Uncertainty in Minimum Cost Multicuts for Image and Motion Segmentation

Amirhossein Kardoost[1]   **Margret Keuper**[1]

[1]Data and Web Science Group, University of Mannheim, Germany

Abstract

The minimum cost lifted multicut approach has proven practically good performance in a wide range of applications such as image decomposition, mesh segmentation, multiple object tracking and motion segmentation. It addresses such problems in a graph-based model, where real valued costs are assigned to the edges between entities such that the minimum cut decomposes the graph into an optimal number of segments. Driven by a probabilistic formulation of minimum cost multicuts, we provide a measure for the uncertainties of the decisions made during the optimization. We argue that the access to such uncertainties is crucial for many practical applications and conduct an evaluation by means of sparsifications on three different, widely used datasets in the context of image decomposition (BSDS-500) and motion segmentation (DAVIS$_{2016}$ and FBMS$_{59}$) in terms of variation of information (VI) and Rand index (RI).

1 INTRODUCTION

The minimum cost (lifted) multicut problem (Chopra and Rao [1993]), also known as correlation clustering, has been widely applied in computer vision for applications ranging from image segmentation (Keuper et al. [2015b]) to multiple person tracking (Keuper et al. [2020]) and pose estimation (Pishchulin et al. [2016]). The formulation is flexible and can easily be adapted to a new clustering problem. Specifically entities are represented by nodes in a graph and real valued costs are assigned to edges where positive costs indicate that the adjacent nodes are similar and negative costs indicate that the adjacent nodes are dissimilar. The minimum cost multicut then provides a decomposition of the graph into an optimal number of segments. Andres et al. [2012], relate the minimum cost multicut to a probabilistic

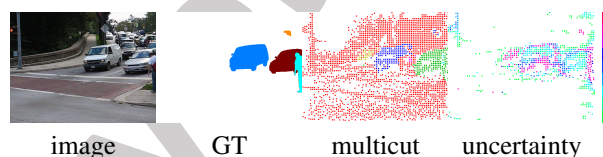
image   GT   multicut   uncertainty

Figure 1: Motion segmentation and the proposed uncertainty measure on a street scene. The uncertainty is high on incorrectly segmented points, specifically on the missed person.

model, which was further extended in Keuper et al. [2015b] to *lifted* multicuts. They show that, if the edge costs are set according to logits of the cut probability between nodes, the minimum cost multicut provides the maximum a posteriori probability (MAP) estimate. Motivated by this finding, edge cost definitions based on cut probability estimates have become common practice, even in settings where the final solution is estimated using primal feasible heuristics Keuper et al. [2015a], Beier et al. [2015] to account for the NP-hardness of the problem (Bansal et al. [2004]).

In this paper, we argue that considering solely the (approximate) MAP solution is not satisfying in all scenarios. For example in an automotive setting, it might be required to assess the uncertainty in the prediction of the number of moving objects and their shapes. Therefore, we propose to employ the probabilistic model from Andres et al. [2012] and derive, for a proposed solution, a measure for the uncertainty of each node-to-label assignment. An example is provided in Fig. 1 for sparse motion segmentation. Based on sparse point trajectories, we apply the motion model derived by Ochs et al. [2014] and provide the minimum cost multicut solution (Keuper et al. [2015a]). Depending on the exact parameters employed, the person standing nearby the street can not be correctly segmented. Yet, the uncertainty indicates potential mistakes in the prediction.

The proposed uncertainty measure is directly derived from the probabilistic multicut formulation and can be applied to *any* given decomposition. We evaluate our approach in the

context of minimum cost multicuts for sparse motion segmentation using the model from Keuper et al. [2015a] on the datasets FBMS$_{59}$ Ochs et al. [2014] and DAVIS$_{2016}$ Perazzi et al. [2016]. Further, we investigate potential benefits of the predicted uncertainties for the generation of dense motion segmentations from sparse ones. Such densifications can be computed using convolutional neural networks trained in a self supervised way for example using the approach by Kardoost et al. [2020]. Last, we evaluate the proposed uncertainty measure in the context of minimum cost *lifted* multicuts for image decomposition on BSDS-500 Arbelaez et al. [2011]. In both applications, motion and image segmentation, we show via sparsification plots that subsequently removing uncertain predictions from the solution improves segmentation metrics. The proposed measure is thus a robust indicator for the uncertainty of a given solution.

2   RELATED WORK

We first review prior work related to the minimum cost multicut problem, then, we summarize related work w.r.t. the uncertainty estimation and last, we summarize prior work on multicuts in the considered application domains.

Correlation Clustering   The minimum cost multicut problem Chopra and Rao [1993], also known as correlation clustering, is a binary *edge* labeling problem defined on a graph with real valued edge costs. The feasible solutions of the multicut problem propose decompositions of the graph and the optimal solution corresponds to the maximum a posteriori probability (MAP) estimate Andres et al. [2012]. Yet, the problem is shown to be APX-hard Bansal et al. [2004]. While optimal solutions or solutions within bounds of optimallity can be found for small instances Demaine et al. [2006], Swoboda and Andres [2017], Andres et al. [2011, 2012], most practical applications depend on heuristic solvers Keuper et al. [2015b], Beier et al. [2014, 2015]. Our proposed uncertainty measure is defined on the formulation of the minimum cost (*lifted*) multicut problem and can be applied on any given solution. We perform a thorough evaluation in the context of the widely used KLj and GAEC heuristics from Keuper et al. [2015b] (compare the supplementary material for GAEC). Probabilistic clustering and segmentation algorithms, such as the minimum cost (*lifted*) multicut have been studied on different applications, such as motion segmentation Keuper et al. [2020], image decomposition Keuper et al. [2015a], Andres et al. [2011], Kappes et al. [2016b], Kardoost and Keuper [2018], multiple object tracking Keuper [2017], Horňáková et al. [2017], Tang et al. [2017], Hornakova et al. [2020], connectomics Beier et al. [2017], cell tracking Rempfler et al. [2017] and instance segmentation Kirillov et al. [2017]. For all of these applications, one can easily imagine use-cases that benefit from a measure of uncertainty. We evaluate the proposed approach on image and motion segmentations.

