise factors $\mathcal{F} = \mathcal{N}$ includes all *edges* $c = \{pq\}$ between pairs of neighboring nodes. Weight w_{pq} defines a penalty for discontinuity between p and q. It could be a constant or may be set by a decreasing function of intensity difference $I_p - I_q$ in order to attract the segmentation boundary to the contrast edges in the image [5].

A useful **bin consistency** constraint corresponds to the P^n-Potts model [3] defined over an arbitrary collection of high-order factors $\mathcal{F}$. Factors $c \in \mathcal{F}$ correspond to predefined subsets of nodes such as *superpixels* [3] or *bins* of pixels with the same color/feature [6,35]. The model penalizes inconsistency in segmentation of each factor

$$\sum_{c \in \mathcal{F}} E_c(S_c) \quad = \quad \sum_{c \in \mathcal{F}} \min\{T, |c| - |S_c|^*\} \tag{10}$$

[3] As well as related continuous methods, *e.g.* [32–34].

where T is some threshold and $|S_c|^* := \max_k |S^k \cap c|$ is the cardinality of the largest segment inside c. Clearly, potential (10) has its lowest value (zero) when all nodes in each factor are within the same segment. In this paper we use such MRF potential for NC-based image collection clustering to enforce *text-tag consistency*.

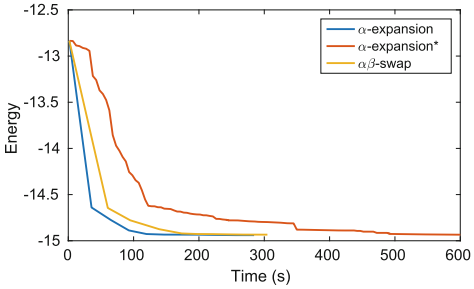
A standard label cost corresponds to an MDL **sparsity potential** [4] defined for a single high-order factor, the whole set of nodes. That is, $\mathcal{F} = \{\Omega\}$. In its simplest form this potential penalizes the number of distinct labels supported by segmentation S

$$\mathbf{E}_\Omega(S) \quad = \quad \sum_k h_k \cdot [|S^k| > 0] \tag{11}$$

where penalty h_k could be a constant or a cost for each specific segment type (label).

All three MRF potentials reviewed above can be optimized by $\alpha\beta$-swap moves and α-expansion moves. We combine these MRF terms with unary bounds for NC (Sect. 2.1). Our kernel cut and spectral cut methods are outlined in Algorithms 1 and 2.

One should decide the order of iterative move making and bound evaluation. In the case of α-expansion, there are at least three options: updating the bound after single expansion step, or after single expansion loop, or after the convergence of α-expansion. More frequent bound recalculation slows down the algorithm, but makes the bound tighter. The particular choice generally depends on the trade-off between the speed and solution quality. However, in our experiments more frequent update does not always improve the energy, see Fig. 1. We recommend updating the bound after a single loop of expansion, see Algorithms 1 and 2. Further by replacing expansion with swap loops in the two algorithm, we can have $\alpha\beta$-swap based version of kernel cut and spectral cut.



(a) Versions of Kernel Cut

Compare	against	# of wins	p-value†
α-expansion	$\alpha\beta$-swap	135/200	10^{-6}
α-expansion	α-expansion*	182/200‡	10^{-34}‡

† The probability to exceed a given number of wins by random chance.

‡ The algorithm terminated due to time limit. This may cause incorrect number of wins.

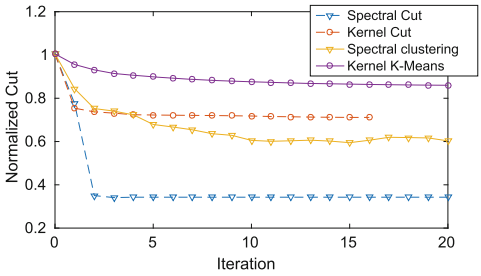
(b) BSDS500 training dataset

Fig. 1. Typical energy evolution wrt different moves and frequency of bound updates. *α-expansion* updates the bound after a round of expansions, *α-expansion** updates the bound after each expansion move. Initialization is a regular 5×5 grid of patches.

