

# Volumetric Bias in Segmentation and Reconstruction: Secrets and Solutions

Yuri Boykov

Hossam Isack

Carl Olsson

Ismail Ben Ayed

Computer Science

University of Western Ontario, Canada

yuri@csd.uwo.ca habdelka@csd.uwo.ca

Mathematical Sciences

Lund University, Sweden

calle@maths.lth.se

École de Technologie Supérieure

University of Quebec, Canada

ismail.benayed@etsmtl.ca

## Abstract

Many standard optimization methods for segmentation and reconstruction compute ML model estimates for appearance or geometry of segments, e.g. Zhu-Yuille [23], Torr [20], Chan-Vese [6], GrabCut [18], Delong et al. [8]. We observe that the standard likelihood term in these formulations corresponds to a generalized probabilistic K-means energy. In learning it is well known that this energy has a strong bias to clusters of equal size [11], which we express as a penalty for KL divergence from a uniform distribution of cardinalities. However, this volumetric bias has been mostly ignored in computer vision. We demonstrate significant artifacts in standard segmentation and reconstruction methods due to this bias. Moreover, we propose binary and multi-label optimization techniques that either (a) remove this bias or (b) replace it by a KL divergence term for any given target volume distribution. Our general ideas apply to continuous or discrete energy formulations in segmentation, stereo, and other reconstruction problems.

## 1. Introduction

Most problems in computer vision are ill-posed and optimization of regularization functionals is critical for the area. In the last decades the community developed many practical energy functionals and efficient methods for optimizing them. This paper analyses a widely used general class of segmentation energies motivated by Bayesian analysis, discrete graphical models (e.g. MRF/CRF), information theory (e.g. MDL), or continuous geometric formulations. Typical examples in this class of energies include a log-likelihood term for models  $P^k$  assigned to image segments  $S^k$

$$E(S, P) = - \sum_{k=1}^K \sum_{p \in S^k} \log P^k(I_p), \quad (1)$$

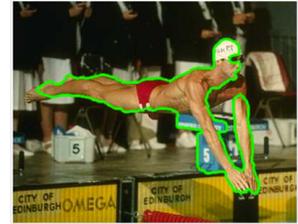
where, for simplicity, we focus on a discrete formulation with data  $I_p$  for a finite set of pixels/features  $p \in \Omega$  and segments  $S^k = \{p \in \Omega | S_p = k\}$  defined by variables/labels

Secrets (1)



(a) GrabCut [18]

Solutions (6), (9-10)



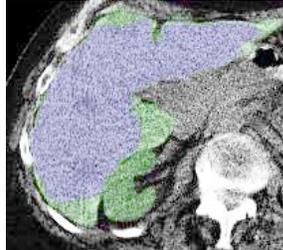
with unbiased data term (10)



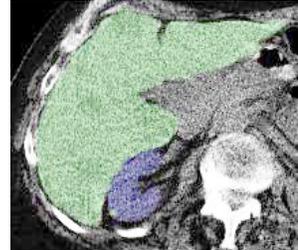
(b) plane fitting [20, 8, 1]



with unbiased data term (10)



(c) Chan-Vese [6] + [7]



with target volumes (6)

Figure 1. *Left*: segmentation and stereo reconstruction with standard likelihoods or *probabilistic K-means* energy  $E(S, P)$  in (1) has bias to equal size segments (2). *Right*: (a-b) corrections due to unbiased data term  $\hat{E}(S, P)$  in (9,10) or (c) weighted likelihoods  $E_W(S, P)$  in (6) biased to proper target volumes, see (5). Sections 3.1-3.3 explain these examples in details.

$S_p \in \mathbf{N}$  indicating the segment index assigned to  $p$ . In different vision problems models  $P^k$  could represent Gaussian intensity models [6], color histograms [2], GMM [23, 18], or geometric models [20, 8, 1] like lines, planes, homographies, or fundamental matrices.

Depending on application, the energies combine likeli-

hoods (1), a.k.a. data term, with different regularization potentials for segments  $S^k$ . One of the most standard regularizers is the Potts potential, as in the following energy

$$E_{Potts}(S, P) = - \sum_{k=1}^K \sum_{p \in S^k} \log P^k(I_p) + \lambda \|\partial S\|,$$

where  $\|\partial S\|$  is the number of label discontinuities between neighboring points  $p$  on a given neighborhood graph or the length of the segmentation boundary in the image grid [3]. Another common regularizer is sparsity or label cost for each model  $P^k$  with non-zero support [20, 23, 1, 8], e.g.

$$E_{sp}(S, P) = - \sum_{k=1}^K \sum_{p \in S^k} \log P^k(I_p) + \gamma \sum_k [S^k \neq \emptyset].$$

In general, energies often combine likelihoods (1) with multiple different regularizers at the same time.

This paper demonstrates practically significant bias to equal size segments in standard energies where models  $P = \{P^k\}$  are variables jointly estimated with segmentation  $S = \{S^k\}$ . This problem comes from likelihood term (1), which we interpret as *probabilistic K-means* energy carefully analyzed in [11] based on the information theory. In particular, [11] decomposes energy (1) as<sup>1</sup>

$$E(S, P) \stackrel{c}{=} \sum_{k=1}^K |S^k| \cdot KL(I^k|P^k) + |\Omega| \cdot (H(S|I) - H(S))$$

where  $KL(I^k|P^k)$  is KL divergence for model  $P^k$  and the true distribution<sup>2</sup> of data  $I^k = \{I_p \mid p \in S^k\}$  in segment  $k$ . Conditional entropy  $H(S|I)$  penalizes “non-deterministic” segmentation if variables  $S_p$  are not completely determined by intensities  $I_p$ . The last term is negative entropy of segmentation variables, which we express via KL divergence

$$-H(S) \stackrel{c}{=} KL(S|U) := \sum_{k=1}^K \frac{|S^k|}{|\Omega|} \log \frac{|S^k|/|\Omega|}{1/K} \quad (2)$$

between the volume distribution for segmentation  $S$

$$V_S := \left\{ \frac{|S^1|}{|\Omega|}, \frac{|S^2|}{|\Omega|}, \dots, \frac{|S^K|}{|\Omega|} \right\} \quad (3)$$

and a uniform distribution  $U := \{\frac{1}{K}, \dots, \frac{1}{K}\}$ . Our KL formulation (2) implies that term  $-H(S)$  represents volumetric bias to equal size segments, as previously noted in [11].

