Proportion Priors for Image Sequence Segmentation

Claudia Nieuwenhuis*
UC Berkeley, ICSI, USA

Evgeny Strekalovskiy
TU Munich, Germany

Daniel Cremers
TU Munich, Germany

Abstract

We propose a convex multilabel framework for image sequence segmentation which allows to impose proportion priors on object parts in order to preserve their size ratios across multiple images. The key idea is that for strongly deformable objects such as a gymnast the size ratio of respective regions (head versus torso, legs versus full body, etc.) is typically preserved. We propose different ways to impose such priors in a Bayesian framework for image segmentation. We show that near-optimal solutions can be computed using convex relaxation techniques. Extensive qualitative and quantitative evaluations demonstrate that the proportion priors allow for highly accurate segmentations, avoiding seeping-out of regions and preserving semantically relevant small-scale structures such as hands or feet. They naturally apply to multiple object instances such as players in sports scenes, and they can relate different objects instead of object parts, e.g. organs in medical imaging. The algorithm is efficient and easily parallelized leading to proportion-consistent segmentations at runtimes around one second.

1. Introduction

1.1. Image Sequence Segmentation

Automatic image sequence segmentation denotes the task of jointly segmenting one or several objects from a series of images taken under different view points, lighting conditions and background scenes. The difficulty lies in the fact that none of the objects' properties is guaranteed to be preserved over time. Both the geometry and the photometry of the objects of interest may change from one image to the next. Changing illumination affects the observed color model, different viewpoints lead to different object scales and possibly self-occlusions, whole parts of the object can even vanish from the image, and objects may occlude each other in the case of multiple foreground objects. Moreover, articulations and non-rigid deformations give rise to substantial shape changes. Modeling such variable conditions

Multi-label segmentation with length regularity
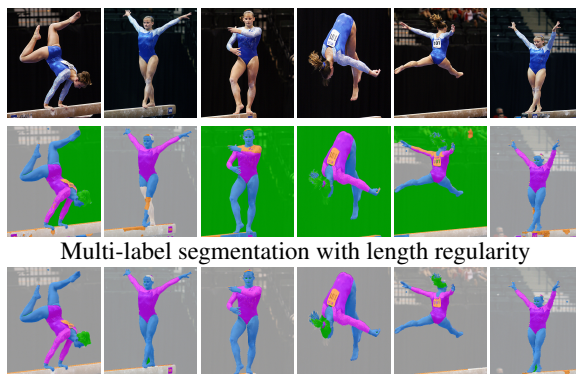
Multi-label segmentation with proportion priors

Figure 1. Proportion priors allow to constrain the relative size between different object parts (e.g. hair, skin, body, 101-sign). They enable stable segmentations over long image sequences, preventing the seeping out of regions into the background (e.g. the green hair region) or the removal of semantically important small parts (e.g. hands or feet).

and exploiting information which is shared among the images is a challenging task.

In general, the co-segmentation problem deals with the situation that we have no knowledge on the object of interest besides that it appears in all images. To make the problem tractable many approaches introduced at least some kind of prior knowledge based on training data or user scribbles [1, 12, 7, 21]. The resulting optimization problems are often iterative and hard to optimize [12, 19, 11, 20] leading to runtimes of several seconds or even minutes. For complex real-world image sequences the task of leveraging relevant shared information for co-segmentation remains an open challenge. In this paper, we propose a Bayesian framework for multi-region co-segmentation which allows to impose learned *proportion priors* — see Figure 1. We show that near-optimal solutions can be efficiently computed by means of convex relaxation techniques.

1.2. Related Work

Existing co-segmentation approaches can be loosely grouped into two classes regarding the additional assumptions they make:

Shape similarity: One class of methods assumes that the segmentations of each image only differ by a rigid body transformation [22] or — somewhat relaxed — that they are highly similar [18]. Unfortunately, this limits the applicability to rigid objects observed from similar viewpoints. Upon strong viewpoint changes, articulations or non-rigid deformations, the arising segmentations are too different in their shape to be accounted for. Multiple foreground objects cannot be handled, either.

Color similarity: On the other extreme are methods which make no assumptions on *shape similarity*, but which strongly rely on *similarity of color or feature distributions* [19, 11, 6, 1, 21, 5]. These methods do not generalize well to certain real-world challenges where firstly object and background may have similar and possibly overlapping color distributions and where secondly these distributions may vary from image to image due to illumination changes, motion blur or scale-dependent color sampling. Recent approaches relax these assumptions, requiring less color similarity [7] and instead derive high level knowledge on object properties such as object subspaces [12] or region correspondences [20]. However, the resulting optimization problems become more difficult to solve.

To increase the stability of segmentation results, it was suggested [8] to impose shape moment constraints into variational segmentation methods. Since these constraints are *absolute*, they are not invariant to changes in scale, viewpoint, occlusions or multiple object instances and are thus not applicable to the general co-segmentation problem.

1.3. Contribution

In this paper we revisit the problem of image sequence segmentation and reconsider the following question: Can we identify aspects of an object's *shape* which are shared among the various images of this object and are preserved despite rigid transformation, despite articulation and despite substantial non-rigid deformation?

The central idea is to tackle the problem in a multilabel framework where an object, say an athlete, is made up of multiple components (the various limbs of the body). While the object may undergo substantial changes — rigid body motion, articulation, non-rigid deformation — what is typically preserved is the relative size of object parts (e.g. the head to the entire body), the object part proportions. We formulate image sequence segmentation as a problem of Bayesian inference and introduce *proportion priors* to restrict the relative size of object parts. This approach comes with the following advantages:

- It can handle overlapping color distributions, moderately variable lighting conditions, various object scales and multiple foreground objects. The proposed ratio constraints preserve small or elongated object parts.