3 Experiments

This section is divided into two parts. The first part (Sect. 3.1) shows the benefits of extra MRF regularization for the normalized cut criterion. We consider pairwise Potts, label cost and robust bin consistency term, as discussed in Sect. 2.2. We compare to spectral clustering [1,7] and kernel K-means [22], which can be seen as degenerated versions for spectral and kernel cuts (respectively) without MRF terms. We show that MRF helps normalized cut in segmentation and image clustering. In the second part (Sect. 3.2) we replace the log-likelihoods in model-fitting methods, *e.g.* GrabCut [11], by NC term. This is particularly advantageous for high dimension features (location, depth, motion).

Implementation Details: The parameters of the algorithms were selected to minimize the average error over datasets. For segmentation we use affinities A_{pq} defined by *KNN* (K-Nearest-Neighbours) graph due to limitations of fixed-width Gaussian kernel for multi-scale data [36]. Assuming $KNN(f_p)$ is the set of K-nearest-neighbors of feature vector f_p, the *KNN* affinity is $A_{pq} = A(f_p, f_q) = [f_p \in KNN(f_q)] + [f_q \in KNN(f_p)]$. The feature f_p can be concatenation of RGB (color), XY (location) and M (motion or optical flow [37]). We choose 400 neighbors and randomly sample 50 neighbors for each pixel. Sampling does not degrade our segmentation but expedites bound evaluation. We also experiment with popular *mPb* contour based affinities [8] for segmentation. The window radius is set to 5 pixels. For image clustering, we extract GIST [38] feature and use Gaussian kernel to build a dense affinities matrix. For GrabCut we use histograms as models and try various bin size for spatial and depth channels.



BSDS500 training dataset

Compare	against	# of wins	p-value*
Spectral Cut	Spectral clustering	184/200	10^{-15}
Kernel Cut	Kernel K-Means	200/200	10^{-60}

* The probability to exceed a given number of wins by random chance.

Fig. 2. Left: typical evolution of the NC term during iterative optimization. Right: adding MRF regularization helps both Spectral and Kernel clustering to escape local minima and to achieve lower value of NC. Initialization is a regular 5×5 grid of patches.

3.1 MRF Helps Normalized Cut

Here we add MRF regulation terms to typical normalized cut applications, such as unsupervised multi-label segmentation [8] and image clustering [39]. Our kernel and spectral cuts are used to optimize the joint energy of normalized cut and MRF (1) or (7).

3.1.1 Normalized Cut with Potts Regularization. Spectral clustering [1,7] solves a (generalized) eigen problem, followed by K-means on the (weighted) eigenvectors. However, it was observed that such paradigm results in undesirable segmentation in large uniform regions [7,8], see examples in Fig. 3. Obviously such edge mis-alignment can be penalized by contrast-sensitive Potts term. Our spectral cut and kernel cut get better segmentation boundaries. As is in [22] we use spectral initialization.

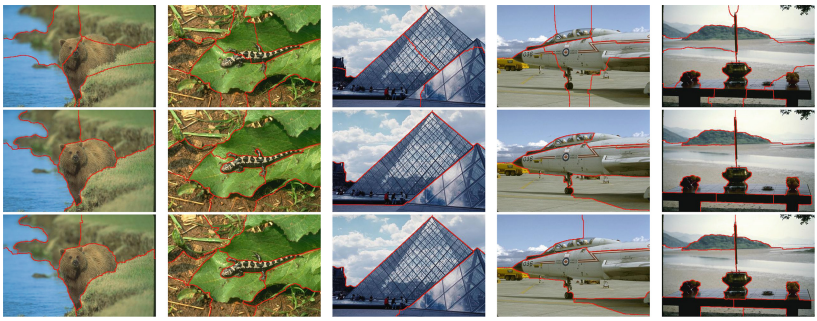


Fig. 3. Sample results on BSDS500 with mPb affinities [8]. Top row: Spectral Clustering; Middle and bottom rows: Kernel and Spectral Cuts giving better edge alignment.