<sup>1</sup>Symbol  $\stackrel{c}{=}$  represents equality up to an additive constant.

<sup>2</sup>The decomposition above applies to either discrete or continuous probability models (e.g. histogram vs. Gaussian). The continuous case relies on Monte-Carlo estimation of the integrals over “true” data density.

## 1.1. Our contributions

Our experiments demonstrate that volumetric bias in probabilistic K-means energy (1) leads to practically significant artifacts for problems in computer vision, where this term is widely used for model fitting in combination with different regularizers, e.g. [23, 20, 6, 18, 8]. Moreover, we propose different approaches addressing this bias in Sec.2.

(A). We propose a simple *weighted* version of the standard likelihood term (1) correcting its equal-size bias to any target distribution of volumes

$$W = \{w^1, w^2, \dots, w^K\} \quad (4)$$

if available. In this case our *weighted likelihoods* (6) have “correct” volumetric bias represented by KL divergence

$$KL(S|W) = \sum_{k=1}^K \frac{|S^k|}{|\Omega|} \log \frac{|S^k|/|\Omega|}{w^k} \quad (5)$$

instead of (2). Correct volumetric bias (5) is implicit for our weighted likelihoods, while practically no changes in optimization are needed compared to standard likelihoods (1). Note that (1) is often combined with a separate volumetric term [9, 17] requiring more advanced optimization. In contrast to our implicit approach, explicit volumetric terms can be freely tuned. Yet, such approaches should also use correctly weighted likelihoods (6) instead of tuning a separate term to overcome wrong bias (2) in standard likelihoods (1).

(B). We propose methods removing the volumetric bias. One standard method uses weighted likelihoods treating weights  $W$  as free variables optimized via block-coordinate descent [11]. We show that this approach is often weaker than the following alternative. We observe that adding extra term  $|\Omega| \cdot H(S)$  to any energy with standard likelihoods (1) also cancels the volumetric bias. Since entropy  $H(S)$  is a concave cardinality function *submodular* for binary problems [15], we can design efficient powerful solvers for our high-order entropy-based correction term: exact one for binary problems and approximate one for multi-label problems using  $\alpha$ -expansion [5]. Similar methods are known for different concave cardinality terms in vision [12, 19].

Our segmentation and stereo tests show that methods (A) and (B) managing volumetric bias (1) improve the results.

## 2. Log-likelihood energy formulations

This section has two goals. First, we present *weighted likelihoods* energy  $E_W(S, P)$  in (6) and derive its volumetric bias  $KL(S|W)$ . Standard data term  $E(P, S)$  in (1) is a special case with  $W = U$ . Then, we present another modification of likelihood energy  $\hat{E}(S, P)$  in (9) and prove that it does not have volumetric bias, as also discussed in [11]. Our analysis of  $\hat{E}$  is needed for completeness and to devise optimization methods for vision problems where likelihoods are only one of the terms in typical regularization energies.

**Weighted likelihoods:** Consider energy

$$E_W(S, P) := - \sum_{k=1}^K \sum_{p \in S^k} \log(w^k P^k(I_p)), \quad (6)$$

motivated by a Bayesian interpretation in [8] where weights  $W$  explicitly come from a volumetric prior. As easy to see

$$\begin{aligned} E_W(S, P) &= E(S, P) - \sum_{k=1}^K |S^k| \log w^k \\ &= E(S, P) + |\Omega| \cdot H(S|W) \end{aligned} \quad (7)$$

where  $H(S|W)$  is a cross entropy between distributions  $V_S$  and  $W$ . As discussed in the introduction, the analysis of probabilistic K-means energy  $E(S, P)$  in [11] implies that

$$\begin{aligned} E_W(S, P) &\stackrel{c}{=} \sum_{k=1}^K |S^k| \cdot KL(I^k|P^k) + |\Omega| \cdot H(S|I) \\ &- |\Omega| \cdot H(S) + |\Omega| \cdot H(S|W). \end{aligned}$$

Combining two terms in the second line gives

$$\begin{aligned} E_W(S, P) &\stackrel{c}{=} \sum_{k=1}^K |S^k| \cdot KL(I^k|P^k) + |\Omega| \cdot H(S|I) \\ &+ |\Omega| \cdot KL(S|W). \end{aligned} \quad (8)$$

In case of given weights  $W$  equation (8) implies that weighted likelihood term (6) has bias to the target volume distribution represented by KL divergence (5).

Note that optimization of weighted likelihood term (6) presents no extra difficulty for regularization methods in vision. Fixed weights  $W$  contribute unary potentials for segmentation variables  $S_p$ , see (7), which are trivial for standard discrete or continuous optimization methods. Nevertheless, examples in Sec. 3 show that indirect minimization of KL divergence (5) substantially improves the results in applications if (approximate) target volumes  $W$  are known.