- It extends recent convex relaxation techniques [9, 2, 23] from multilabel segmentation to *multilabel sequence segmentation*. We present an efficient optimization scheme which can be parallelized with runtimes of around one second to compute pixel-accurate segmentations.

- The approach yields state-of-the-art results on the ICoseg benchmark for subdivisible object sequences.

2. Multilabel Image Sequence Segmentation

Let $I : \Omega \to \mathbb{R}^3$ denote an input image of the sequence defined on the domain $\Omega \subseteq \mathbb{R}^2$. The task of segmentation is to partition the image plane into a set of n pairwise disjoint regions Ω_i s.t. $\Omega = \bigcup_{i=1}^{n} \Omega_i$ and $\Omega_i \cap \Omega_j = \emptyset \ \forall i \neq j$. It can be solved by computing a labeling $l : \Omega \to \{1, \ldots, n\}$ indicating which of the n regions each pixel belongs to: $\Omega_i = \{x \mid l(x) = i\}$. In the framework of Bayesian inference, one can compute such a segmentation by maximizing the conditional probability

$$\arg \max_l \mathcal{P}(l \mid I) = \arg \max_l \mathcal{P}(I \mid l) \, \mathcal{P}(l). \quad (1)$$

It combines the observation likelihood $\mathcal{P}(I \mid l)$ (typically favoring a color-based region association) with prior knowledge $\mathcal{P}(l)$ regarding what kinds of partitionings are more or less likely.

In the case of image sequence segmentation, commonly used color and boundary length priors are often insufficient to obtain good results. Firstly, the object boundaries are not necessarily short. Secondly, the color distributions of object and background may have substantial overlap and may exhibit strong variations across images. In order to stabilize the segmentation process against color and lighting variations, pose changes and substantial non-rigid deformations, and in order to leverage it to a parsing of objects into their semantic components, we propose to introduce proportion priors into the optimization problem.

2.1. Framework for Proportion Preserving Priors

In the following, we will introduce proportion priors as a means to impose information on the *relative size* of respective object parts. Whereas the absolute size of parts will vary with viewing angle and distance from the camera, their relative size is typically well preserved – i.e. the size of the head is typically 10% of the size of the entire body.

Let us assume that the object we want to segment can be divided into $n-1$ sub-regions (e.g. head, feet, body and hands) with the n-th region denoting the background of the image. Then in the Bayesian framework, the prior $\mathcal{P}(l) = \mathcal{P}(\Omega_1, \ldots, \Omega_n)$ can be expressed in the following way:

$$
\begin{aligned}
\mathcal{P}(l) &= \mathcal{P}(\Omega_1, \ldots, \Omega_{n-1} \mid \Omega_n) \, \mathcal{P}(\Omega_n) \\
&= \prod_{i=1}^{n-1} \mathcal{P}(\Omega_i \mid \Omega_n) \, \mathcal{P}(\Omega_n),
\end{aligned}
\quad (2)
$$

where we assume conditional independence of the segments $\Omega_1, .., \Omega_{n-1}$ given the background Ω_n. A ratio constraint relates the size Ω_i of the i-th region to the size $\sum_{j=1}^{n-1} |\Omega_j|$ of the whole object:

$$r_i = \frac{|\Omega_i|}{\sum_{j=1}^{n-1} |\Omega_j|} = \frac{|\Omega_i|}{|\Omega| - |\Omega_n|}, \quad 1 \le i < n. \quad (3)$$

For segmentation we want to impose regions of short boundary length $\text{Per}(\Omega_i)$, whose ratios additionally follow a learned (or specified) ratio probability distribution $\mathcal{P}_p$

$$\mathcal{P}(\Omega_i|\Omega_n) = \frac{1}{C} \exp\left(-\frac{\lambda}{2}\text{Per}(\Omega_i)\right) \cdot \mathcal{P}_p(r_i), \quad 1 \le i < n, \quad (4)$$

where C is a normalization constant and λ a weighting parameter. Finally, we assume each background Ω_n to be equally likely a priori, so that $\mathcal{P}(\Omega_n) = const$.

Instead of maximizing $\mathcal{P}(I|l)\mathcal{P}(l)$ in (1) we minimize its negative logarithm and obtain the energy

$$E(\Omega_1, \ldots, \Omega_n, r_1, .., r_{n-1}) = \quad (5)$$

$$\sum_{i=1}^{n}\left\{\int_{\Omega_i} f_i(x)\,dx + \tfrac{\lambda}{2}\text{Per}(\Omega_i)\right\} - \sum_{i=1}^{n-1}\log\mathcal{P}_p(r_i)$$

$$\text{s.t. } r_i = \frac{|\Omega_i|}{|\Omega| - |\Omega_n|} \quad 1 \le i < n,$$

where $f_i(x) := -\log\mathcal{P}(I(x)|l(x) = i)$.

2.2. Contribution: Proportion Preserving Priors

In this section we propose two different proportion preserving priors: the uniform distribution prior and the Laplace distribution prior. Both assume that the ratios of the object parts follow a specific distribution $\mathcal{P}_p(r_i)$ whose parameters are estimated from sample data, i.e. sample segmentations from which ratio samples can be obtained.