The table on the right gives quantitative results on BSDS500 dataset. Number of segments in ground truth is provided to each method. It also shows that kernel and spectral cuts give better covering, PRI (probabilistic rand index) and VOI (variation of information) than spectral clustering. Figure 3 gives sample results. Kernel K-means [22] gives results similar to spectral clustering and hence are not shown.

method	Covering	PRI	VOI
Spectral relax	0.34	0.76	2.76
Our kernel cut	0.41	**0.78**	2.44
Our spectral cut	**0.42**	**0.78**	**2.34**

3.1.2 Normalized Cut with Label Cost [4]. Unlike spectral clustering, our kernel and spectral cuts do not need the number of segments beforehand. We optimize a combination of the normalized cut, Potts model and label costs terms. The label cost (11) penalizes each label by constant h_k. The energy is minimized by α-expansion and $\alpha\beta$-swap moves in Sect. 2.2. We sample initial models from patches, as in [4]. Results with different label cost are shown in Fig. 4. Due to sparsity prior for normalized cut, our kernel and spectral cuts automatically prune *weak* models and determine the number of segments, yet yield regularized segmentation. We use *KNN* affinity for normalized cut and mPb [8] based Potts regularization.

3.1.3 Normalized Cut with High Order Consistency Term [3,6,35]. It is common that images come with multiple tags, such as those in Flickr platform or the LabelMe dataset [38]. We study how to utilize tag-based group prior for image clustering [39].

Fig. 4. Segmentation using our kernel cut with label cost. We experiment with increasing value of label cost h_k for each label (from left to right).

We experiment on the LabelMe dataset [38] which contains 2,600 images of 8 scene categories (coast, mountain, forest, open country, street, inside city, tall buildings and highways). We use the same GIST feature, affinity matrix and group prior as used in [39]. We found the group prior to be noisy. The dominant category in each group occupies only 60 %–90 % of the group. The high-order consistency term is defined on each group. For each group, we introduce an energy term that is akin to the *robust* P^n-Potts [3], which can be exactly minimized within a single $\alpha\beta$-swap or α-expansion move. Notice that here we have to use robust consistency potential instead of rigid ones.

Our kernel cut minimizes NC plus the *robust* P^n-Potts term. Spectral cut minimizes energy of (7). Normalized mutual information (NMI) is used as the measure of clustering quality. Perfect clustering with respect to ground truth has NMI value of 1.

Spectral clustering and kernel K-means [22] give NMI value of 0.542 and 0.572 respectively. Our kernel cut and spectral cut significantly boost the NMI to 0.683 and 0.681. Figure 5 shows the results with respect to different amount of image tags used. The left most points correspond to the case when no group prior is given. We optimize over the weight of high order consistency term, see

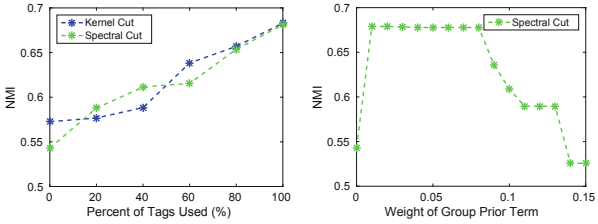


Fig. 5. Incorporating group prior achieves better NMI for image clustering. Here we use tags-based group prior. Our method achieved better NMI when more images are tagged. The right plot shows how the weight of bin consistency term affects our method.

Fig. 5. Note that it's not the case the larger the weight the better since the grouping prior is noisy.