**Unbiased data term:** If weights  $W$  are treated as unknown parameters in likelihood energy (6) they can be optimized out. In this case decomposition (8) implies that the corresponding energy has no volumetric bias:

$$\begin{aligned} \hat{E}(S, P) &:= \min_W E_W(S, P) \\ &= \sum_{k=1}^K |S^k| \cdot KL(I^k|P^k) + |\Omega| \cdot H(S|I). \end{aligned} \quad (9)$$

Weights  $V_S$  in (3) are ML estimate of  $W$  that minimize (8) by achieving  $KL(S|W) = 0$ . Putting optimal weights  $W = V_S$  into (7) confirms that volumetrically unbiased data term (9) is a combination of standard likelihoods (1) with a

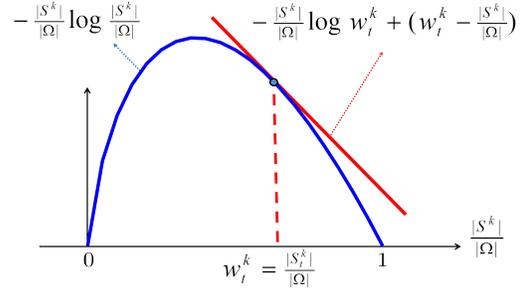


Figure 2. (*Entropy - bound optimization*) According to (7,10) energy  $E_{W_t}(S, P)$  is a bound for  $\hat{E}(S, P)$  since cross entropy  $H(S|W_t)$  is a bound for entropy  $H(S)$  with equality at  $S = S_t$ . This standard fact is easy to check: function  $-z \log z$  (blue curve) is concave and its 1st-order approximation at  $z_t = w_t^k = |S_t^k|/|\Omega|$  (red line) is a tight upper-bound or *surrogate function* [14].

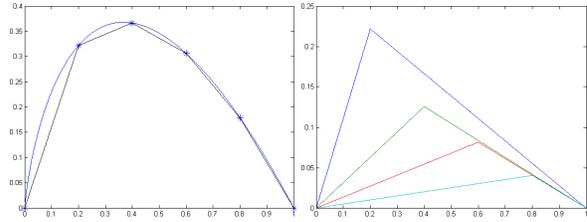


Figure 3. (*Entropy - high order optimization*) (a) polygonal approximation for  $-z \log z$ . (b) “triangle” functions decomposition.

high-order correction term  $H(S)$ :

$$\begin{aligned} \hat{E}(S, P) &= E(S, P) - \sum_{k=1}^K |S^k| \log \frac{|S^k|}{|\Omega|} \\ &= E(S, P) + |\Omega| \cdot H(S). \end{aligned} \quad (10)$$

Note that unbiased data term  $\hat{E}(S, P)$  should be used with caution in applications where allowed models  $P^k$  are highly descriptive. In particular, this applies to Zhu&Yuille [23] and GrabCut [18] where probability models are histograms or GMM. According to (9), optimization of model  $P^k$  will over-fit to data, i.e.  $KL(I^k|P^k)$  will be reduced to zero for arbitrary  $I^k = \{I_p \mid p \in S^k\}$ . Thus, highly descriptive models reduce  $\hat{E}(S, P)$  to conditional entropy  $H(S|I)$ , which only encourages consistent labeling for points of the same color. While this could be useful in segmentation, see *bin consistency* in [19], trivial solution  $S^0 = \Omega$  becomes good for energy  $\hat{E}(S, P)$ . Thus, bias to equal size segments in standard likelihoods (1) is important for histogram or GMM fitting methods [23, 18].

Many techniques with unbiased data term  $\hat{E}(S, P)$  avoid trivial solutions. Over-fitting is not a problem for simple models, e.g. Gaussians [6], lines, homographies [20, 8]. Label cost could be used to limit model complexity. Trivial solutions could also be removed by specialized regional terms added to the energy [19]. Indirectly, optimization methods that stop at a local minimum help as well.

**Bound optimization for (9-10):** Block-coordinate descent for energy  $E(W, S, P) := E_W(S, P)$  as in [11] can be seen as a *bound optimization* [14] for our unbiased energy  $\hat{E}(S, P)$ . According to (8) the optimal weights at any current solution  $S_t$  are  $W_t = \{\frac{|S_t^1|}{|\Omega|}, \dots, \frac{|S_t^K|}{|\Omega|}\}$  since they minimize  $KL(S_t|W)$ . The algorithm in [11] iteratively optimizes  $E_{W_t}(S, P)$  over  $P, S$  and resets to energy  $E_{W_{t+1}}(S, P)$  at each step until convergence. Figure 2 shows that at any given  $S_t$  energy  $E_{W_t}(S, P)$  is an *upper bound* for  $\hat{E}(S, P)$ , that is

$$\begin{aligned} \hat{E}(S, P) &\leq E_{W_t}(S, P) \quad \forall S \\ \hat{E}(S_t, P) &= E_{W_t}(S_t, P). \end{aligned}$$

This iterative approach suggests a simple method for removing volumetric bias in vision problems: use weighted likelihoods  $E_W(S, P)$  in your regularization energy (with standard optimization) and iteratively re-estimate  $W$ . However, our experiments show it is weaker than the following.

**High-order optimization for entropy in (10):** Alternatively, optimization of unbiased term  $\hat{E}(S, P)$  could be based on equation (10). Since term  $E(S, P)$  is unary for  $S$  the only issue is optimization of high-order entropy  $H(S)$ . The entropy combines terms  $-z \log z$  for  $z = |S^k|/|\Omega|$ . Each of these is a concave function of cardinality, which are known to be *submodular* [15]. As explained below, entropy is amenable to efficient discrete optimization techniques both in binary (Sec.3.2) and multi-label cases (Sec.3.2-3.3).