Uncertainty Estimation   Kohli and Torr [2006] proposed a measure of uncertainty for the graph cut solutions using the min-marginals associated with the label assignments in a MRF. With respect to minimum cost multicuts, Kappes et al. [2016a] measure the uncertainty of image partitions by an approximate marginal distribution using cost perturbations and induced *edge* label flips. The proposed approach is complementary to this method as it does not measure the uncertainty of a binary edge labeling but assesses the uncertainty of the induced *node* labeling. In Tomczak et al. [2018], the Perturb and MAP (PM) approach is used for learning a restricted Boltzmann machine. In this method, each of the observable and hidden variables are flipped in order to check the change in the energy function. We propose a measure of uncertainty on the cut/join decisions of the graph in minimum cost (*lifted*) multicut formulation. To show the applicability of our uncertainty measure in real-world scenarios we evaluate on two *motion segmentation* and one *image decomposition* benchmarks.

Motion Segmentation   Formulations of minimum cost multicut problems have been successfully used for motion segmentation for example in Keuper et al. [2015a], Keuper [2017]. The goal in this application is to segment motion patterns of the foreground objects with respect to the scene and irrespective of the camera motion, scaling movements and out-of-plane rotation of the objects Ochs et al. [2014]. One widely used paradigm to tackle this problem is to define spatio-temporal curves, called *motion trajectory*. The trajectories are created by tracking the points through consecutive frames using the optical flow estimation. Due to the high computational complexity, one often considers sparse motion segmentation (Ochs et al. [2014]), where point trajectories are not sampled at every point but at a defined density. Therefore, the segmentations do not cover all the image pixels. The trajectories are then cast to the nodes in a graph and their affinities are used to compute costs on the edges. In this setting, the solution to the motion segmentation problem is sparse. There are different methods to provide dense motion segmentations from the sparse results, like the variational approach from Ochs et al. [2014] and the self-supervised deep learning based approach from Kardoost et al. [2020]. We study the proposed uncertainty measure on motion segmentation in the datasets FBMS$_{59}$ Ochs et al. [2014] and DAVIS$_{2016}$ Perazzi et al. [2016]. Additionally, we use uncertainties to improve the training of the densification model of Kardoost et al. [2020].

Image Decomposition.   For image decomposition, the minimum cost multicut problem is defined over sets of pixels or superpixels for example in Arbelaez et al. [2011], Keuper et al. [2015b], Kappes et al. [2016b, 2011], Andres et al. [2011]. In this scenario, the pixels act as nodes in the graph and their connectivity defined by edges Andres et al. [2011], i.e. directly neighboring pixels are connected. The extension of this problem using *lifted* mulitucts is de-

fined in Keuper et al. [2015b], where lifted edges are used to encode long-range information while the connectivity of the original graph is preserved. This can lead to an improved pixel-level clustering behavior. Kappes et al. [2016b] use a higher order multicut model on superpixels for the task. Orbanz and Buhmann Orbanz and Buhmann [2008] provide a non-parametric Bayesian model for histogram clustering to determine the number of image segments, utilizing Dirichlet process mixture model. Cutting plane and integer programming techniques are used in Kappes et al. [2011] for image decomposition. Further they proposed an approximate solution by solving a polynomial LP.

3 UNCERTAINTIES IN MINIMUM COST (LIFTED) MULTICUTS

In this section, we formally define the minimum cost multicut problem (MP) (Chopra and Rao [1993]) and its generalization, the minimum cost *lifted* multicut problem (LMP) (Keuper et al. [2015b]). Then, we summarize the probabilistic model from Andres et al. [2012] and deduce the proposed uncertainty measure.

3.1 THE MINIMUM COST MULTICUT PROBLEM

Given a graph $G = (V, E)$, a cost function $c : E \to \mathbb{R}$ and *edge* labels $y : E \to \{0, 1\}$, the optimization problem in (1) is an instance of the minimum cost multicut problem (MP) with respect to the graph G and real-valued costs c

$$\min_{y \in \{0,1\}^E} \sum_{e \in E} c_e y_e \tag{1}$$

$$s.t. \quad \forall C \in cycles(G) \quad \forall e \in C : y_e \leq \sum_{e' \in C \setminus \{e\}} y_{e'}.$$

The edge labeling y induces a decomposition of graph G, which is ensured by the inequality constraints stated over all cycles of G. Chopra and Rao [1993] showed that is sufficient to consider all chordless cycles. The minimum cost multicut problem allows to assign a cost (c_e) for every edge $e \in E$, where a positive cost encourages the edge to be joined while a negative cost encourages the edge to be cut.

3.2 MINIMUM COST LIFTED MULTICUTS

The minimum cost *lifted* multicut problem (LMP) is defined with respect to a graph $G = (V, E)$ and a lifted graph $G' = (V, E')$ with $E \subseteq E'$ and a cost function $c' : E' \to \mathbb{R}$ (Keuper et al. [2015b]). The cost function c' allows to assign to every edge in E' a cost for being cut. The decompositions of graph G relate one-to-one to the feasible solutions of the problem, similar to the minimum cost multicut problem.
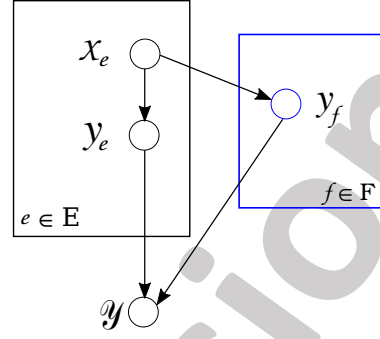

Figure 2: Bayesian Network from Keuper et al. [2015b], defining a set of probability measures on multicuts (black) and *lifted* multicuts (blue).