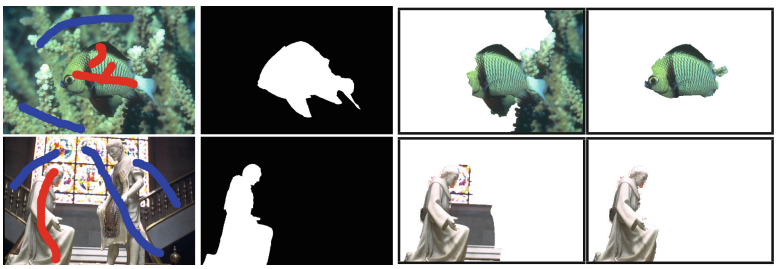
3.2 Normalized Cut Helps MRF

In typical MRF applications we replace the log-likelihood terms by the normalized cut. We test various applications including separating similar objects, RGBD and motion segmentation.

3.2.1 Similar Objects Separation. Even though objects may have similar appearances or look similar to the background (*e.g.* the top row in Fig. 6), we assume that the objects of interest are compact and have different locations. This assumption motivates using XY coordinates of pixels as extra feature for distinguishing similar or camouflaged objects[4]. Let $f_p \in \mathcal{R}^5$ be the augmented color-location feature $f_p = [L_p, a_p, b_p, \beta x_p, \beta y_p]$ at pixel p where $[L_p, a_p, b_p]$ is its color, $[x_p, y_p]$ are its image coordinates, and β is a scaling parameter.

Note that the edge-based Potts model [5] also uses the XY information. Location features in the clustering and regularization terms have complementary effect: the former solves appearance camouflage while the latter gets edge alignment. We report quantitative results on 18 images with similar objects and camouflage selected from the Berkeley database [41]. We set strokes to select one of the objects, see Fig. 6.

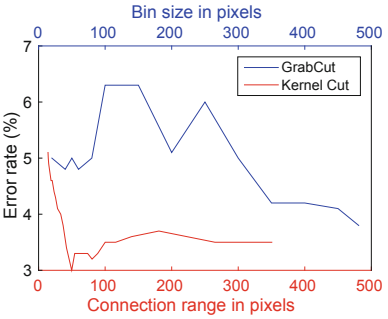
We test the effect of adding XY into feature space for GrabCut and Kernel Cut. We try various β for Kernel Cut. Figure 7(a) shows the effect of different β on *KNN*s of a pixel. For histogram-based GrabCut we change spatial bin size for the XY channel, ranging from 30 pixels to the image size. Figure 7(a) compares GrabCut and Kernel Cut.



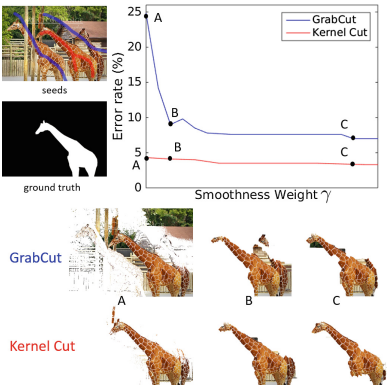
(a) seeds (b) ground truth (c) GrabCut (d) Our Kernel Cut

Fig. 6. Sample results on Berkeley dataset.

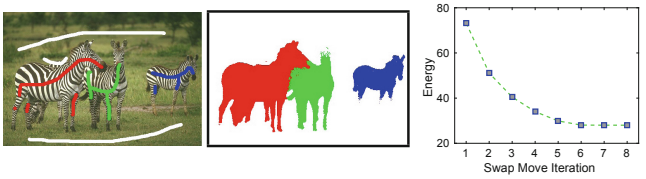
[4] XY feature has also been used in [40] to build space-variant color distribution. However, such distribution used in MRF-MAP inference [40] would still overfit the data [13].



(a) Effect of XY channel weight β



(b) Effect of smoothness weight γ



(c) Sample results of our multi-label kernel cut and energy convergence

Fig. 7. (a) Average errors for multi-object dataset. We vary the spatial bin size for GrabCut and β for Kernel Cut. The connection range is the average geometric distance between a pixel and its k^{th} nearest neighbor. **The right-most point of the curves corresponds to the absence of XY features.** GrabCut does not benefit from XY features. Kernel Cut achieves the best error rate of 2.9 % for the connection range of 50 pixels. **(b)** Our kernel cut is robust to smoothness weight γ. **(c)** Multi-objects segmentation.

