Les Cahiers du GERAD ISSN: 0711–2440

Data association via set packing for computer vision applications

J. Yarkony, Y. Adulyasak,
M. Singh, G. Desaulniers

G–2019–42

June 2019

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2019
– Bibliothèque et Archives Canada, 2019

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2019
– Library and Archives Canada, 2019

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Data association via set packing for computer vision applications

Julian Yarkony [a]

Yossiri Adulyasak [b,c]

Maneesh Singh [a]

Guy Desaulniers [b,d]

[a] *Verisk AI, Jersey City, New Jersey, USA, 07310–1686*

[b] *GERAD, Montréal (Québec), Canada, H3T 2A7*

[c] *Department of Logistics and Operations Management, HEC Montréal, Montréal (Québec), Canada, H3T 2A7*

[d] *Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montréal (Québec) Canada, H3C 3A7*

julian.yarkony@verisk.com
yossiri.adulyasak@hec.ca
msingh@verisk.com
guy.desaulniers@gerad.ca

June 2019
Les Cahiers du GERAD
G–2019–42

Abstract: Significant progress has been made in the field of computer vision, due to the development of supervised machine learning algorithms, which efficiently extract information from high-dimensional data such as images and videos. Such techniques are particularly effective at recognizing the presence or absence of entities in the domains, where labeled data is abundant. However, supervised learning is not sufficient in applications where one needs to annotate each unique entity in crowded scenes respecting known domain specific structures of those entities. This problem, known as data association, provides fertile ground for the application of combinatorial optimization. In this paper, we present the computer vision applications, namely, multi-person tracking, multi-person pose estimation, and multi-cell segmentation, which can be formulated as integer linear programs with a massive number of variables. In order to solve this problem, column generation algorithms are applied to circumvent the need to enumerate all variables explicitly. To enhance the solution process, we provide a general approach for applying subset-row inequalities to tighten the formulations, and introduce novel dual optimal inequalities to reduce the dual search space. The proposed algorithms and their enhancements are successfully applied to solve the three aforementioned computer vision problems and achieve superior performance compared to benchmark approaches.

Keywords: Data association, computer vision, set packing, column generation

Acknowledgments: We would like to thank the stockholders of Verisk for supporting this work.

1 Introduction

Artificial neural networks (ANN) (Rumelhart et al., 1985) excel at learning functions that map input data vectors (e.g., images of objects and living beings) to output semantic labels (dog, horse, car, etc.) using large amounts of labeled training data. An ANN learns a function, that generalizes beyond the training data set, so as to produce the correct label as output, on test data not part of the training data set. One popular application of ANNs is object recognition in which an ANN learns to recognize the presence of objects in images. Large data sets facilitate learning such functions and include the image-net data set (Deng et al., 2009), which provides fourteen million training images, each associated with the labels of the objects present in the image.

Localizing each unique instance of objects, which is called instance segmentation (Silberman et al., 2014), in crowded images is an important related task to object recognition. The naive approach to instance segmentation iterates over all possible rectangles of pixels (also called bounding boxes) in the image, and predicts the presence of each object in that rectangle. However, combining the hypotheses generated in each rectangle to describe each unique instance of objects, which we refer as data association, is challenging as the hypotheses need not be mutually consistent. For example, multiple predicted hypotheses may share a common pixel but clearly multiple objects can not be associated with the same pixel in the ground truth. Heuristics called non-max suppression (Dalal and Triggs, 2005) are often used to remove conflicts between predicted hypotheses. Non-max suppression removes from consideration all but one of each set of "similar" and/or overlapping predictions. Combinatorial optimization provides a principled alternative to non-max suppression heuristics (Desai et al., 2011).