For any undirected graph $G = (V, E)$, any $F = \binom{V}{2} \setminus E$ and any $c' : E' = E \cup F \to \mathbb{R}$, the linear program written in Eq. (2) - (5) is an instance of the minimum cost *lifted* multicut problem with respect to the graph G, lifted edges F and the edge costs c' Keuper et al. [2015b].

$$\min_{y \in \{0,1\}^{E'}} \sum_{e \in E'} c'_e y_e \tag{2}$$

$$s.t. \quad \forall C \in cycles(G) \quad \forall e \in C : y_e \leq \sum_{e' \in C \setminus \{e\}} y_{e'} \tag{3}$$

$$\forall vw \in F \quad \forall P \in vw\text{-paths}(G) : y_{vw} \leq \sum_{e \in P} y_e \tag{4}$$

$$\forall vw \in F \quad \forall C \in vw\text{-cuts}(G) : 1 - y_{vw} \leq \sum_{e \in C} (1 - y_e) \tag{5}$$

The linear inequalities in Eq. (3) - (5) constrain y such that $\{e \in E | y_e = 1\}$ is a multicut of G. Further, they ensure that, for any edge $uv \in F, y_{uv} = 0$ if and only if there exists a path (uv-path) in the original graph G, along which all edges are connected, i.e. labeled as 0. The set F of lifted edges is utilized to modify the cost function of the multicut problem without changing the set of feasible solutions. The lifted edges can thus connect non-neighboring nodes in graph G and are commonly used to introduce non-local information.

3.3 PROBABILITY MEASURES

Andres et al. [2012] show that minimum cost multicuts can be derived from a Bayesian model (see Fig.2) and their solutions correspond to the maximum a posteriori estimates. In the following, we summarize this probabilistic model and its extension to lifted multicuts from Keuper et al. [2015b], upon which we build the proposed uncertainty estimation.

Probability Measures on MP and LMP For a graph $G = (V, E)$, Andres et al. [2012] assume that likelihoods $p_{\mathcal{X}_e | \mathcal{Y}_e}$ are computed based on an affinity definition of the nodes in graph G. Further, the costs assigned to the edges are assumed to be independent of each other and the topology of the graph G. Moreover, the prior $p_{(\mathcal{Y}_e)}$ is assumed to be identical for all edges $e \in E$ and is specified by a value $\beta \in (0, 1)$, so that $p_{\mathcal{Y}_e = 1} = \beta$ and $p_{\mathcal{Y}_e = 0} = 1 - \beta$. In this setting, they show that the minimum cost multicut maximizes the posterior probability $p_{\mathcal{Y} | \mathcal{X}, \mathcal{Y}}$ of a joint labeling $y \in \{0, 1\}^{|E|}$. By definition, the Maximum a Posteriori (MAP) is then

$$p_{\mathcal{Y}|\mathcal{X},\mathcal{Y}} \propto p_{\mathcal{Y}|y} \cdot \prod_{e \in E} p_{\mathcal{Y}_e | \mathcal{X}_e} \cdot p_{\mathcal{Y}_e} \quad . \tag{6}$$

for the MP, Fig. 2 (the black part) (Andres et al. [2012]), and it can be extended to the LMP, Fig. 2 (the blue part) as follows (Keuper et al. [2015b])

$$p_{\mathcal{Y}|\mathcal{X},\mathcal{Y}} \propto p_{\mathcal{Y}|y} \cdot \prod_{e \in E} p_{\mathcal{Y}_e | \mathcal{X}_e} \cdot p_{\mathcal{Y}_e} \cdot \prod_{f \in F} p_{\mathcal{Y}_f | \mathcal{X}_E} \cdot p_{\mathcal{Y}_f} \quad . \tag{7}$$

In both cases, $p_{\mathcal{Y}|y}$ indicates the feasibility of a solution, i.e.

$$p_{\mathcal{Y}|y}(Y_{E'}, y) \propto \begin{cases} 1 & \text{if } y \in Y_{E'} \\ 0 & \text{otherwise} \end{cases} \quad . \tag{8}$$

where $E' = E$ for the MP and $E' = E \cup F$ for the LMP.

Equations (6) and (7) can be maximized by minimizing their negative log-likelihoods. This leads to the definition of instances of the MP (Eq. 1) and LMP (Eq. 2) by setting edge costs according to

$$\forall e \in E: \quad c_e = \log \frac{p_{\mathcal{Y}_e | \mathcal{X}_e}(0, x_e)}{p_{\mathcal{Y}_e | \mathcal{X}_e}(1, x_e)} + \log \frac{1 - \beta}{\beta} \quad . \tag{9}$$

The value of the scalar β is the cut prior and assumed to be 0.5 for a bias-free case.

3.4 UNCERTAINTY ESTIMATION MODEL

Based on the probabilistic formulation of the MP (6) and LMP (7), we can study the uncertainty of given feasible solutions. More specifically, we aim to assign to every node in V a confidence reflecting the certainty of the assigned label. The simplest attempt to this goal would be to directly assess the posterior probability of the solution. In the following, we briefly sketch this baseline approach before introducing the proposed measure.

Baseline Approach For every node $v_i \in V$ we propose, as a baseline approach, to draw an uncertainty measure from

the proxi to the posterior probability (e.g. Eq. (6)). Thus, the confidence of the given label A of node v_i is

$$\prod_{e=(v_i,v_j),v_j \in A} p_{\mathcal{Y}_e | \mathcal{X}_e}(0, x_e) \cdot \prod_{e=(v_i,v_j),v_j \in V \setminus A} p_{\mathcal{Y}_e | \mathcal{X}_e}(1, x_e). \tag{10}$$

This measure intuitively aggregates the local join probabilities of all edges that are adjacent to v_i and joined (set to 0) in the current solution, and the local cut probabilities of all edges adjacent to v_i and cut (set to 1) in the current solution. Accordingly, Eq. (10) yields a low value if the local probabilities for the given solution are low.

A potential issue with this simple approach from Eq. (10) is that it under-estimates the confidence in many practical scenarios, especially when the local probabilities are uncertain and the connectivity is dense. Then, the product of local probabilities will issue a low value even if all local cues agree. Therefore, we describe in the following the derivation of the proposed, calibrated uncertainty measure.