We study the effect of MRF smoothness weight γ on the algorithms, see Fig. 7. Kernel Cut is more robust w.r.t. smoothness weight compared to GrabCut. If smoothness term is omitted the Kernel Cut is significantly better (4.6 % vs 24.6 % errors in Fig. 7b). MRF benefits from having NC instead of log-likelihoods since model fitting gets highly sensitive to local minima for higher dimensional features [13]. Figure 7(c) shows multi-label segmentation of similar objects using our algorithm. We show energy convergence for the swap moves discussed in Sect. 2.2.

3.2.2 Interactive RGBD Images Segmentation. Depth sensor are widely used in vision for 3D modelling [44,45], semantic segmentation [42,46–48], motion flow [49]. We selected 64 indoor RGBD images from semantic segmentation database NYUv2 [42] and provided bounding boxes and ground truth. In contrast to [11], the prepared dataset consists of low-quality images: there are camera motion artifacts, underexposed and overexposed regions. Such artifacts make color-based segmentation harder.

We compare GrabCut to Kernel Cut over joint features $f_p = [L_p, a_p, b_p, \beta D_p]$ as in Sect. 3.2.1. Figure 8 shows the error statistics and segmentation examples. While Kernel Cut takes advantage of the additional channel, GrabCut fails to improve.

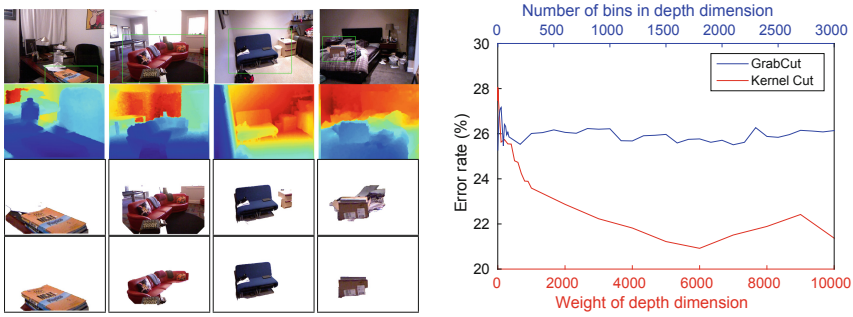


Fig. 8. NYUv2 database [42] **Left:** The first two rows show original images with bounding box and color-coded depth channel. The third row shows results of Grabcut, the forth row shows results of Kernel Cut. **Right:** The average errors of GrabCut and Kernel Cut methods over 64 images randomly selected and labeled.

3.2.3 Motion Segmentation. Besides location and depth features, we also test segmentation with motion features. Figures 10, 11 and 9 compare motion segmentations using different feature spaces: RGB, XY, M (optical flow) and their combinations (RGBM or RGBXY or RGBXYM). Abbreviation **+XY** means Potts regularization. Here we use kernel cut (Algorithm 1) for the combination of normalized cut with the Potts term.

Challenging Video Examples: For videos in FBMS-59 dataset [50], our algorithm runs on individual frames instead of 3D volume. Segmentation of previous frame initializes the next frame. The strokes are provided *only* for the first frame. We use the optical flow algorithm in [37] to generate M features. Selected frames are shown in Figs. 10 and 11. Instead of tracks from all frames in [51], our segmentation of each frame uses only motion estimation between two consecutive frames. Our approach jointly optimizes normalized cut and Potts model. In contrast, [51]