To convert the energy in (5) to a convex optimization problem in Section 2.3, the key challenge is to express the terms $-\log\mathcal{P}_p(r_i)$ in (5) as a convex function of the variables $a_i := |\Omega_i|/|\Omega|$, which denote the fraction of the size of region Ω_i with respect to the image size. It holds that $a_i \in [0,1]$ and $\sum_{i=1}^n a_i = 1$. The ratios r_i in (3) can be easily expressed in terms of a_i by

$$r_i = \frac{a_i}{1 - a_n}. \quad (6)$$

2.2.1 Uniform Distribution Prior

As a first case, we assume a uniform distribution of the ratios r_i over a specific interval $[l_i, h_i]$. The left and right boundaries l_i and h_i are computed from training data by means of maximum likelihood estimation, which assigns l_i

and h_i the minimum and maximum values of the sample ratios, respectively. We obtain for the ratio probabilities

$$\mathcal{P}_p(r_i) = \begin{cases} \frac{1}{h_i - l_i} & \text{if } l_i \le r_i \le h_i, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

Since $-\log\mathcal{P}_p$ is either constant or infinity, this prior corresponds to $2(n-1)$ constraints in the optimization problem:

$$l_i(1 - a_n) \le a_i \le h_i(1 - a_n) \quad \forall 1 \le i < n. \quad (8)$$

These relative ratio constraints are linear and thus convex in terms of the a_i. The advantage of this prior is the simple computation and the convexity of the constraints. Yet, in the case of large variations of the ratios r_i in the sample data (i.e. sample outliers) we can obtain very large intervals $[l_i, h_i]$, which hardly limit the resulting segmentations. Therefore, we propose Laplace distribution priors.

2.2.2 Laplace Distribution Prior

The Laplace distribution prior penalizes deviations of the ratios r_i from their median $\bar{r}_i$. In this way, the influence of ratio sample outliers on the constraints is limited. We assume the following Laplace distribution

$$\mathcal{P}_p(r_i) = \frac{1}{2\sigma_i} \exp\left(-\frac{|r_i - \bar{r}_i|}{\sigma_i}\right). \quad (9)$$

Given a set of sample ratios $s_{i1}, \ldots, s_{iM}$ for segment i from M training segmentations, the parameters $\bar{r}_i$ are obtained by means of maximum likelihood estimation as the median of the samples for each ratio, and $\sigma_i = \frac{1}{M}\sum_{j=1}^M |s_{ij} - \bar{r}_i|$.

Taking the negative log-likelihood in energy (5) and multiplying it by a parameter $\mu > 0$ for balancing the prior with respect to the other terms, we get the energy

$$E_p(r_i) = -\mu\log\mathcal{P}_p(r_i) = \frac{\mu}{\sigma_i}|r_i - \bar{r}_i|. \quad (10)$$

Unfortunately, after replacing r_i by (6), this function is not convex in a_i and a_n. For example, for $r_i < \bar{r}_i$ we obtain $E_p(r_i) = \frac{\mu}{\sigma_i}(\bar{r}_i - \frac{a_i}{1 - a_n})$, which is not convex in a_n for fixed a_i. To make global optimization possible, in the following we propose two methods to convexify this prior.

Convex Relaxation. First, we consider the convex relaxation of E_p as a function $E_p(a_i, a_n)$, i.e. the tightest possible convex *lower bound*. The definition domain of E_p is naturally given by $a_i, a_n \ge 0$ and $a_i + a_n \le 1$, where the latter inequality follows from $\sum_{j=1}^n a_j = 1$. We obtain the convex relaxation

$$E_1 := \frac{\mu}{\sigma_i}\big|a_i - \bar{r}_i(1 - a_n)\big|. \quad (11)$$

Proof. See supplementary material. $\quad\square$

In contrast to the Laplace prior (10) this approximation is convex. However, E_1 is minimal not only for $a_i = \bar{r}_i(1 - a_n)$ but also for $a_n = 1$, $a_i = 0$, i.e. when the segmentation only consists of the background. Thus, this prior is biased towards smaller foreground object areas. This can be understood as an additional compactness prior which removes cluttered background regions, but sometimes also parts of the objects.

Convex Upper Bound. As shown in the last paragraph, even the *greatest* convex lower bound E_1 is too small. This suggests to go in the other direction. As a second convexification method we propose a convex *upper bound* on E_p. Note that E_p does not have a lowest upper bound, in contrast to the convex relaxation case. To arrive at one possible solution, the idea is to write E_p in (10) by replacing the ratios r_i by the a_i in (6)

$$E_p = \frac{\mu}{\sigma_i} \frac{\left| a_i - \bar{r}_i(1 - a_n) \right|}{\sqrt{1 - a_n}} \cdot \frac{1}{\sqrt{1 - a_n}}$$

and apply Young's inequality $pq \leq \frac{1}{4\varepsilon}p^2 + \varepsilon q^2$, valid for all $p, q \in \mathbb{R}$ and arbitrary $\varepsilon > 0$ (we choose $\varepsilon = 10$ in the experiments). It is equivalent to $\left(\frac{1}{2\sqrt{\varepsilon}}p - \sqrt{\varepsilon}q \right)^2 \geq 0$. This leads to $E_p \leq E_2$ with

$$E_2 := \frac{\mu^2}{4\varepsilon\sigma_i^2} \frac{\left(a_i - \bar{r}_i(1 - a_n) \right)^2}{1 - a_n} + \frac{\varepsilon}{1 - a_n}. \quad (12)$$

This energy is convex: the first addend is a linear transformation of the convex function $(x, y) \mapsto \frac{x^2}{y}$, $x \in \mathbb{R}$, $y > 0$, and the second one is obviously convex in $a_n \in [0, 1)$.

The main advantage is that E_2 always favors the relation $a_i \approx \bar{r}_i(1 - a_n)$ regardless of the size of the object. It also avoids large background regions as $E_2 \to \infty$ for $a_n \to 1$, which can be alleviated by using a small ε.