Optimization of concave cardinality functions was previously used in vision, *e.g.* for *label consistency* [12] or *bin consistency* [19]. We propose similar optimization methods in the context of entropy. We use a polygonal approximation with triangle functions as illustrated in Figure 3. Each triangle function is the minimum of two affine cardinality functions, yielding an approximation of the type

$$-\frac{|S^k|}{|\Omega|} \log \frac{|S^k|}{|\Omega|} \approx \sum_l \min(a_l^L |S^k|, a_l^U |S^k| + b_l^U). \quad (11)$$

Optimization of each ‘‘triangle’’ term in this summation can be done as follows. Cardinality functions like  $a_l^L |S^k|$  and  $a_l^U |S^k| + b_l^U$  are unary. Evaluation of their minimum can be done with an *auxiliary* variable  $y_l \in \{0, 1\}$  as in

$$\min_{y_l} y_l (a_l^L |S^k|) + \bar{y}_l (a_l^U |S^k| + b_l^U) \quad (12)$$

which is a pairwise energy. Indeed, consider binary segmentation problems  $S_p \in \{0, 1\}$ . Since

$$|S^k| = \begin{cases} \sum_{p \in \Omega} S_p, & \text{if } k = 1 \\ \sum_{p \in \Omega} (1 - S_p), & \text{if } k = 0 \end{cases} \quad (13)$$

(12) breaks into *submodular*<sup>3</sup> pairwise terms for  $y_l$  and  $S_p$ . Thus, each ‘‘triangle’’ energy (12) can be globally optimized

<sup>3</sup>Depending on  $k$ , may need to switch  $y_l$  and  $\bar{y}_l$ .

with graph cuts [13]. For more general multi-label problems  $S_p \in N$  energy terms (12) can be iteratively optimized via binary graph-cut moves like  $\alpha$ -expansion [5]. Indeed, let variables  $x_p \in \{0, 1\}$  represent  $\alpha$ -expansion from a current solution  $S_t = \{S_t^k\}$  to a new solution  $S$ . Since

$$|S^k| = \begin{cases} \sum_{p \in \Omega} x_p, & \text{if } k = \alpha \\ \sum_{p \in S_t^k} (1 - x_p), & \text{if } k \neq \alpha \end{cases} \quad (14)$$

(12) also reduces to submodular pairwise terms for  $y_l, x_p$ .

The presented high-order optimization approach makes stronger moves than the simpler bound optimization method in the previous sub-section. However, both methods use block coordinate descent iterating optimization of  $S$  and  $P$  with no quality guarantees. The next section shows examples with different optimization methods.

### 3. Examples

This sections considers several representative examples of computer vision problems where regularization energy uses likelihood term (1) with re-estimated models  $P^k$ . We empirically demonstrate bias to segments of the same size (2) and show advantages of different modifications of the data term proposed in the previous section.

#### 3.1. Segmentation with target volumes

In this section we consider a biomedical example with  $K = 3$  segments:  $S_1$  background,  $S_2$  liver,  $S_3$  substructure inside liver (blood vessels or cancer), see Fig.4. The energy combines standard data term  $E(S, P)$  from (1), boundary length  $|\partial S|$ , an inclusion constraint  $S_3 \subset S_2$ , and a penalty for  $L_2$  distance between the background segment and a given shape template  $T$ , as follows

$$E(S, P) + \lambda |\partial S| + [S_3 \subset S_2] + \beta \|S_1 - T\|_{L_2}. \quad (15)$$

For fixed models  $P^k$  this energy can be globally minimized over  $S$  as described in [7]. In this example intensity likelihood models  $P^k$  are histograms treated as unknown parameters and estimated using block-coordinate descent for variables  $S$  and  $P$ . Figure 4 compares optimization of (15) in (b) with optimization of a modified energy replacing standard likelihoods  $E(S, P)$  with a weighted data term in (6)

$$E_W(S, P) + \lambda |\partial S| + [S_3 \subset S_2] + \beta \|S_1 - T\|_{L_2} \quad (16)$$

for fixed weights  $W$  set from specific target volumes (c-d).

The teaser in Figure 1(c) demonstrates a similar example for separating a kidney from a liver based on Gaussian models  $P^k$ , as in Chan-Vese [6], instead of histograms. Standard likelihoods  $E(P, S)$  in (15) show equal-size bias, which is corrected by weighted likelihoods  $E_W(P, S)$  in (16) with approximate target volumes  $W = \{0.05, 0.95\}$ .