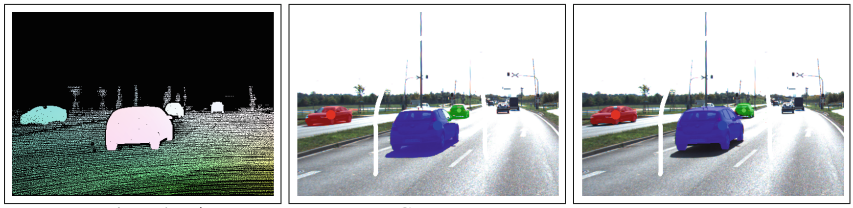
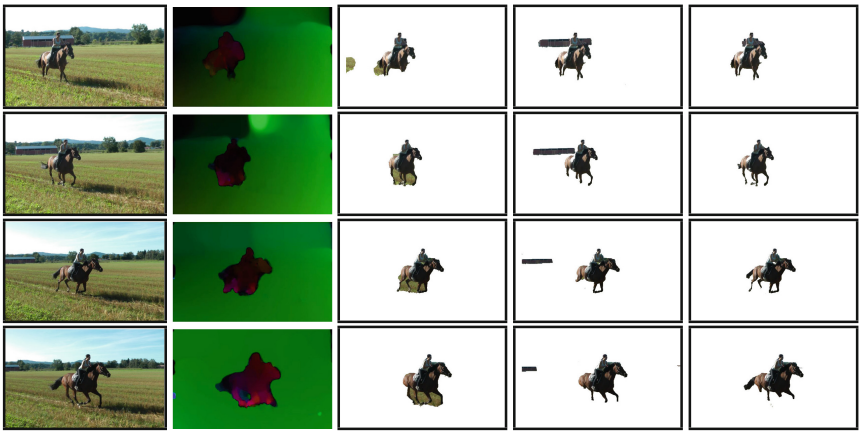
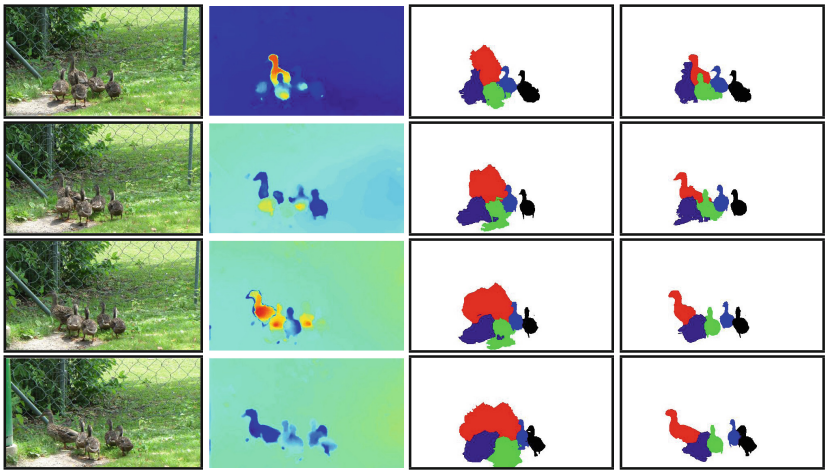


Fig. 9. Kernel cut on image 000079_10 from *KITTI* [43]. The images show the motion flow, color-based segmentation (RGB+XY), motion based segmentation with location features (MXY+XY). * Black color shows pixels that lack motion information.



(a) frames (b) optical flow [37] (c) **M+XY** (d) **RGB+XY** (e) **RGBM+XY**

Fig. 10. Motion segmentation using our framework for the sequence *horses01* in FBMS-59 dataset [50]. **+XY** means with Potts model. Motion feature alone (**M+XY** in (c)) is not sufficient to obtain fine segmentation. Our framework successfully utilize motion feature to separate the horse from the barn, which have similar appearances.



(a) frames (b) optical flow [37] (c) **RGBXY+XY** (d) **RGBXYM+XY**

Fig. 11. Multi-label motion segmentation for the sequence *ducks01* in FBMS-59 dataset [50]. This video is challenging since the ducks here have similar appearances and even spatially overlap with each other. However, different ducks come with different motions, which helps our framework to better separate individual ducks.

first clusters semi-dense tracks via spectral clustering [50] and then obtains dense segmentation via regularization.

Kitti Segmentation Example: We also experiment with Kitti dataset [43]. Figure 9 shows the multi-label segmentation using either color information RGB+XY (first row) or motion MXY+XY (second row). The ground-truth motion field works as M channel. Note that the motion field is known only for approximately 20 % of the pixels. To build an affinity graph, we construct a *KNN* graph from pixels that have motion information. The regularization over 8-neighborhood on the pixel grid interpolates the segmentation labels during the optimization procedure.