Since (12) is used for all $1 \leq i < n$, for optimization it will be convenient to decouple the a_is from a_n using the dual representation $\frac{1}{4\widehat{\varepsilon}} \frac{x^2}{y} = \sup_{\beta \geq \widehat{\varepsilon}\alpha^2} \alpha x - \beta y$ for all $x, y \in \mathbb{R}$, $\widehat{\varepsilon} > 0$, where the left hand side is defined as ∞ for $y < 0$, as well as for $x \neq 0$ and $y = 0$, and as zero for $x = y = 0$. Applying this dual formulation for $1 \leq i < n$, we obtain the following dual representation of the convex upper bound for the Laplace proportion preserving prior:

$$\sup_{\alpha, \beta} \sum_{i=1}^{n-1} \left(\alpha_i \big(a_i - \bar{r}_i(1 - a_n) \big) - \beta_i(1 - a_n) \right) + \frac{\varepsilon(n - 1)}{1 - a_n}. \quad (13)$$

with the constraints $\beta_i \geq \frac{\varepsilon\sigma_i^2}{\mu^2}\alpha_i^2$ for each $1 \leq i < n$.

We note that a Gaussian distribution instead of a Laplace distribution (9) would still yield a non-convex E_p in (10), while leading to more complex and slower relaxations. In addition, we observed that the Laplace prior better represents the training data than the Gaussian prior.

2.3. Conversion to a Convex Optimization Problem

Several approaches to efficiently compute minimum energy solutions for respective cost functions based on convex relaxation techniques have been proposed [4, 9, 16, 17]. The idea is to introduce label indicator functions $u_i : \Omega \to \{0, 1\}$ for each label $1 \leq i \leq n$, defined by

$$u_i(x) = \begin{cases} 1 & \text{if } x \in \Omega_i, \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

At each image point $x \in \Omega$ the label uniqueness constraint $\sum_{i=1}^{n} u_i(x) = 1$ is imposed. Energy (5) can now be rewritten in a convex way in terms of u as follows:

First, the (weighted) boundary length $\mathrm{Per}(\Omega_i)$ is given by the weighted total variation of u_i [23, 14]:

$$\frac{\lambda}{2} \mathrm{Per}_g(\Omega_i) = \frac{\lambda}{2} \int_{\Omega} g|\nabla u_i| \, dx = \sup_{\xi_i \in \mathcal{K}} \int_{\Omega} -u_i \, \mathrm{div} \, \xi_i \, dx$$

with $\mathcal{K} := \left\{ \xi \in C_c^1\big(\Omega; \mathbb{R}^2\big) \,\middle|\, |\xi(x)| \leq \frac{\lambda}{2}g(x) \right\}$. Here C_c^1 is the space of smooth functions with compact support and the weighting $g(x) := \exp(-\gamma|\nabla I(x)|)$ for some $\gamma > 0$ favors the coincidence of object and image edges. We set $\gamma = 5$.

Next, the variables $a_i = |\Omega_i|/|\Omega|$ used to replace r_i in (6) can be written in terms of the indicator functions u_i as

$$a_i = \frac{1}{|\Omega|} \int_{\Omega} u_i(x) \, dx. \quad (15)$$

These are linear and thus convex constraints in a and u.

The final step is to relax the binary and therefore non-convex constraints $u_i(x) \in \{0, 1\}$ to the convex ones $u_i(x) \in [0, 1]$. The domain of u becomes $\mathcal{D} := \left\{ u : \Omega \to [0, 1]^n \,\middle|\, \sum_{i=1}^{n} u_i(x) = 1 \; \forall x \in \Omega \right\}$.

Convex Energy. We obtain the following energy

$$\min_{\substack{u \in \mathcal{D} \\ a}} \sup_{\xi_i \in \mathcal{K}} \sum_{i=1}^{n} \int_{\Omega} u_i(f_i - \mathrm{div} \, \xi_i) \, dx - \sum_{i=1}^{n-1} \log \mathcal{P}_p(a_i, a_n)$$
$$\text{s.t. } (15), \quad (16)$$

which is to be minimized w.r.t. u and a and maximized w.r.t. ξ. For the uniform prior the constraints (8) replace the second term. For the Laplace convex relaxation prior it is given by (11), and for the Laplace convex upper bound prior by (12) respectively (13). The relaxed convex optimization problem can be minimized globally yielding results which are independent of the initialization.

3. Implementation

To find the globally optimal solution of the saddle-point optimization problem (16) we employ the fast primal-dual algorithm [3], which guarantees convergence to a global

minimum. Basically, optimization is done by alternating a gradient descent step in u and a and a gradient ascent step in ξ, with reprojections onto the constraint sets, respectively application of more general proximal operators [3].

For the Laplace convex relaxation prior (11), in order to simplify the resulting proximal operators it is convenient to decouple the $a_1, \ldots, a_{n-1}$ from a_n by using the dual representation $E_1 = \sup_{|\alpha_i| \leq \mu/\sigma} \alpha_i(a_i - \bar{r}_i(1 - a_n))$. This introduces new dual variables α_i, $1 \leq i < n$, into the optimization. For the Laplace convex upper bound prior (12), it is also advisory to decouple the a_is to simplify the computations by formulation (13). For the projection of α_i, β_i and the proximal operator for a_n, see supplementary material.