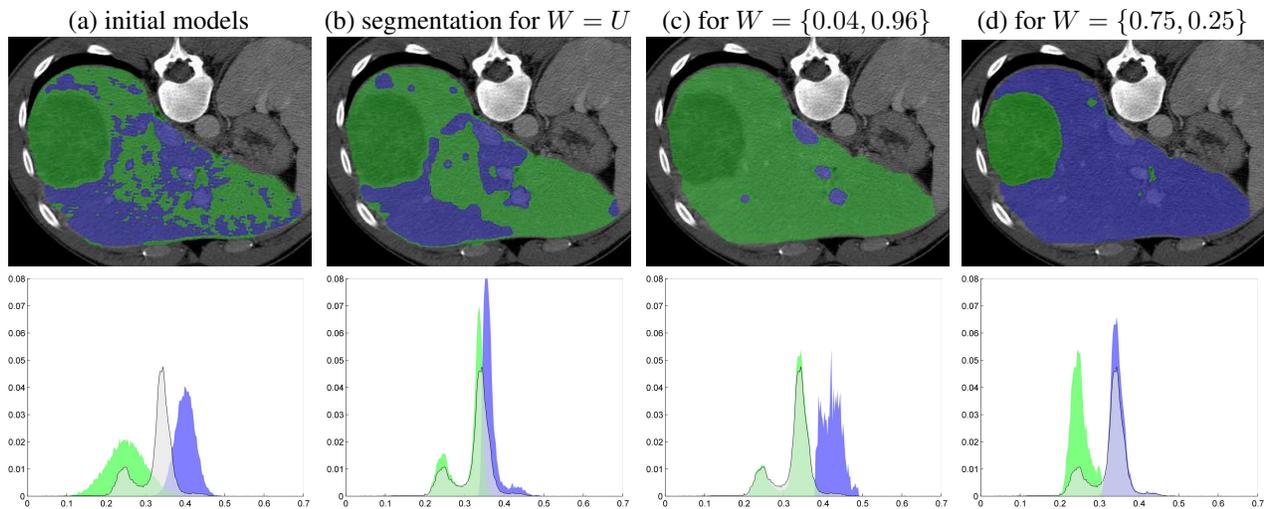


Figure 4. Equal volumes bias  $KL(S|U)$  versus target volumes bias  $KL(S|W)$ . Grey histogram is a distribution of intensities for the ground truth liver segment including normal liver tissue (the main mode), blood vessels (the small mode on the right), and cancer tissue (the left mode). (a) Initial (normalized) histograms for two liver parts. Initial segmentation shows which histogram has larger value for each pixel’s intensity. (b) The result of optimizing energy (15). The solid blue and green (normalized) histograms at the bottom row are for intensities at the corresponding segments. (c-d) The results of optimizing energy (16) for fixed weights  $W$  set for specific target volumes.

### 3.2. Segmentation without volumetric bias

We demonstrate in different applications a practically significant effect of removing the volumetric bias, i.e., using our functional  $\hat{E}(S, P)$ . We first report comprehensive comparisons of binary segmentations on the GrabCut data set [18], which consists of 50 color images with ground-truth segmentations and user-provided bounding boxes<sup>4</sup>. We compared three energies: high-order energy  $\hat{E}(S, P)$  (10), standard likelihoods  $E(S, P)$  (1), which was used in the well-known GrabCut algorithm [18], and  $E_W(S, P)$  (6), which constrains the solution with true target volumes (i.e., those computed from ground truth). The appearance models in each energy were based on histograms encoded by 16 bins per channel, and the image data is based color specified in RGB coordinates. For each energy, we added a standard contrast-sensitive regularization term [18, 2]:  $\lambda \sum_{p,q \in \mathcal{N}} \alpha_{pq} [S_p \neq S_q]$ , where  $\alpha_{pq}$  denote standard pairwise weights determined by color contrast and spatial distance between neighboring pixels  $p$  and  $q$  [18, 2].  $\mathcal{N}$  is the set neighboring pixels in a 8-connected grid.

We further evaluated two different optimization schemes for high-order energy  $\hat{E}(S, P)$ : (i) bound optimization and (ii) high-order optimization of concave cardinality potential  $H(S)$  using polygonal approximations; see Sec.2 for details. Each energy is optimized by alternating two iterative steps: (i) fixing the appearance histogram models and optimizing the energy w.r.t  $S$  using graph cut [4]; and (ii) fixing segmentation  $S$  and updating the histograms from cur-

rent solution. For all methods we used the same appearance model initialization based on a user-provided box<sup>5</sup>.

The error is evaluated as the percentage of mis-classified pixels with respect to the ground truth. Table 1 reports the best average error over  $\lambda \in [1 \dots 30]$  for each method. As expected, using the true target volumes yields the lowest error. The second best performance was obtained by  $\hat{E}(S, P)$  with high-order optimization; *removing the volumetric bias substantially improves the performance of standard log-likelihoods reducing the error by 6%*. The bound optimization obtains only a small improvement as it is more likely to get stuck in weak local minima. We further show representative examples for  $\lambda = 16$  in the last two rows of Table 1, which illustrate clearly the effect of both equal-size bias in (1) and the corrections we proposed in (10) and (6).

It is worth noting that the error we obtained for standard likelihoods (the last column in Table 1) is significantly higher than the 8% error previously reported in the literature, e.g., [21]. The lower error in [21] is based on a different (more recent) set of tighter bounding boxes [21], where the size of the ground-truth segment is roughly half the size of the box. Therefore, the equal-size bias in  $\hat{E}(S, P)$  (10) for this particular set of boxes has an effect similar to the effect of true target volumes  $W$  in  $E_W(S, P)$  (6) (the first column in Table 1), which significantly improves the performance of standard likelihoods (the last column). In practice, both 50/50 boxes and true  $W$  are equally unrealistic

<sup>4</sup><http://research.microsoft.com/en-us/um/cambridge/projects/visionimagevideoediting/segmentation/grabcut.htm>

<sup>5</sup>The data set comes with two boxes enclosing the foreground segment for each image. We used the outer bounding box to restrict the image domain and the inner box to compute initial appearance models.

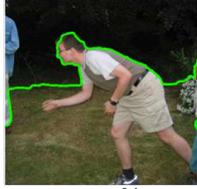
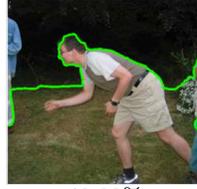
Energy	$E_W(S, P)$ (6) true target volumes $W$	$\hat{E}(S, P)$ (10) high-order optimization	$\hat{E}(S, P)$ (9) bound optimization	$E(S, P)$ (1) standard likelihoods
Overall Error (50 images)	5.29%	7.87%	13.41%	14.41%
Examples	 error: 4.75%  error: 2.29%	 error: 6.85%  error: 4.95%	 error: 9.64%  error: 41.20%	 error: 14.69%  error: 40.88%

Table 1. Comparisons on the GrabCut data set.

assumptions that require knowledge of the ground truth.