References

1. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 888–905 (2000)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **23**(11), 1222–1239 (2001)
3. Kohli, P., Ladicky, L., Torr, P.H.S.: Robust higher order potentials for enforcing label consistency. Int. J. Comput. Vis. (IJCV) **82**(3), 302–324 (2009)
4. Delong, A., Osokin, A., Isack, H., Boykov, Y.: Fast approximate energy minization with label costs. Int. J. Comput. Vis. (IJCV) **96**(1), 1–27 (2012)
5. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: ICCV, vol. I, pp. 105–112 (2001)
6. Tang, M., Gorelick, L., Veksler, O., Boykov, Y.: Grabcut in one cut. In: International Conference on Computer Vision (ICCV), Sydney, Australia, December 2013
7. Malik, J., Belongie, S., Leung, T., Shi, J.: Contour and texture analysis for image segmentation. Int. J. Comput. Vis. **43**(1), 7–27 (2001)
8. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **33**(5), 898–916 (2011)
9. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. Image Process. **10**(2), 266–277 (2001)
10. Zhu, S.C., Yuille, A.: Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. IEEE Trans. PAMI **18**(9), 884–900 (1996)
11. Rother, C., Kolmogorov, V., Blake, A.: Grabcut - interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. (SIGGRAPH) **23**(3), 309–314 (2004)
12. Kearns, M., Mansour, Y., Ng, A.: An information-theoretic analysis of hard and soft assignment methods for clustering. In: Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI), August 1997
13. Tang, M., Ayed, I.B., Marin, D., Boykov, Y.: Secrets of grabcut and kernel k-means. In: International Conference on Computer Vision (ICCV), Santiago, Chile, December 2015
14. Yu, S.X., Shi, J.: Segmentation given partial grouping constraints. IEEE Trans. Pattern Anal. Mach. Intell. **26**(2), 173–183 (2004)
15. Eriksson, A., Olsson, C., Kahl, F.: Normalized cuts revisited: a reformulation for segmentation with linear grouping constraints. J. Math. Imaging Vis. **39**(1), 45–61 (2011)

16. Chew, S.E., Cahill, N.D.: Semi-supervised normalized cuts for image segmentation. In: IEEE International Conference on Computer Vision (ICCV), December 2015
17. Ayed, I.B., Gorelick, L., Boykov, Y.: Auxiliary cuts for general classes of higher order functionals. In: CVPR, pp. 1304–1311 (2013)
18. Tang, M., Ben Ayed, I., Boykov, Y.: Pseudo-bound optimization for binary energies. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part V. LNCS, vol. 8693, pp. 691–707. Springer, Heidelberg (2014)
19. Von Luxburg, U.: A tutorial on spectral clustering. Stat. Comput. **17**(4), 395–416 (2007)
20. Tang, M., Ayed, I.B., Marin, D., Boykov, Y.: Kernel Cuts: MRF meets kernel & spectral clustering. In: (July 2016, also submitted to PAMI). arXiv:1506.07439
21. Bach, F., Jordan, M.: Learning spectral clustering. Adv. Neural Inf. Process. Syst. (NIPS) **16**, 305–312 (2003)
22. Dhillon, I., Guan, Y., Kulis, B.: Kernel k-means, spectral clustering and normalized cuts. In: KDD (2004)
23. Roth, V., Laub, J., Kawanabe, M., Buhmann, J.: Optimal cluster preserving embedding of nonmetric proximity data. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) **25**(12), 1540–1551 (2003)
24. Dhillon, I., Guan, Y., Kulis, B.: Weighted graph cuts without eigenvectors: a multilevel approach. IEEE Trans. Pattern Anal. Mach. Learn. (PAMI) **29**(11), 1944–1957 (2007)
25. Belongie, S., Malik, J.: Finding boundaries in natural images: a new method using point descriptors and area completion. In: Burkhardt, H.-J., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1406, pp. 751–766. Springer, Heidelberg (1998)
26. Cox, T.F., Cox, M.A.: Multidimensional Scaling. CRC Press, Boca Raton (2000)
27. Chitta, R., Jin, R., Havens, T.C., Jain, A.K.: Approximate kernel k-means: solution to large scale kernel clustering. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 895–903. ACM (2011)
28. Fowlkes, C., Belongie, S., Chung, F., Malik, J.: Spectral grouping using the nystrom method. IEEE Trans. Pattern Anal. Mach. Intell. **26**(2), 214–225 (2004)
29. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. IEEE Trans. Pattern Anal. Mach. Intell. **28**(10), 1568–1583 (2006)
30. Werner, T.: A linear programming approach to max-sum problem: a review. IEEE Trans. Pattern Anal. Mach. Intell. **29**(7), 1165–1179 (2007)
31. Kappes, J.H., Andres, B., Hamprecht, F.A., Schnörr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B.X., Kröger, T., Lellmann, J., et al.: A comparative study of modern inference techniques for structured discrete energy minimization problems. Int. J. Comput. Vis. **115**(2), 155–184 (2015)
32. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vis. **20**(1–2), 89–97 (2004)
33. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**(1), 120–145 (2011)
34. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. Int. J. Comput. Vis. **72**(2), 195–215 (2007)
35. Park, K., Gould, S.: On learning higher-order consistency potentials for multi-class pixel labeling. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 202–215. Springer, Heidelberg (2012)