The linear constraints (15) and (8) can be implemented by means of Lagrange multipliers, adding the terms $\sup_{\eta_i \in \mathbb{R}} \eta_i \left(|\Omega| a_i - \int_\Omega u_i(x)\, dx \right)$, respectively e.g. $\sup_{\eta_i \leq 0} \eta_i(a_i - l_i(1 - a_n))$ to the energy. Finally, projection of $\xi \in \mathcal{K}$ is done by clipping of the absolute value, while that of the variable u is a simplex-projection [10].

In order to obtain a binary solution to the original optimization problem, we assign at each pixel x the label $l(x)$ with the maximum value u, i.e. $l(x) = \arg\max_i\{u_i(x)\}$. This yields a solution, which is not globally optimal but within computable bounds of the optimum. Let u^* be the global minimizer of the original binary problem, u^{conv} the solution of the relaxed problem obtained by the proposed algorithm, and $u^{\text{conv}}_{\text{bin}}$ the binarized version of u^{conv}. Then $E(u^{\text{conv}}_{\text{bin}}) - E(u^*) \leq E(u^{\text{conv}}_{\text{bin}}) - E(u^{\text{conv}})$ holds and the right hand side is computable a posteriori.

4. Experiments

We have developed an approach for image sequence segmentation which preserves object proportions by imposing relative size priors on different object components. In this way we allow for arbitrary scaling of the objects, spatially varying color distributions and the recovery of easily missed object components. To evaluate the proposed prior we selected several different image series from the ICoseg-database [1] and the web, requiring that the foreground objects can be subdivided into several parts, see Fig. 3 and 4.

The image series in the ICoseg dataset consist of up to 40 images in each class. To obtain a small set of proportion samples[1] and separate color models for each object part, we created a small training set for each class consisting of five images partitioned into object part labels. The color models f_i in (5) for each part can then either be learned from these images or derived from user scribbles on one or several of them, e.g. by means of the Parzen density estimator [15, 13]. We used the CIELab color space and a Gaussian kernel with standard deviation 8. The remaining

images of the series are then segmented automatically. For most image sequences the parameter set $\lambda = 15$ and $\mu = 200$ yielded optimal results.

4.1. Benchmark Results

In Table 1 we compare the average accuracy (ratio of correctly labeled pixels with respect to total number of pixels) of the proposed proportion priors on the sequences of the ICoseg benchmark, which contain subdivisible objects. First, we compare the proportion prior results to the results of the same algorithm without proportion priors. The table shows drastic improvements of up to 40 or 50%.

We also compare against state-of-the-art image sequence and co-segmentation approaches. Almost all of these approaches are at least partially supervised just as ours since they use training data from a small set of images or scribbles to learn common object properties or classifiers, e.g. Mukherjee et al. [12] learn dictionaries of appearance models, Vicente et al. [21] train a random forest based on appearance model features, Batra et al. [1] learn Gaussian Mixture Models for foreground and background from user scribbles, and Joulin et al. [7] train a supervised classifier for object class separation. Others such as Rubio et al. [20] and Collins et al. [5] do not make use of prior knowledge. However, the former build on an algorithm to measure 'objectness' that again requires learning on general images. Not all of these methods indicate accuracy per sequence or not for all sequences we tested. Missing results are marked by '-'. Collins et al. [5] evaluated only on two sequences of our test set, so we indicate the average score reported in [5], and Batra et al. [1] only gave the average accuracy on the whole benchmark. The results show that we outperform all other methods on the given test set.

Finally, we compare the different types of proportion prior formulations we proposed in Section 2.2. The results show that all three types outperform the other approaches. However, two points need to be mentioned: 1) The uniform prior (8) is too weak in case of outliers since all ratios within the maximum and minimum sample range are equally likely. This can be observed in the kite panda series in the benchmark, where the accuracy drops to 82% since most images contain only parts of the kite with strongly varying proportions (for most other scenes the object is usually fully visible). 2) The relaxed Laplace prior (11) is biased towards large background regions and thus comes with a shrinking bias which tries to minimize the area of the whole foreground object, see (11) and the remark thereafter. Figure 2 illustrates both points. Hence, we can conclude that the convex upper bound Laplace prior (12) yields the best and most stable performance over all test sequences.

---

[1] Note that optimizing for the proportions by alternative minimization is not reasonable: They will be set to match the initial segmentation so that each subsequent segmentation will be driven towards this first (bad) one.

Dataset	w/o proportions	uniform	relaxed L.	bounded L.	[21]	[7]	[20]	[12]	[5]	[1]
Gymnastics	55.13	98.35	97.86	98.51	91.7	90.9	87.1	92.18	-	-
Ferrari	49.71	97.83	97.72	98.19	89.9	85.0	84.3	89.95	96.4	-
Balloon	83.53	96.46	96.74	96.73	90.1	85.2	89.0	95.17	-	-
KitePanda	78.77	82.41	89.69	92.09	90.2	73.2	78.3	93.37	-	-
Baseball	63.01	95.85	96.20	95.01	90.9	73.0	90.5	95.66	-	-
Skating	93.72	96.42	96.55	96.60	77.5	82.1	76.8	96.64	-	-
Liverpool	70.22	95.44	95.55	95.50	87.5	76.4	82.6	-	93.9	-
Ice-skating	56.05	99.33	99.41	99.41	-	-	-	-	-	-
Mean	68.77	95.26	96.22	**96.51**	88.26	80.83	85.7	93.83	93.14	92.8