Data association can use combinatorial optimization to partition the observations in a data set (e.g., pixels in an image) into a set of hypotheses (e.g., unique instances of objects or background), each associated with a subset of the observations, that are consistent with the statistical properties of the known structure of hypothesis. We motivate the use of data association with the examples of multi-person tracking and pose estimation, which are important for self-driving car and personal robot assistant applications. Here the set of observations is the set of all pixels, and the set of possible hypotheses is the power set of pixels. The statistical support for a hypothesis is defined in terms of how well a classifier (such as an ANN) scores the quality of a single person dominating the corresponding pixels.

The use of combinatorial optimization in computer vision/machine learning has developed largely without influence from the operations research community, and has focused on network flows (called graph cuts (Boykov and Kolmogorov, 2004)), primal dual methods (the most prominent of which is message passing (Sontag et al., 2008; Kolmogorov, 2006; Komodakis et al., 2007)), and compact linear programming (LP) relaxations augmented with cutting planes (Andres et al., 2011; Pishchulin et al., 2016; Insafutdinov et al., 2016). This often leads to less efficient solvers than desirable. However, and perhaps more importantly, the capacity of the associated approaches is limited by ignoring decades of research in combinatorial optimization in the operations research community.

Recently the operations research techniques of column generation (CG) (Gilmore and Gomory, 1961; Barnhart et al., 1996) and (nested) Benders decomposition ((N)BD) (Birge, 1985; Benders, 1962) have been introduced to the machine learning and computer vision communities (Yarkony and Fowlkes, 2015; Wang et al., 2017b, 2018, 2017a; Yarkony et al., 2012; Wang et al., 2017c; Zhang et al., 2017). However, the application of these techniques and the construction of models to support the use of CG and (N)BD is in its infancy. The goal of this document is to introduce data association problems from the computer vision community to the operations research community so as to catalyze joint research efforts.

In this paper, we focus on the very flexible minimum weight set packing (MWSP) formulation (Karp, 1972) of data association. A MWSP instance is parameterized by a set of possible hypotheses, each of which is associated with a real valued cost that describes the sensibility of the belief that the members of the hypothesis correspond to a common cause. MWSP then selects the least total cost set

of hypotheses, such that no two selected hypotheses share a common observation. Observations that are not included in any selected hypothesis define the set of false observations, which are observations not-explained by our model, and can be thought of as noise.

This document makes the following contributions to research in combinatorial optimization for computer vision. First, we introduce a common and comprehensive treatment of the literature on MWSP formulations for computer vision problems, targeted to an operations research audience. Our paper provides a framework and motivating examples to allow the combinatorial optimization community to apply their methodologies to these problems. Second, our paper extends the work of Wang et al. (2017b) to produce a general approach for applying subset-row inequalities (Jepsen et al., 2008), so as to exploit the initial structure of the pricing problem. Third, we introduce novel dual optimal inequalities (DOIs, Ben Amor et al., 2006) that are tighter than the current baseline for MWSP formulations in computer vision. These DOIs depend on the current set of columns in the restricted master problem (RMP), and loosen with the addition of new columns to the RMP. They are easy to compute and provably never looser than the baseline.

We outline our document as follows. In Section 2, we review the existing literature on combinatorial optimization in the context of computer vision. In Section 3, we discuss compact and extended formulations of MWSP problems, along with the solution of extended formulations via CG. In Section 4, we describe the computer vision applications that we consider and provide application-specific MWSP formulations for data association. In Section 5, we consider pricing for CG in the context of our applications. In Section 6, we apply subset-row inequalities to tighten the LP relaxation of the MWSP formulations. In Section 7, we consider the use of DOIs to bound the dual variables and accelerate optimization. In Section 8, we provide computational results for our problem domains regarding computation time, tightness of bounds, and accuracy relative to the ground truth. In Section 9, we conclude and discuss extensions.

2 Literature review

The successful solution of large-scale integer linear programs (ILPs) requires solving LP relaxations that give good approximations to the convex hull of feasible integer solutions. Compact relaxations in terms of the number of variables/constraints of classic ILPs are often associated with extremely loose relaxations and high levels of symmetry so that branch-and-bound operations do not tend to rapidly tighten them (Barnhart et al., 1996). To attack such ILPs successfully, LP relaxations that a have huge number of variables are employed. The cardinality of such variable sets is often too large to enumerate them and, much less, to consider them in a linear program.