Uncertainty Estimation. Given an instance of the (lifted) multicut problem and its solution, we employ the probability measures in equations (6) (*MP*) and (7) *(LMP)*. We iterate through nodes $v_i \in \{1, \ldots, |V|\}$ in vicinity of a cut, i.e. $\exists e \in \mathcal{N}_E(v_i)$ with $e \in E$ and $y_e = 1$. Assuming that v_i belongs to segment A and its neighbour v_j according to E belongs to the segment B, the amount of cost change γ_B is computed in the linear cost function (defined in (1) and (2)) by moving v_i from cluster A to cluster B as

$$\gamma_B = \sum_{v_j \in \mathcal{N}_{E'}(v_i) \cap A} c_{(v_i, v_j)} - \sum_{v_j \in \mathcal{N}_{E'}(v_i) \cap B} c_{(v_i, v_j)}. \tag{11}$$

Thus, in γ_B, we accumulate all costs of edges from v_i that are not cut in the current decomposition and subtract all costs of edges that are cut in the current decomposition but would not be cut if v_i is moved from A to B. Note that, while the cost change is computed over all edges in E' for lifted graphs, only the uncertainty of nodes with an adjacent cut edge in E can be considered in order to preserve the feasibility of the solution. For each node v_i the number of possible moves depends on the labels of its neighbours $\mathcal{N}_E(v_i)$, and Eq. (11) allows us to assign a cost to any such node-label change. Altogether, we assess the uncertainty of a given node label by the cheapest, i.e. the most likely, possible move

$$\gamma_i = \min_B \gamma_B. \tag{12}$$

and set γ_i to ∞ if no move is possible. The minimization in Eq. (12) corresponds to considering the local move of v_i which maximizes

$$\prod_{e=(v_i,v_j),v_j \in A} \frac{p_{\mathcal{Y}_e | \mathcal{X}_e}(1, x_e)}{p_{\mathcal{Y}_e | \mathcal{X}_e}(0, x_e)} \cdot \prod_{e=(v_i,v_j),v_j \in B} \frac{p_{\mathcal{Y}_e | \mathcal{X}_e}(0, x_e)}{p_{\mathcal{Y}_e | \mathcal{X}_e}(1, x_e)}. \tag{13}$$

To produce an uncertainty measure for each node in the graph, we apply the logistic function on (12)

$$\text{uncertainty} = \frac{1}{1 + \exp\left(-\gamma_i\right)} \qquad (14)$$

as it is the inverse of the logit function used in the cost computation in (9). As shown in the supplementary material, this uncertainty measures the below ratio

$$\frac{\prod\limits_{e,A} p_{\mathcal{Y}_e|\mathcal{X}_e}(0, x_e) \cdot \prod\limits_{e,B} p_{\mathcal{Y}_e|\mathcal{X}_e}(1, x_e)}{\prod\limits_{e,A} p_{\mathcal{Y}_e|\mathcal{X}_e}(0, x_e) \cdot \prod\limits_{e,B} p_{\mathcal{Y}_e|\mathcal{X}_e}(1, x_e) + \prod\limits_{e,B} p_{\mathcal{Y}_e|\mathcal{X}_e}(0, x_e) \cdot \prod\limits_{e,A} p_{\mathcal{Y}_e|\mathcal{X}_e}(1, x_e)} \qquad (15)$$

where the nominator is the product of the local posterior probabilities of the observed solution A at node v_i (compare Eq. (6) for $p_{\mathcal{Y}_e}$ const.) and is proportional to the posterior of the chosen node label if v_i has at most two labels in the local neighborhood. The denominator sums trivially to one in the case of $|\mathcal{N}_{E'}(v_i)| = 1$. For the more common case of $|\mathcal{N}_{E'}(v_i)| \geq 1$, expression (15) can be interpreted as a "calibrated" probability. It normalizes the posterior probability of the MAP solution with the sum of posteriors of this solution and the second most likely one. Intuitively, if all solutions have a low absolute posterior probability, the relatively best one can still have a high certainty.

4 EVALUATION AND RESULTS

We evaluate the proposed uncertainty measure in two common application scenarios of minimum cost multicuts: motion segmentation and image segmentation. For motion segmentation, we conduct experiments on the datasets Freiburg-Berkeley Motion Segmentation FBMS$_{59}$ (Ochs et al. [2014]) and Densely Annotated Video Segmentation (DAVIS$_{2016}$) (Perazzi et al. [2016]). On both datasets, we compute point trajectories as well as the segmentation using the method from Keuper et al. [2015a], which computes edges between point trajectories as pseudo-probabilities from a simple motion model.

Additionally, we highlight the potential benefit of a robust uncertainty measure: It not only allows us to produce several hypotheses of solutions. We can also employ the estimated uncertainties in the densification framework from Kardoost et al. [2020] to compute improved, dense video segmentations from sparse ones.

Last, we evaluate the proposed uncertainty measure on the image segmentation task posed by the BSDS-500 dataset by Arbelaez et al. [2011], employing the minimum cost lifted multicut instances from Keuper et al. [2015b].

Freiburg-Berkeley Motion Segmentation The FBMS$_{59}$ (Ochs et al. [2014]) dataset contains 29 train and 30 test sequences. They are between 19 to 800 frames long
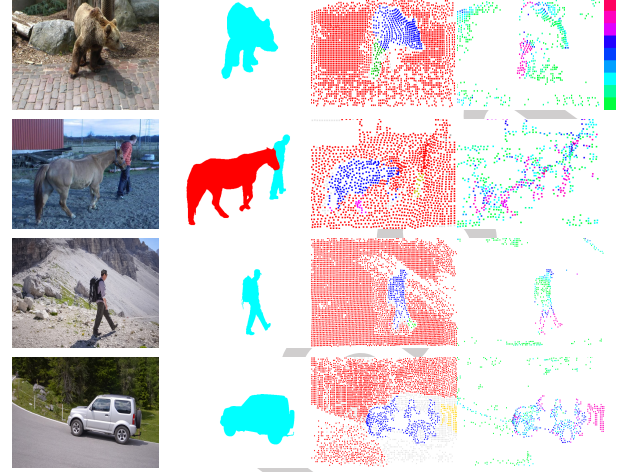

Figure 3: Uncertainty measure on point trajectories on two sequences from FBMS$_{59}$ (first two rows) and DAVIS$_{2016}$ (last two rows). In each row, from left to right, we provide an image, its ground-truth, the segmentation and our uncertainty estimation. The uncertainty values are discretized and color coded for visualization purposes. White areas correspond to the trajectories with high certainty. The uncertainty on thin, articulated object parts is high.

and show motion of possibly multiple foreground objects. The sequences contain camera shaking, rigid/non-rigid object motion as well as occlusion/dis-occlusion events. The ground-truth segmentation is provided for a subset of the frames in each sequence.