Table 1. Proportion prior accuracy. Comparison of the accuracy (in %) of the proposed proportion prior formulations (uniform, relaxed Laplace, bounded Laplace prior) to the same algorithm without proportion priors and to state-of-the-art segmentation results on the ICoseg benchmark. The upper bound formulation of the Laplace distribution performs best.
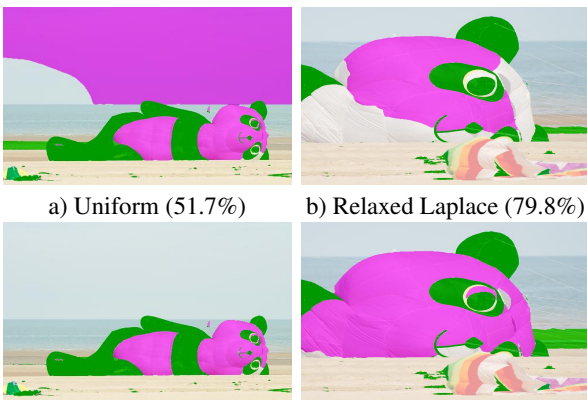
a) Uniform (51.7%)   b) Relaxed Laplace (79.8%)

c) Bounded Laplace (98.8%) d) Bounded Laplace (94.6%)

Figure 2. Proportion prior comparison. a) The uniform prior (8) is weakened in case of strongly varying sample proportions (the panda kite is often only partially visible in the training images). b) The Laplace convex relaxation prior (11) is biased towards larger backgrounds and thus comes with a shrinking bias, which sometimes yields suboptimal results. c+d) The Laplace convex upper bound prior (12) yields stable results.

4.2. Qualitative Results

In this section we show qualitative results for the proportion prior algorithm on the ICoseg dataset, for single objects in Figure 3 and for multiple foreground objects in Figure 4. For each series, we show the result of the proposed segmentation algorithm without proportion priors in the left column and with proportion priors based on the convex upper bound Laplace relaxation (12) in the right column. From the results we can draw four conclusions.

Higher Accuracy. The imposed proportion prior yields segmentation results of much higher accuracy than the original approach without size constraints. In several of the series in Figure 3 it is a common problem for object regions to seep out into the background and cover large parts of it, e.g. the background of the gymnast series is sometimes completely assigned to the 'hair' label, or the ice background is marked as 'skates' in the figure-skating se-

ries. Larger parts of the car background are assigned as various car elements as well, and the sky in the panda kite series is often misclassified as kite. Without the application of proportion priors, further problems appear with different regions of similar color, which are easily mixed up due to low energy differences. Take for example the shoulders of the gymnast in the leftmost image of Figure 3, which are marked as '101-sign'.

Scale Invariance. Depending on the viewpoint of the camera, the objects have different sizes. Take for example the balloon or the car in Figure 3, which appear at different distances to the camera. Absolute size constraints would not allow for a correct segmentation in these cases without changing the size constraints for each image, which would be very tedious for the user. In contrast, relative proportion constraints relating different parts of the object, are naturally invariant with respect to object scale.

Preservation of Small or Elongated Object Parts. Another advantage of the proposed approach is that it preserves semantically meaningful small-scale objects. If these parts are either similar to the background in color or small or elongated, the algorithm without size constraints will remove these parts to minimize the segment boundary length. Examples can be seen in the first and third image of the ice-skater series in Figure 3, where the shoes are assigned to the background. By defining proportion constraints for each object part separately, we can learn and impose ratio likelihoods, which prevent such regions from disappearing.

Multiple Object Instances. Most segmentation algorithms are limited to single foreground objects. Since the proportion priors within the multilabel framework naturally apply to multiple object instances in the same image we can also apply our approach to image series with many foreground objects such as baseball or soccer scenarios, see Figure 4. In addition, proportion priors are not limited to different parts of the same object but can also relate different objects, e.g. the size of a mouse with respect to a dog or the heart with respect to the lungs in medical imaging.

| Image | Without proportion prior | With proportion prior |

Figure 3. Segmentation results for different image sequences containing a single object. The left three columns of each sequence show the segmentation results without proportion prior, the right three columns after imposing proportion priors.

4.3. Runtime

An important advantage of the proposed method is its efficiency, especially for large sequences of images. Previous and current state-of-the-art approaches demand runtimes per image of 45 seconds [7], 10 seconds [5] or 25–100 seconds [12]. Due to the inherent parallel structure of the algorithm, each pixel can be updated independently. Hence, the proposed method can be easily parallelized. Using a single NVIDIA GTX 680 GPU we obtained an average runtime of 2.2 seconds per image. Using three GPUs in parallel we reduced the average runtime to one second.

5. Conclusion

We introduced the concepts of part decompositions and proportion priors into a framework for multilabel co-segmentation. The problem is formulated as a variational approach together with a convex relaxation which can be globally optimized. Extensive evaluations on various image series show that proportion priors provide accurate segmentations despite changing illumination, despite viewpoint or background changes, and despite substantial articulations and non-rigid deformations. The relativity of the proportion constraints allows for stable, scale-invariant segmentations over long ranges of images (see suppl. material) containing single or multiple foreground objects and helps to recover small-scale semantically important object parts which are frequently suppressed in existing approaches. With an average runtime of about one second on graphics hardware the algorithm is more than an order of magnitude faster than existing techniques.