For certain problem classes, CG (Dantzig and Wolfe, 1960; Gilmore and Gomory, 1961; Desaulniers et al., 2005) is used to solve to optimality the LP relaxation over a huge number of variables, without explicitly enumerating most of them. The use of CG often leads to both fast optimization and (near) tight relaxations. CG proceeds as follows. It iterates between solving an LP defined over a subset of the variables (initialized heuristically or as empty) and adding variables with negative reduced costs to that subset. This process iterates until no negative reduced cost variables exist. The identification of negative reduced cost variables, which is called pricing, often corresponds to solving a dynamic program or another tractable integer program. CG is commonly applied in diverse domains including: vehicle routing (Desrochers et al., 1992; Desaulniers et al., 1998; Costa et al., 2019), crew scheduling (Gamache et al., 1999; Kasirzadeh et al., 2017), material cutting (Gilmore and Gomory, 1961; Delorme et al., 2016), and web search (Abrams et al., 2007).

More recently CG has been used in computer vision applications including: multi-object tracking (Wang et al., 2017b; Leal-Taixe et al., 2012), multi-person pose estimation (Wang et al., 2017a,c), multi-cell instance segmentation (Zhang et al., 2017), and (hierarchical) image segmentation (Yarkony et al., 2012; Yarkony and Fowlkes, 2015; Zhang et al., 2014b). As in operations research domains, the purpose of CG in computer vision is to circumvent the loose LP relaxations, which are often induced by compact formulations.

2.1 Mixed integer programming in computer vision/machine learning

Below, we provide a mapping from approaches in operations research to applications of those approaches in computer vision.

2.1.1 Network flows

Network flow techniques (Boykov and Kolmogorov, 2004) are used to solve ILP formulations of the task of providing each pixel in an image with a semantic label (dog, horse, car, etc.). The underlying model exploits the statistical observation that pixels in close proximity tend to share the same label. The use of network flow techniques is not restricted to pixel labeling. Indeed, they have been used for tasks including multi-object tracking (Zhang et al., 2008; Butt and Collins, 2013).

2.1.2 Dual ascent methods

Many tasks in machine learning, including protein structure prediction, are formulated as ILPs where the objective consists of minimizing the sum of functions over pairs of variables (Sontag et al., 2008). The set of functions is enumerable (though large), but the ILP has no easily exploitable special structure to facilitate optimization. To attack these problems, coordinate/sub-gradient ascent methods in the Lagrangian dual (Kolmogorov, 2006; Komodakis et al., 2007) are used. The corresponding LP relaxation can be tightened using cutting planes (Sontag et al., 2008) that facilitate the use of coordinate/sub-gradient ascent methods.

2.1.3 Correlation clustering methods with cutting planes

Many problems in computer vision can be formulated as correlation clustering problems, including image segmentation (breaking an image into semantically meaningful parts) (Andres et al., 2011), and connectomics (producing the wiring diagram of the brain) (Andres et al., 2012). Optimization is attacked using a linear programming/cutting plane method.

2.1.4 Column generation

In Wang et al. (2017b), CG is applied to the problem of tracking many people (or objects) as they move in video. CG is employed to solve a MWSP formulation, where elements correspond to detections of people in frames of video, and sets correspond to complete tracks of people moving across time. The cost of a set is real valued and is produced using a K-th order Markov model of the person as he moves across space-time. Pricing is solved by dynamic programming. The LP relaxation of the MWSP formulation is not tighter than that of a compact formulation. However, the former formulation permits the use of subset-row inequalities (Jepsen et al., 2008) to tighten the relaxation. Wang et al. (2017b) solves a tighter relaxation than the baseline dual-decomposition approach of Butt and Collins (2013), yielding faster computational times. In Leal-Taixe et al. (2012), CG is employed in a branch-and-price framework for tracking objects across multiple cameras. The corresponding pricing problem is an unstructured ILP.