Densely Annotated Video Segmentation Dataset The DAVIS$_{2016}$ dataset by Perazzi et al. [2016] has been introduced for binary video object segmentation. It contains 30 train and 20 validation sequences where objects undergo occlusion and dis-occlusion and rigid as well as non-rigid movements. The pixel-wise binary ground truth segmentation is provided per frame for each sequence. The DAVIS$_{2016}$ dataset has been proposed for *video instance segmentation*. Yet, as each sequence only contains one annotated object with eminent motion with respect to the scene, the dataset can be used to evaluate motion segmentation tasks. Our method is evaluated on train and test sequences of FBMS$_{59}$ and train and validation sequences of DAVIS$_{2016}$.

Berkeley Segmentation Dataset The BSDS-500 dataset (Arbelaez et al. [2011]) consists of 200 train, 200 test and 100 validation images, where for each images five different human made annotations are provided. We evaluate our method on the test images.

Parameter Setting In the MP and LMP, the value β (as in the Eq. (9)) is set to 0.5 (unbiased) to be comparable with the baseline methods in the evaluated datasets. The LMP on images requires the value τ to be set, which corresponds to
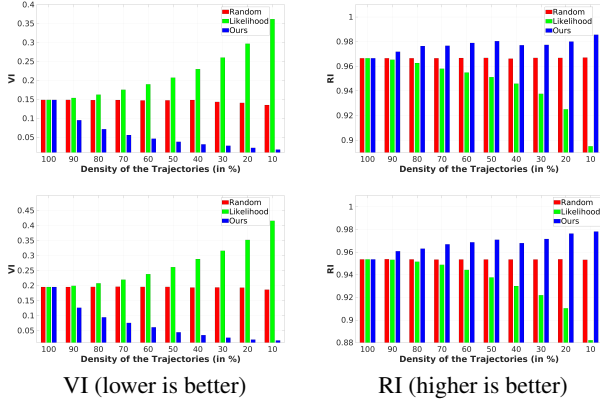
VI (lower is better)   RI (higher is better)

Figure 4: Study on motion trajectory uncertainties on VI (*left*) and RI (*right*) on the train set (top row) and on the test set (bottom row) of FBMS$_{59}$. The metrics improve by removing trajectories according to the proposed uncertainty measure. Notice that removing uncertain trajectories according to the likelihood baseline deteriorates both VI and RI.

the radius until which to insert lifted edges. We set $\tau = 20$ as in Keuper et al. [2015b].

Metrics   In both applications, we evaluate our uncertainty measure based on the variation of information (VI) (Meilă [2007]) and Rand index (RI) (Meilă [2007]). We study the effect of measured uncertainties by means of sparsification. The most uncertain nodes are removed subsequently and in each step VI and RI are measured. Ideally, VI should drop and RI should increase monotonically as more uncertain nodes are removed. In Meilă [2007] it is shown that VI is a more reliable indicator than the RI with respect to the density of the results, since the RI depends heavily on the granularity of the decomposition. Specifically, the more pixels or motion trajectories are removed during sparsification, the less reliable the RI becomes, which is why both metrics are reported in all our experiments. It is important to notice that high uncertainty of a node indicates that the label assigned to the node after the termination of the solver has a tendency to flip. Therefore, removing those nodes from the decomposition should improve the quality of the decomposition in terms of VI and RI. We compared our approach with the proposed baseline approaches.

4.1   MOTION SEGMENTATION RESULTS

To study the proposed uncertainty measure, we first compute minimum cost multicuts on the motion segmentation instances of FBMS$_{59}$ and DAVIS$_{2016}$. On these decompositions, we compute the proposed uncertainties. In Fig. 3, we depict example images from sequences of FBMS$_{59}$ and DAVIS$_{2016}$, their ground truth segmentation, trajectory segmentation and the proposed uncertainty. The uncertainty

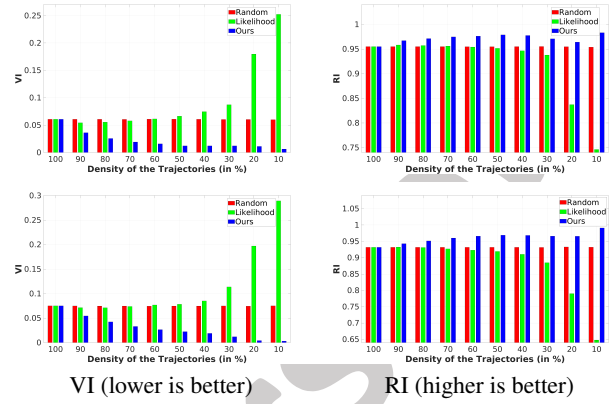VI (lower is better)   RI (higher is better)

Figure 5: Study on the motion trajectory uncertainties on VI (*left*) and RI (*right*) on train set (top row) and validation set (bottom row) of DAVIS$_{2016}$ Perazzi et al. [2016]. Our results improve significantly over the baselines.

measures are in the range of 0 (lowest uncertainty) to 1 (highest uncertainty). As can be seen, the sparse segmentations show good overall accuracy but tend to have issues on dis-occlusion areas for example behind the driving car in the last row, and on object parts under articulated motion such as the legs of the horse in the second row and the legs of the the person in the third row. In all these regions, the estimated uncertainty is high. Next, we assess the proposed uncertainties in terms of sparsification plots as done in Ilg et al. [2018], i.e. by subsequently removing nodes in the order of their decreasing uncertainty from the solution. We compare our results to the *Likelihood* baseline defined in Eq. (10) as well as to a random baseline. In the random baseline, the trajectories are removed subsequently at *random*. In this case, neither improvement or decay of the segmentation metric should be expected.