Fig. 5 depicts a different application, where we segment a magnetic resonance image (MRI) of the brain into multiple regions ( $K > 2$ ). Here we introduce an extension of  $\hat{E}(S, P)$  using a positive factor  $\gamma$  that weighs the contribution of entropy against the other terms:

$$\hat{E}_\gamma(S, P) = E(S, P) + \gamma|\Omega|H(S). \quad (17)$$

This energy could be written as

$$\sum_{k=1}^K |S^k| KL(I^k | P^k) + |\Omega|H(S|I) + (\gamma - 1)|\Omega|H(S)$$

using the high-order decomposition of likelihoods  $E(S, P)$  from [11] presented in the intro. Thus, the bias introduced by  $H(S)$  has two cases:  $\gamma \leq 1$  (volumetric *equality* bias) and  $\gamma \geq 1$  (volumetric *disparity* bias), as discussed below.

We use Chan-Vese data term [6] assuming the appearance models in  $E(S, P)$  are Gaussian distributions  $-\log P^k(I_p) \stackrel{c}{=} (I_p - \mu^k)^2 / 2\sigma^2$  where  $\mu^k$  is the mean of intensities in segment  $S^k$  and  $\sigma$  is a given constant. We also added a standard total-variation term [22] for regularization.

The solution is sought following the bound optimization strategy discussed earlier, see Fig. 2. The algorithm iterates over two steps: (i) optimizing a bound of  $\hat{E}_\gamma(S, P)$  w.r.t segmentation  $S$  via a continuous convex-relaxation technique [22] with fixed model parameters and (ii) update parameters  $\mu^k$  and  $w^k$  using fixed current solution  $S$ . We set the initial number of models to 5 and fixed  $\lambda = 0.1$  and  $\sigma = 0.05$ . We run the method for  $\gamma = 0$ ,  $\gamma = 1$  and  $\gamma = 3$ . Figure 5 displays the results using colors encoded by the

region means obtained at convergence. Column (a) demonstrates the equal-size bias for  $\gamma = 0$ ; notice that the yellow, red and brown components have approximately the same size. Setting  $\gamma = 1$  in (b) removes this bias, yielding much larger discrepancies in size between these components. Example (c) shows that larger weight  $\gamma > 1$  has a *sparsity* effect; it reduces the number of distinct segments/labels from 5 to 3. At the same time, for  $\gamma > 1$  this energy introduces *disparity* bias: the gap between the volumes of orange and brown segments increases compared to  $\gamma = 1$  in (b) where no volumetric bias is present. This disparity bias is opposite to the equality bias for  $\gamma < 1$  in (a). Note that  $\gamma = 1$  in combination with a separate *label cost* term [8] should also give *sparsity*, but without the *disparity* bias.

### 3.3. Geometric model fitting

Energy minimization methods for geometric model fitting problems have recently gained popularity due to [10]. Similarly to segmentation these methods are often driven by a maximum likelihood based data term measuring model fit to the particular feature. The theory presented in Section 2 applies to these problems as well and they therefore exhibit the same kind of volumetric bias.

Figure 1 (b) shows a simple homography estimation example. Here we captured two images of a scene with two planes and tried to fit homographies to these (the right image with results is shown in Figure 1). For this image pair SIFT [16] generated 3376 matches on the larger plane (paper and floor) and 135 matches on the smaller plane (book). For a pair of matching points  $I_p = \{x_p, y_p\}$  we use the log

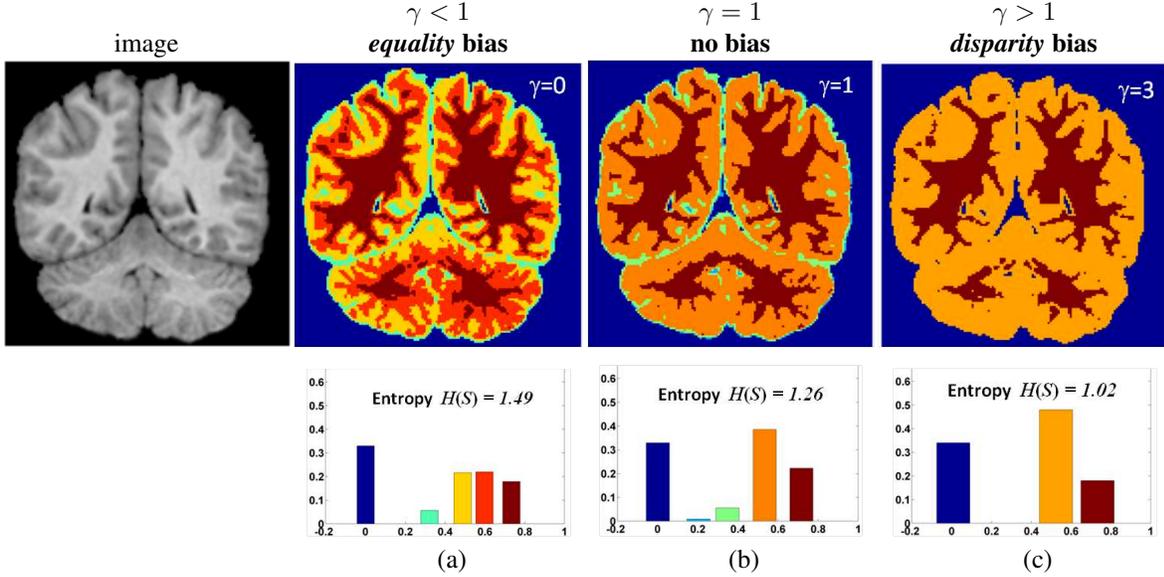


Figure 5. Segmentation using energy (17) combined with standard TV regularization [22]. We use the Chan-Vese model [6] as appearance term and bound optimization computing a local minimum of the energy, see Fig.2. At each iteration, the bound is optimized w.r.t segmentation using the convex-relaxation technique in [22]. Initial number of models is 5,  $\lambda = 0.1$ , and  $\sigma = 0.05$ . Upper row (from left to right): image data and results for  $\gamma = 0$ ,  $\gamma = 1$  and  $\gamma = 3$ . Lower row: histograms of the number of assignments to each label and the entropies at convergence. Note that the *sparsity* effect observed for  $\gamma > 1$  could be obtained without *disparity* bias using  $\gamma = 1$  and *label costs* [8].