In Yarkony et al. (2012) and Yarkony and Fowlkes (2015), CG is applied to image segmentation, on the planar problems found commonly in computer vision. In Yarkony et al. (2012), image segmentation is formulated as correlation clustering on a planar graph, where nodes correspond to (super) pixels (Ren and Malik, 2003) and edges indicate adjacency. Correlation clustering is described by a superposition of 2-colorable partitions of the planar graph and optimization is attacked using CG. The corresponding pricing problem is a max-cut problem on a planar graph, which is solved fast via a reduction to minimum-cost perfect matching (Shih et al., 1990). In addition, Yarkony et al. (2012) independently develop DOIs (though they are not referred to as such), which accelerate optimization dramatically. In Wang et al. (2017c) an extended formulation of multi-person pose estimation solvable by CG is presented. The skeletons of people and the descriptions of their body parts correspond to separate sets of variables. Their corresponding pricing problems are dynamic programs and small-scale ILPs, respectively.

2.1.5 Benders decomposition

In Wang et al. (2017a), Benders decomposition is employed to solve the same problem as in Wang et al. (2017c). The master problem corresponds to creating the skeletons of people, and is solved using CG. There is one Benders subproblem for each body part, which provides the descriptions for that body part to the skeletons assembled in the master problem. Because the subproblems are independent of each other, Benders multicuts are added to the master problem. A variant of Magnanti and Wong (1981)'s cuts is presented though not labeled as such.

In Wang et al. (2018), an extended formulation solved by CG is proposed for multi-person pose estimation. This work improves on that of Wang et al. (2017c) by jointly modeling the skeletons of people and the descriptions of their body parts. However, the corresponding pricing problem is a dynamic program with a huge state space that is challenging to solve. To circumvent this difficulty, Wang et al. (2018) solves the pricing problem using NBD and employs a variant of the Magnanti-Wong cuts. The NBD algorithm exploits Benders cuts generated in previous calls to the pricing problem, and is shown to accelerate CG substantially.

2.2 Learning cost terms

For the problems discussed in the preceding section, the cost coefficients in the objective functions of the ILPs are often produced by training a standard linear classifier to determine the probability that a variable/pair of variables takes on a given label/pair of labels (Wang et al., 2018; Zhang et al., 2017). The output probabilities are converted to cost terms by taking the negative logarithm of the probabilities (Insafutdinov et al., 2016). However, this is not a mathematically principled approach since it does not consider the use of the ILP.

To correctly model the use of the ILP, structured support vector machines (SVM) are employed (Tsochantaridis et al., 2005). A structured SVM learns a mechanism to produce cost terms for ILPs, such that the optimal solution to that ILP is similar to the ground truth. The structured SVM takes as input features selected according to a fixed recipe that does not learn, and has proven challenging to integrate into ANN frameworks, which dominate modern machine learning. Learning a structured SVM from large amounts of labeled data is modeled as a quadratic program and solved using a cutting plane approach. It, thus, requires repeatedly solving ILPs (or LPs), making learning on large data sets challenging.

Computing cost terms is, however, not the focus of this paper as we assume that they are provided. For detailed studies on the calculation of cost terms, we refer the reader to Insafutdinov et al. (2016); Pishchulin et al. (2016) and Wang and Fowlkes (2015).

3 Compact and extended formulations

This section introduces and motivates the use of extended representations for MWSP problems, and their solution via CG. In Section 3.1, we describe the correlation clustering problem (Bansal et al., 2004), which can be seen as a basic data association problem. In Section 3.2, we formulate correlation clustering as a compact ILP with an easily enumerated set of variables and constraints. This ILP is shown to have a very weak LP relaxation. In Section 3.3, we introduce an extended MWSP formulation for correlation clustering. The LP relaxation of this formulation is tighter than that of the compact formulation. In Section 3.4, we use CG to solve the extended formulation.