FBMS$_{59}$ The Variation of Information (VI) and the Rand index (RI) are provided in Fig. 4 and FBMS$_{59}$-train and FBMS$_{59}$-test. The most uncertain trajectories are removed until 10% of the density of the trajectories. Interestingly, removing highly uncertain trajectories accounts for reduction in VI (lower is better) and improves the RI indicating the effectiveness of the proposed uncertainty measure. The VI continuously improves as more uncertain trajectories are removed. Yet, the RI becomes unsteady when getting closer to the lowest density. Recall that the RI depends on the granularity of the solution (see Meilă [2007]) so this behavior is to be expected as more and more ground truth segments are not considered in the sparse solution.

DAVIS$_{2016}$ The evaluation on the DAVIS$_{2016}$ train and validation sets in Fig. 5 indicates a similar behaviour as seen on FBMS$_{59}$ and shows the effect of our approach. Notice that our approach provides better uncertainty measures than

image   uncertainty

-127910   -127894   -127894   -127205
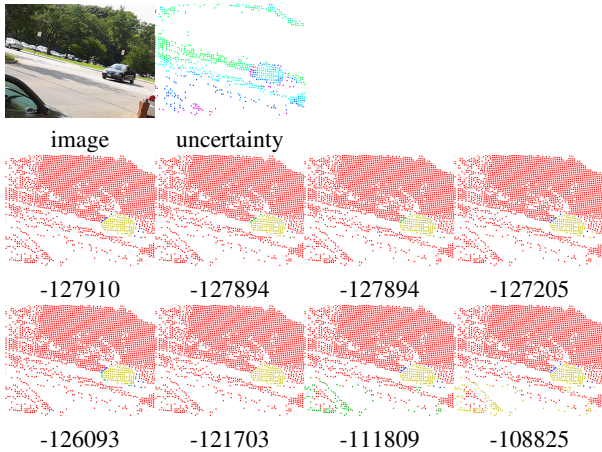
-126093   -121703   -111809   -108825

Figure 6: Visualization of eight different likely solutions and their energies (Eq. (2), lower is better), as they can be generated by the proposed method. The different solution candidates vary mostly along object boundaries. The best segmentation w.r.t. the ground truth corresponds to the second image (from left) in the last row.
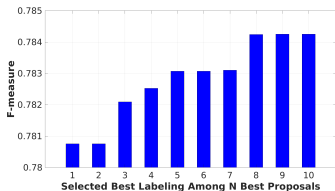

Figure 7: F-measure on the train set of $FBMS_{59}$ when selecting n best segmentation proposals. The F-measure improves as the number of segmentation candidates increases.

the baselines (*Random Baseline* and *Likelihood Approach* (refer to 3.4)). Again, the likelihood baseline even shows an increase in VI during sparsification, proving the importance of using the denominator in the Eq. (15).

4.2   MULTIMODAL MOTION SEGMENTATION

In the following, we briefly sketch how the proposed approach can be leveraged to generate multiple, likely solutions. After the termination of the heuristic solver, the nodes in the graph $G = (V, E)$ are moved between different clusters to compute the cost differences in Eq. (11). Given the N clusters, the node from cluster A is moved to other $N-1$ clusters and the cost change is computed. Therefore, for each node, a vector with the number of different partitions minus one $(N-1)$ is computed. This leads to $\frac{N \times (N-1)}{2}$ vectors, each with $|V|$ values. To produce multiple segmentations, such vectors are ordered based on their magnitude, i.e. their associated energy increase. Each such vector holds

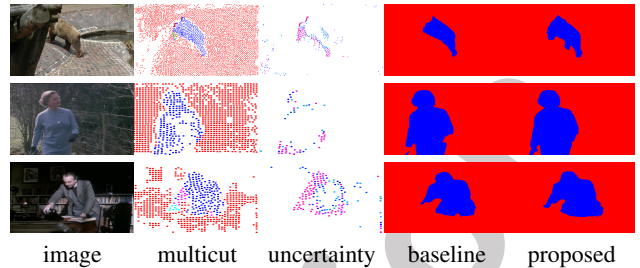image   multicut   uncertainty   baseline   proposed

Figure 8: Densification of sparse segmentations using uncertainties. We compare the result from Kardoost et al. [2020] (baseline) with the proposed method, which uses uncertainties in the model training. Improvements can be observed especially on thin structures such as limbs.

Table 1: Densification of the sparse trajectory segmentations on the $FBMS_{59}$ train set. We compare the model of Kardoost et al. [2020] to the variant trained using uncertainties.

	Precision	Recall	F-measure
Kardoost et al. [2020]	89.35	67.67	77.01
Ours (uncertainty-aware)	88.17	68.96	77.40

the costs for moving nodes from cluster A to cluster B. The label of the node v is changed from cluster A to cluster B if $\Delta cost(A, B)_v \approx 0$. With this approach, it is possible to generate n best clusterings, which differ in the labels of uncertain points. In Fig. 6, we visualize an example of eight such potential solutions and their energies. The variance is strongest on the object boundaries and on the car in the foreground. It coincides with the estimated uncertainties. In Fig. 7, we allow the evaluation script of $FBMS_{59}$ to choose the best among the n most likely solutions for $n \in \{1, ..., 10\}$. By taking more solutions into account, the F-measure improves.