References

[1] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen. icoseg: Interactive co-segmentation with intelligent scribble guid-

| | | |
| Image | Without proportion prior | With proportion prior |

Figure 4. Segmentation results for different image sequences containing multiple objects (see Figure 3 for explanation).

ance. In *Proc. CVPR*, 2010. 1, 2, 5, 6

[2] A. Chambolle, D. Cremers, and T. Pock. A convex approach for computing minimal partitions. Tech. rep. TR-2008-05, Univ. of Bonn, 2008. 2

[3] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.*, 40:120–145, 2011. 4, 5

[4] T. Chan, S. Esedoḡlu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.*, 66(5):1632–1648, 2006. 4

[5] M. Collins, J. Xu, L. Grady, and V. Singh. Random walks based multi-image segmentation: Quasiconvexity results and gpu-based solutions. In *Proc. CVPR*, 2012. 2, 5, 6, 7

[6] D. Hochbaum and V. Singh. An efficient algorithm for co-segmentation. In *Proc. ICCV*, 2009. 2

[7] A. Joulin, F. Bach, and J. Ponce. Discriminative clustering for image co-segmen-tation. In *Proc. CVPR*, 2010. 1, 2, 5, 6, 7

[8] M. Klodt and D. Cremers. A convex framework for image segmentation with moment constraints. In *Proc. ICCV*, 2011. 2

[9] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. Tech. rep., Univ. Heidelberg, 2008. 2, 4

[10] C. Michelot. A finite algorithm for finding the projection of a point onto the canonical simplex of R^n. *J. Optimiz. Theory Appl.*, 50, 1986. 5

[11] L. Mukherjee, V. Singh, and C. Dyer. Half-integrality based algorithms for cosegmentation of images. In *Proc. CVPR*, 2009. 1, 2

[12] L. Mukherjee, V. Singh, J. Xu, and M. Collins. Analyzing the subspace structure of related images: Concurrent segmentation of image sets. In *Proc. ECCV*, volume 7575, pages 128–142, 2012. 1, 2, 5, 6, 7

[13] C. Nieuwenhuis and D. Cremers. Spatially varying color distributions for interactive multi-label segmentation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 2013. 5

[14] C. Nieuwenhuis, E. Toeppe, and D. Cremers. A survey and comparison of discrete and continuous multilabel segmentation approaches. *Int. J. of Comp. Vis.*, 2013. 4

[15] E. Parzen. On the estimation of a probability density function and the mode. *Annals of Math. Stat.*, 33:1065–1076, 1962. 5

[16] T. Pock, A. Chambolle, H. Bischof, and D. Cremers. A convex relaxation approach for computing minimal partitions. In *Proc. CVPR*, Miami, Florida, 2009. 4

[17] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the piecewise smooth mumford-shah functional. In *ICCV*, 2009. 4

[18] T. Riklin Raviv, N. Sochen, and N. Kiryati. Shape-based mutual segmentation. *Int. J. of Comp. Vis.*, 79:231–245, 2008. 2

[19] C. Rother, T. Minka, A. Blake, and V. Kolmogorov. Cosegmentation of image pairs by histogram matching-incorporating a global constraint into MRFs. In *Proc. CVPR*, 2006. 1, 2

[20] J. C. Rubio, J. Serrat, A. M. Lopez, and N. Paragios. Unsupervised co-segmentation through region matching. In *Proc. CVPR*, pages 749–756, 2012. 1, 2, 5, 6

[21] S. Vicente, C. Rother, and V. Kolmogorov. Object cosegmentation. In *Proc. CVPR*, 2011. 1, 2, 5, 6

[22] A. Yezzi, L. Zöllei, and T. Kapur. A variational framework to integrate segmentation and registration through active contours. *Medical Image Analysis*, 7:171–185, 2003. 2

[23] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. In *Workshop VMV*, October 2008. 2, 4

Proportion Priors for Image Sequence Segmentation

Claudia Nieuwenhuis	Evgeny Strekalovskiy	Daniel Cremers
UC Berkeley, ICSI, USA	TU Munich, Germany	TU Munich, Germany

A. Convex Relaxation for the Laplace Prior

In Section 2.2.2 of the paper, the Laplace distribution proportion prior energy

$$E_p(r_i) = \frac{\mu}{\sigma_i} |r_i - \bar{r}_i| = \frac{\mu}{\sigma_i} \left| \frac{a_i}{1 - a_n} - \bar{r}_i \right| \qquad (1)$$

is introduced and the following is stated:

Proposition 1. *The convex relaxation of* (1) *on the domain* $a_i, a_n \geq 0$ *and* $a_i + a_n \leq 1$ *is given by*

$$E_1(a_i, a_n) := \frac{\mu}{\sigma_i} |a_i - \bar{r}_i(1 - a_n)|. \qquad (2)$$

Here we give a proof of this proposition:

Proof. W.l.o.g. let $\frac{\mu}{\sigma_i} = 1$. First, E_1 is a convex lower bound on E_p since it is convex with $E_1 = E_p \cdot (1 - a_n) \leq E_p$. For any other such bound $\widehat{E}_1$, by $\widehat{E}_1 \leq E_p$ it follows

$$\widehat{E}_1(0, a_n) \leq \bar{r}_i, \qquad (3)$$
$$\widehat{E}_1(\bar{r}_i(1 - a_n), a_n) \leq 0. \qquad (4)$$

From this, $\widehat{E}_1(0, 0) \leq \bar{r}_i$ and $\widehat{E}_1(0, 1) \leq 0$, and therefore

$$\widehat{E}_1(0, a_n) \leq a_n \widehat{E}_1(0, 1) + (1 - a_n) \widehat{E}_1(0, 0) \\ \leq \bar{r}_i(1 - a_n). \qquad (5)$$