likelihood costs

$$\sum_{p \in S^k} -\log(w^k \cdot P^{H_k, \Sigma_k}(I_p)), \quad (18)$$

where  $P^{H_k, \Sigma_k}(I_p) = \frac{1}{(2\pi)^2 \sqrt{|\Sigma_k|}} e^{-\frac{1}{2} d_{H_k, \Sigma_k}(x_p, y_p)^2}$  and  $d_{H_k, \Sigma_k}$  is the symmetric mahalanobis transfer distance. The solution to the left in Figure 1 (b) was generated by optimizing over homographies and covariances while keeping the priors fixed and equal ( $w^1 = w^2 = 0.5$ ). The volume bias makes the smaller plane (blue points) grab points from the larger plane. For comparison Figure 1 (b) also shows the result obtained when reestimating  $w^1$  and  $w^2$ . Note that the two algorithms were started with the same homographies and covariances. Figure 6 shows an independently computed 3D reconstruction using the same matches as for the homography experiment.

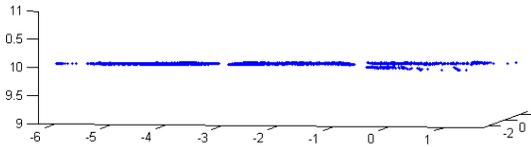


Figure 6. 3D-Geometry of the book scene in Figure 1 (b).

### 3.3.1 Multi Model Fitting

Recently discrete energy minimization formulations have been shown to be effective for geometric model fitting tasks [10, 8]. These methods effectively handle regularization terms needed to produce visually appealing results. The typical objective functions are of the type

$$E(S, W, \Theta) = V(S) + D(S, W, \Theta) + L(S), \quad (19)$$

where  $V(S) = \sum_{(p,g) \in \mathcal{N}} V_{pq}(S_p, S_q)$  is a smoothness term and  $L(S)$  is a label cost preventing over fitting by penalizing the number of labels. The data term

$$D(S, W, \Theta) = -\sum_k \sum_{S_p=k} (\log(w^k) + P(m_p | \Theta^k)) \quad (20)$$

consists of log-likelihoods for the observed measurements  $m_p$ , given the model parameters  $\Theta$ . Typically the prior distributions  $w^k$  are ignored (which is equivalent to letting all  $w^k$  be equal) hence resulting in a bias to equal partitioning. Because of the smoothness and label cost terms the bias is not as evident in practical model fitting applications as in k-means, but as we shall see it is still present.

Multi model fitting with variable priors presents an additional challenge. The PEARL (Propose, Expand And Reestimate Labels) paradigm [10] naturally introduces and removes models during optimization. However, when reestimating priors, a model  $k$  that is not in the current labeling will have  $w^k = 0$  giving an infinite log-likelihood penalty.

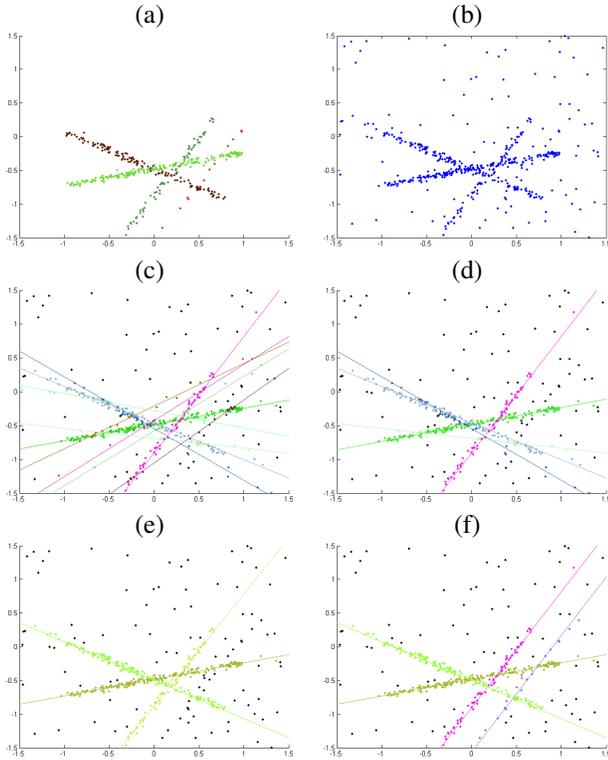


Figure 7. *Line fitting*: (a) data generated from three lines, (b) data with outliers, (c) fixed  $W$  and  $h = 100$ , (d) fixed  $W$  and  $h = 200$ , (e) fixed  $W$  and  $h = 300$ , (f) variable  $W$  and  $h = 5$ .

Therefore a simple alternating approach (see *bound optimization* in Sec.2) will be unable to add new models to the solution. For sets of small cardinality it can further be seen that the entropy bound in Figure 2 will become prohibitively large since the derivative of the entropy function is unbounded (when approaching  $w = 0$ ). Instead we use  $\alpha$ -expansion moves with higher order interactions to handle the entropy term, as described in Section 2.

Figure 7 shows the result of a synthetic line fitting experiment. Here we randomly sampled points from four lines with different probabilities, added noise with  $\sigma = 0.025$  and added outliers. We used energy (19) without smoothness and with label cost  $h$  times the number of labels (excluding the outlier label). The model parameters  $\Theta$  consist of line location and orientation. We treated the noise level for each line as known. Although the volume bias manifests itself more clearly when the variance is reestimated, it is also present when only the means are estimated.