4.3   DENSIFIED MOTION SEGMENTATION

To create actual pixel-level segmentations from sparse segmentations a separate model needs to be learned. For example Ochs et al. [2014] propose a variational model while Kardoost et al. [2020] employs a U-Net Ronneberger et al. [2015] based model, which is trained in a self-supervised manner, using the sparsely segmented trajectories as labels. In the following, we employ our uncertainties in order to enhance the self-supervised training signal in this model in the FBMB59 dataset. As a pre-processing, Kardoost et al. [2020] remove small segments from the training data such that the remaining uncertain points are mostly on thin, articulated object parts. For each segment A in the decomposition of graph G, the mean uncertainty is computed as μ_A. The highly uncertain nodes, where $u_v > \mu_A$, in the highly uncertain segments $\mu_A > \phi$ are assumed to be hard examples

for the network and are therefore used more often than other points during training (with factor 1000). In Tab. 1, we report Precision, Recall and F-measure of the proposed training and the original model from Kardoost et al. [2020]. Using this simple trick, the F-Measure can be improved by 0.4%. Qualitative results are provided in Fig. 8.

4.4 IMAGE DECOMPOSITION

As a last application, we evaluate the proposed uncertainty measure on the image segmentation task in the context of minimum cost lifted multicuts. The evalutation is computed on the BSDS500 dataset using the problem instances and solutions from Keuper et al. [2015b]. Fig. 9 visualizes the node (pixel) level uncertainty on several example images from the BSDS-500 Arbelaez et al. [2011] dataset. As expected, the node level uncertainty is higher along the object boundaries, where label changes are likely to happen.

Removing the most uncertain pixels corresponds to removing object boundaries as it is depicted in Fig. 10 for an example. As the segmentation becomes sparser, entire object parts such as the tail of the squirrel will get removed from the solution. In Fig. 11, the VI and RI metrics are provided with respect to the pixel densities. As expected, removing highly uncertain pixels results in increasing the RI and decreasing VI. Thus, an improvement can be observed in both metrics. Again, we compare our to the probability driven baseline and to a random baseline. Unlike before, the probability driven baseline performs reasonably in this setting. Yet, the proposed measure can achieve a faster decrease in VI and a higher increase in RI and thus shows more robust behavior. The relatively good performance of the baseline in comparison to its results on motion segmentation can be explained by the way local cut probabilities are computed in this setting. Specifically, they are derived by Keuper et al. [2015b] from edge maps and thus have consistently low values in homogeneous regions. Therefore, the normalization by the denominator in the proposed measure (refer to Eq. (15)) becomes less important.

5 CONCLUSION

The minimum cost *lifted* multicut problem has been widely studied in different computer vision and data analysis applications such as motion segmentation and image decomposition. Due to the probabilistic behavior of the multicut problem the final decomposition of the nodes in the proposed graph can be seen as relaxed decisions rather than hard decisions on label assignment to the nodes of the graph. We studied this probabilistic behavior and provide an informative uncertainty measure in the node uncertainties. We showed that removing uncertain nodes according to the proposed measure leads to an improvement in the variation of information (VI) and the Rand index (RI) in two appli-
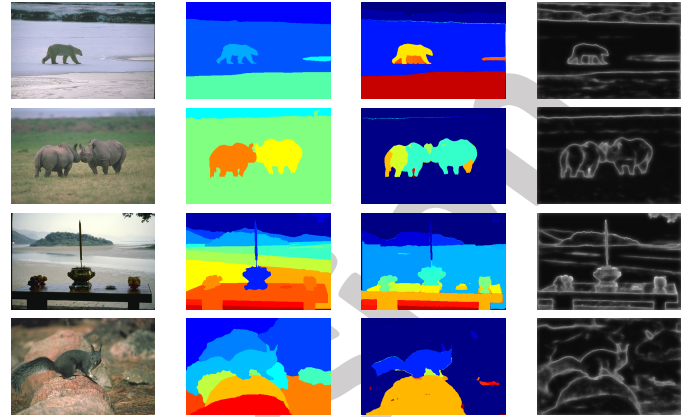

Figure 9: Exemplary images and segmentation uncertainties on the BSDS-500 Arbelaez et al. [2011] dataset. In each row, from left to right, the original images, ground-truth segmentation, the resulting minimum cost lifted multicut segmentation and the proposed uncertainties are given. Bright areas in the uncertainty images represent uncertain pixels.
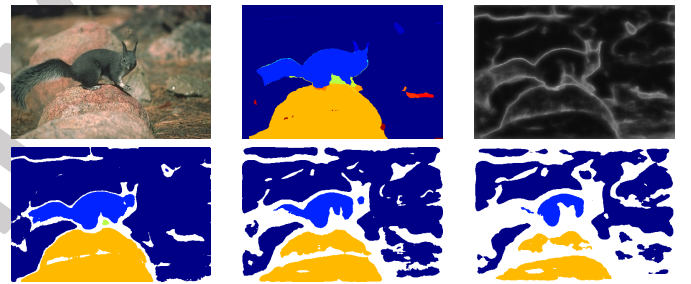

Figure 10: Visualization of removing uncertain pixels in BSDS-500 images (Arbelaez et al. [2011]). Notice that removing uncertain pixels corresponds to removing pixels along the object boundaries. The original image (left), its multicut solution (middle) and the uncertainty measure (right) based on our model are shown in the first row. The second row visualizes (from left to right) removing 10, 30 and 50 % most uncertain pixels.

cations of motion segmentation and image decomposition. Further, we showed the application of such uncertainties to train a self-supervised model for motion segmentation. The proposed uncertainty measure can be combined with any minimum cost multicut based formulation.
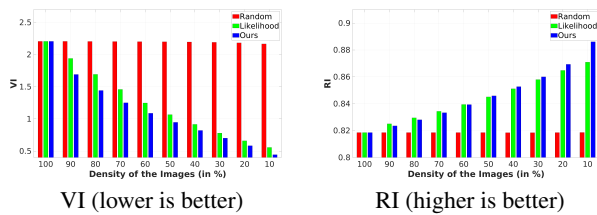
VI (lower is better)   RI (higher is better)

Figure 11: Sparsification analysis in VI (*left*) and RI (*right*) on the BSDS-500 (Arbelaez et al. [2011]) test data. The proposed method shows a faster decrease in VI than the baseline and reaches a higher RI.