For $a_i \leq \bar{r}_i(1 - a_n)$ we can define $\alpha := \frac{a_i}{\bar{r}_i(1 - a_n)} \in [0, 1]$. By convexity of $\widehat{E}_1$, and from (5) and (4) we get

$$\widehat{E}_1(a_i, a_n) = \widehat{E}_1((1 - \alpha) \cdot 0 + \alpha \cdot \bar{r}_i(1 - a_n), a_n) \\ \leq (1 - \alpha) \widehat{E}_1(0, a_n) + \alpha \widehat{E}_1(\bar{r}_i(1 - a_n), a_n) \\ \leq (1 - \alpha) \cdot \bar{r}_i(1 - a_n) + \alpha \cdot 0 \\ = \bar{r}_i(1 - a_n) - a_i = E_1(a_i, a_n).$$

Similarly, one can show $\widehat{E}_1 \leq E_1$ also for $a_i \geq \bar{r}_i(1 - a_n)$. Thus, E_1 is the greatest convex lower bound on E_p. $\square$

B. Implementation Details

In Section 2.2.2 of the paper we give the following dual formulation for the convex upper bound of the Laplace distribution prior E_2:

$$E_2 = \sup_{\alpha, \beta} \sum_{i=1}^{n-1} \Big(\alpha_i \big(a_i - \bar{r}_i(1 - a_n)\big) - \beta_i(1 - a_n) \Big) \\ + \frac{\varepsilon(n - 1)}{1 - a_n}. \qquad (6)$$

The duals α, β are constrained to be in the convex set

$$A := \left\{ (\alpha, \beta) \in \mathbb{R}^{2(n-1)} \mid \beta_i \geq \frac{\varepsilon \sigma_i^2}{\mu^2} \alpha_i^2 \quad \forall 1 \leq i < n \right\}. \qquad (7)$$

Section 3 of the paper contains implementation details for the employed primal-dual algorithm.

B.1. Proximal Operator for a_n

For the primal-dual algorithm, in each iteration one must compute the proximal operator

$$\arg \min_{a_n} \left\{ \frac{(a_n - a_n^0)^2}{2\tau} + \frac{\varepsilon(n - 1)}{1 - a_n} \right\}, \qquad (8)$$

where $a_n^0 \in \mathbb{R}$ and $\tau > 0$ are constants. Setting the derivative w.r.t. a_n to zero, one has to solve a cubic equation. We use the method of [1] for this. Define $c := \tau \varepsilon(n - 1)$, $v := \frac{1 - a_n^0}{3}$, $w := v^3$ and $D := \frac{c}{4} + w$.

The case $D \geq 0$. In this case the solution is given by

$$a_n = 1 - v - z - \frac{v^2}{z} \qquad (9)$$

with $z := \sqrt[3]{\frac{c}{2} + w + \sqrt{cD}} > 0$.

The case $D < 0$. Otherwise, the solution is

$$a_n = 1 - v + 2v \cos \left(\tfrac{1}{3} \arccos \left(1 - \tfrac{2D}{w} \right) \right). \qquad (10)$$

B.2. Proximal Operator for α, β

The proximal operator is here

$$\underset{(\alpha,\beta)\in A}{\arg\min} \sum_{i=1}^{n-1} \frac{(\alpha - \alpha_i^0)^2}{2\tau} + \sum_{i=1}^{n-1} \frac{(\beta - \beta_i^0)^2}{2\tau} \qquad (11)$$
$$+ \sum_{i=1}^{n-1} \left(-\alpha_i \bar{r}_i - \beta_i \right).$$

for some $\alpha_i^0, \beta_i^0 \in \mathbb{R}$ and $\tau > 0$ with the set A in (7). Define $\widehat{\alpha}_i := \alpha_i^0 + \tau \bar{r}_i$ and $\widehat{\beta}_i := \beta_i^0 + \tau$. Then the solution is given by the projection onto a parabola, separately for each i:

$$(\alpha_i, \beta_i) = \text{proj}_{\beta_i \geq \widehat{\varepsilon}_i \alpha_i^2} \left(\widehat{\alpha}_i, \widehat{\beta}_i \right) \qquad (12)$$

with $\widehat{\varepsilon}_i := \frac{\varepsilon \sigma_i^2}{\mu^2}$. Considering the optimality conditions for this projection leads to a cubic equation, which we again solve by the method of [1]:

If already $\widehat{\beta}_i \geq \widehat{\varepsilon}_i \widehat{\alpha}_i^2$, the solution is $(\alpha_i, \beta_i) = \left(\widehat{\alpha}_i, \widehat{\beta}_i \right)$. Otherwise, with $a := 2\widehat{\varepsilon}_i |\widehat{\alpha}_i|$, $b := \frac{2}{3}(1 - 2\widehat{\varepsilon}_i \widehat{\beta}_i)$ and $d := a^2 + b^3$ set

$$v := \begin{cases} c - \frac{b}{c} & \text{with } c = \sqrt[3]{a + \sqrt{d}} & \text{if } d \geq 0, \\ 2\sqrt{-b} \, \cos\left(\frac{1}{3} \arccos \frac{a}{\sqrt{-b}^3} \right) & \text{if } d < 0. \end{cases} \qquad (13)$$

If $c = 0$ in the first case, set $v := 0$. The solution is then given by

$$\alpha_i = \begin{cases} \frac{v}{2\widehat{\varepsilon}_i} \frac{\widehat{\alpha}_i}{|\widehat{\alpha}_i|} & \text{if } \widehat{\alpha}_i \neq 0 \\ 0 & \text{else} \end{cases}, \quad \beta_i = \widehat{\varepsilon}_i \alpha_i^2. \qquad (14)$$

References

[1] J. P. McKelvey. Simple transcendental expressions for the roots of cubic equations. *Amer. J. Phys.*, 52(3):269–270, 1984. 1, 2

[2] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. Global solutions of variational models with convex regularization. *SIAM J. Imaging Sci.*, 3:1122–1145, 2010.