Using random sampling we generated 200 line proposals to be used by both methods (fixed and variable  $W$ ). Figure 7 (c), (d) and (e) show the results with fixed  $W$  for three different strengths of label cost. Both the label cost and the entropy term want to remove models with few assigned points. However, the label cost does not favor any assignment when



Figure 8. *Homography fitting*: fixed (left) and variable  $W$  (right).

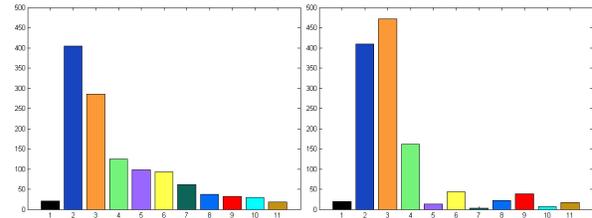


Figure 9. Histogram of the number of assignments to each label (model) in Figure 8. Fixed  $W$  (left) and variable  $W$  (right).

it is not strong enough to remove a model. Therefore it cannot counter the volume bias of the standard data term favoring more assignments to weaker models. In the line fitting experiment of Figure 7 we varied the strength of the label cost (three settings shown in (c), (d) and (e)) without being able to correctly find all the 4 lines. Reestimation of  $W$  in Figure 7 (f) resulted in a better solution.

Figures 8 and 9 show the results of a homography estimation problem with the smoothness term  $V(S)$ . For the smoothness term we followed [10] and created edges using a Delaunay triangulation with weights  $e^{-d^2/5^2}$ , where  $d$  is the distance between the points. For the label costs we used  $h = 100$  with fixed  $W$  and  $h = 5$  with variable  $W$ . We fixed the model variance to  $5^2$  (pixels<sup>2</sup>).

The two solutions are displayed in Figure 8 and Figure 9 shows a histogram of the number of assigned points to each model (black corresponds to the outlier label). Even though smoothness and label costs mask it somewhat, the bias to equal volume can be seen here as well.

## Acknowledgements

This research was generously supported by several grants from Canadian NSERC Discovery and RTI programs, The Fields Institute for Research in Mathematical Sciences, US National Institute of Health (no. 1 12074 00 02), Swedish Research Council (grant no. 2012-4213), and The Crafoord Foundation.

## References

- [1] O. Barinova, V. Lempitsky, and P. Kohli. On the Detection of Multiple Object Instances using Hough Transforms. In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, June 2010. 1, 2
- [2] Y. Boykov and M.-P. Jolly. *Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images*. In *International Conference on Computer Vision*, volume I, pages 105–112, July 2001. 1, 5
- [3] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *International Conference on Computer Vision*, volume I, pages 26–33, 2003. 2
- [4] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, September 2004. 5
- [5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, November 2001. 2, 4
- [6] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001. 1, 2, 3, 4, 6, 7
- [7] A. Delong and Y. Boykov. Globally Optimal Segmentation of Multi-Region Objects. In *International Conference on Computer Vision (ICCV)*, 2009. 1, 4
- [8] A. Delong, A. Osokin, H. Isack, and Y. Boykov. Fast Approximate Energy Minimization with Label Costs. *International Journal of Computer Vision (IJCV)*, 96(1):1–27, January 2012. 1, 2, 3, 6, 7
- [9] L. Gorelick, F. R. Schmidt, and Y. Boykov. Fast trust region for segmentation. In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1714–1721, Portland, Oregon, June 2013. 2
- [10] H. N. Isack and Y. Boykov. Energy-based Geometric Multi-Model Fitting. *International Journal of Computer Vision (IJCV)*, 97(2):123–147, April 2012. 6, 7, 8
- [11] M. Kearns, Y. Mansour, and A. Ng. An Information-Theoretic Analysis of Hard and Soft Assignment Methods for Clustering. In *Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, August 1997. 1, 2, 3, 4, 6
- [12] P. Kohli, L. Ladicky, and P. H. S. Torr. Robust Higher Order Potentials for Enforcing Label Consistency. *International Journal of Computer Vision (IJCV)*, 82(3):302–324, 2009. 2, 4
- [13] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, February 2004. 4
- [14] K. Lange, D. R. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics*, 9(1):1–20, 2000. 3, 4
- [15] L. Lovasz. Submodular functions and convexity. *Mathematical programming: the state of the art*, pages 235–257, 1983. 2, 4
- [16] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 6
- [17] C. Nieuwenhuis, E. Strekalovskiy, and D. Cremers. Proportion priors for image sequence segmentation. In *International Conference on Computer Vision (ICCV)*, pages 2328–2335, 2013. 2
- [18] C. Rother, V. Kolmogorov, and A. Blake. Grabcut - interactive foreground extraction using iterated graph cuts. In *ACM transactions on Graphics (SIGGRAPH)*, August 2004. 1, 2, 3, 5
- [19] M. Tang, L. Gorelick, O. Veksler, and Y. Boykov. From GrabCut to One Cut. In *International Conference on Computer Vision (ICCV)*, December 2013. 2, 3, 4
- [20] P. Torr. Geometric motion segmentation and model selection. *Philosophical transactions of the Royal Society A*, 356:1321–1340, 1998. 1, 2, 3
- [21] S. Vicente, V. Kolmogorov, and C. Rother. Joint optimization of segmentation and appearance models. In *IEEE International Conference on Computer Vision (ICCV)*, pages 755–762, 2009. 5
- [22] J. Yuan, E. Bae, X. Tai, and Y. Boykov. A continuous max-flow approach to potts model. In *European Conference on Computer Vision (ECCV), Part VI*, pages 379–392, 2010. 6, 7
- [23] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, September 1996. 1, 2, 3