References

B. Andres, J. H. Kappes, T. Beier, U. Köthe, and F. A. Hamprecht. Probabilistic image segmentation with closedness constraints. In *ICCV*, 2011.

B. Andres, T. Kröger, K. L. Briggman, W. Denk, N. Korogod, G. Knott, U. Köthe, and F. A. Hamprecht. Globally optimal closed-surface segmentation for connectomics. In *ECCV*, 2012.

P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE TPAMI*, 33(5):898–916, May 2011. ISSN 0162-8828.

N. Bansal, A. Blum, and S. Chawla. Correlation clustering. *Machine Learning*, 56(1):89–113, Jul 2004. ISSN 1573-0565.

T. Beier, T. Kroeger, J. H. Kappes, U. Köthe, and F. A. Hamprecht. Cut, glue, & cut: A fast, approximate solver for multicut partitioning. In *CVPR*, 2014.

T. Beier, F. A. Hamprecht, and J. H. Kappes. Fusion moves for correlation clustering. In *CVPR*, 2015.

T. Beier, C. Pape, N. Rahaman, T. Prange, B. Stuart, D. Bock, A. Cardona, G. W. Knott, S. M. Plaza, L. K. Scheffer, K. Ullrich, A. Kreshuk, and F. A. Hamprecht. Multicut brings automated neurite segmentation closer to human performance. *Nature Methods*, 14:101–102, 2017.

S. Chopra and M. R. Rao. The partition problem. *Mathematical Programming*, 59(1–3):87–115, 1993.

E. D. Demaine, D. Emanuel, A. Fiat, and N. Immorlica. Correlation clustering in general weighted graphs. *Theoretical Computer Science*, 361(2–3):172–187, 2006.

A. Hornakova, R. Henschel, B. Rosenhahn, and P. Swoboda. Lifted disjoint paths with application in multiple object tracking. 07 2020.

A. Horňáková, J. H. Lange, and B. Andres. Analysis and optimization of graph decompositions by lifted multicuts. In *ICML*, 2017.

E. Ilg, Ö. Çiçek, S. Galesso, A. Klein, O. Makansi, F. Hutter, and T. Brox. Uncertainty estimates and multi-hypotheses networks for optical flow. In *ECCV*, 2018. https://arxiv.org/abs/1802.07095.

J. Kappes, P. Swoboda, B. Savchynskyy, T. Hazan, and C. Schnörr. Multicuts and perturb & map for probabilistic graph clustering. *Journal of Mathematical Imaging and Vision*, 56, 01 2016a.

J. H. Kappes, M. Speth, B. Andres, G. Reinelt, and C. Schnörr. Globally optimal image partitioning by multicuts. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 31–44. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-23094-3.

J. H. Kappes, M. Speth, G. Reinelt, and C. Schnörr. Higher-order segmentation via multicuts. *Computer Vision and Image Understanding*, 143c, 2016b.

A. Kardoost and M. Keuper. Solving minimum cost lifted multicut problems by node agglomeration. In *ACCV*, Perth, Australia, 2018.

A. Kardoost, K. Ho, P. Ochs, and M. Keuper. Self-supervised sparse to dense motion segmentation. In *ACCV*, November 2020.

M. Keuper. Higher-order minimum cost lifted multicuts for motion segmentation. In *ICCV*, Oct 2017.

M. Keuper, B. Andres, and T. Brox. Motion trajectory segmentation via minimum cost multicuts. In *ICCV*, 2015a.

M. Keuper, E. Levinkov, N. Bonneel, G. Lavoué, T. Brox, and B. Andres. Efficient decomposition of image and mesh graphs by lifted multicuts. *ICCV*, pages 1751–1759, 2015b.

M. Keuper, S. Tang, B. Andres, T. Brox, and B. Schiele. Motion segmentation multiple object tracking by correlation co-clustering. *TPAMI*, 42(1):140–153, 2020.

A. Kirillov, E. Levinkov, B. Andres, B. Savchynskyy, and C. Rother. Instancecut: From edges to instances with multicut. In *CVPR*, pages 7322–7331, 2017.

P. Kohli and P. H. S. Torr. Measuring uncertainty in graph cut solutions – efficiently computing min-marginal energies using dynamic graph cuts. In *ECCV*, pages 30–43, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-33835-2.

M. Meilă. Comparing clusterings—an information based distance. *J. Multivar. Anal.*, 98(5):873–895, May 2007. ISSN 0047-259X.

P. Ochs, J. Malik, and T. Brox. Segmentation of moving objects by long term video analysis. *IEEE TPAMI*, 36(6): 1187 – 1200, Jun 2014.

P. Orbanz and J. Buhmann. Nonparametric bayesian image segmentation. *ICCV*, 77:25–45, 05 2008.

F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. pages 724–732, 06 2016.

L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. Gehler, and B. Schiele. Deepcut: Joint subset partition and labeling for multi person pose estimation. In *CVPR*, pages 4929–4937, Las Vegas, NV, USA, 2016. IEEE. ISBN 978-1-4673-8851-1.

M. Rempfler, J. H. Lange, F. Jug, C. Blasse, E. Myers, B. Menze, and B. Andres. Efficient algorithms for moral lineage tracing. pages 4705–4714, 10 2017.

O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. (available on arXiv:1505.04597 [cs.CV]).

P. Swoboda and B. Andres. A message passing algorithm for the minimum cost multicut problem. In *CVPR*, 2017.

S. Tang, M. Andriluka, B. Andres, and B. Schiele. Multiple people tracking by lifted multicut and person re-identification. In *CVPR*, pages 3701–3710, Washington, DC, USA, July 2017. IEEE Computer Society.

J. Tomczak, S. Zareba, S. Ravanbakhsh, and R. Greiner. Low-dimensional perturb-and-map approach for learning restricted boltzmann machines. *Neural Processing Letters*, 10 2018.