36. Zelnik-Manor, L., Perona, P.: Self-tuning spectral clustering. In: Advances in Neural Information Processing Systems, pp. 1601–1608 (2004)
37. Brox, T., Malik, J.: Large displacement optical flow: descriptor matching in variational motion estimation. IEEE Trans. Pattern Anal. Mach. Intell. **33**(3), 500–513 (2011)
38. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comput. Vis. **42**(3), 145–175 (2001)
39. Collins, M.D., Liu, J., Xu, J., Mukherjee, L., Singh, V.: Spectral clustering with a convex regularizer on millions of images. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part III. LNCS, vol. 8691, pp. 282–298. Springer, Heidelberg (2014)
40. Nieuwenhuis, C., Cremers, D.: Spatially varying color distributions for interactive multilabel segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **35**(5), 1234–1247 (2013)
41. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of Eighth IEEE International Conference on Computer Vision, ICCV 2001, vol. 2, pp. 416–423. IEEE (2001)
42. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from RGBD images. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part V. LNCS, vol. 7576, pp. 746–760. Springer, Heidelberg (2012)
43. Menze, M., Geiger, A.: Object scene flow for autonomous vehicles. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
44. Dou, M., Taylor, J., Fuchs, H., Fitzgibbon, A., Izadi, S.: 3D scanning deformable objects with a single RGBD sensor. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 493–501 (2015)
45. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A.: Kinectfusion: real-time dense surface mapping and tracking. In: 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 127–136. IEEE (2011)
46. Deng, Z., Todorovic, S., Latecki, L.J.: Semantic segmentation of RGBD images with mutex constraints. In: International Conference on Computer Vision (ICCV), Santiago, Chile, December 2015
47. Gulshan, V., Lempitsky, V., Zisserman, A.: Humanising grabcut: learning to segment humans using the kinect. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1127–1133. IEEE (2011)
48. Ren, X., Bo, L., Fox, D.: Rgb-(d) scene labeling: features and algorithms. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2759–2766. IEEE (2012)
49. Gottfried, J.M., Fehr, J., Garbe, C.S.: Computing range flow from multi-modal kinect data. In: Bebis, G., et al. (eds.) ISVC 2011. LNCS, vol. 6938, pp. 758–767. Springer, Heidelberg (2011)
50. Brox, T., Malik, J.: Object segmentation by long term analysis of point trajectories. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 282–295. Springer, Heidelberg (2010)
51. Ochs, P., Brox, T.: Object segmentation in video: a hierarchical variational approach for turning point trajectories into dense regions. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 1583–1590. IEEE (2011)