3.1 Correlation clustering

Consider a graph with node set $\mathcal{D}$ and edge set $\mathcal{E}$, where edge $(d_1, d_2) \in \mathcal{E}$ has a weight $\theta_{d_1 d_2} \in \mathbb{R}$. Correlation clustering partitions the nodes in $\mathcal{D}$ into clusters so as to minimize the sum of the weights of the intra-cluster edges, i.e., those linking nodes in a same cluster. Correlation clustering is known to be NP-Hard (Bansal et al., 2004).

3.2 Compact form of correlation clustering

Let us formulate correlation clustering as a compact ILP. Let $\mathcal{J} = \{1, 2, \ldots, |\mathcal{D}|\}$ be the set of possible clusters. We use decision variables $x \in \{0,1\}^{|\mathcal{D}|\times|\mathcal{J}|}$, where $x_{dj} = 1$ if and only if node $d \in \mathcal{D}$ is in cluster $j \in \mathcal{J}$. To describe co-association, we use variables $y \in \{0,1\}^{|\mathcal{D}|\times|\mathcal{D}|\times|\mathcal{J}|}$, where $y_{d_1d_2j} = 1$ if and only if nodes $d_1, d_2 \in \mathcal{D}$ are part of a common cluster $j \in \mathcal{J}$. The proposed ILP is

$$\min_{\substack{x\geq 0\\ y\geq 0}} \quad \sum_{\substack{(d_1,d_2)\in\mathcal{E}\\ j\in\mathcal{J}}} \theta_{d_1d_2} y_{d_1d_2j} \tag{1}$$

$$\text{s.t.:} \quad \sum_{j\in\mathcal{J}} x_{dj} = 1 \qquad \forall d \in \mathcal{D} \tag{2}$$

$$y_{d_1d_2j} \leq x_{d_1j} \qquad \forall (d_1,d_2) \in \mathcal{E}, j \in \mathcal{J} \tag{3}$$

$$y_{d_1d_2j} \leq x_{d_2j} \qquad \forall (d_1,d_2) \in \mathcal{E}, j \in \mathcal{J} \tag{4}$$

$$x_{d_1j} + x_{d_2j} - y_{d_1d_2j} \leq 1 \qquad \forall (d_1,d_2) \in \mathcal{E}, j \in \mathcal{J} \tag{5}$$

$$x_{dj} \in \{0,1\} \qquad \forall d \in \mathcal{D}, j \in \mathcal{J}. \tag{6}$$

The objective function (1) aims at minimizing the sum of the weights of the intra-cluster edges. Constraints (2) ensure that each node is assigned to exactly one cluster. Constraints (3)–(5) collectively enforce that $y_{d_1d_2j} = 1$ if and only if $x_{d_1j} = 1$ and $x_{d_2j} = 1$. Constraints (6) express the binary requirements on x, which also ensure that y is binary. Model (1)–(5) is referred to as the compact relaxation of correlation clustering.

Let see why it is highly inefficient to solve (1)–(6) using a standard branch-and-bound algorithm, where the LP relaxation provides a lower bound in any given branch. Observe that an optimal solution to the compact relaxation is given by: $x_{dj} = \frac{1}{|\mathcal{D}|}$ for all $d \in \mathcal{D}$, $j \in \mathcal{J}$, and $y_{d_1d_2j} = |\frac{1}{\mathcal{D}}|$ if $\theta_{d_1d_2} < 0$ and 0 otherwise, for all $d_1 \in \mathcal{D}$, $d_2 \in \mathcal{D}$, $j \in \mathcal{J}$. The ensuing lower bound is equal to the sum of all negative weights in θ; thus, (1)–(6) has a very loose LP relaxation. It is well established that using a loose relaxation in a branch-and-bound algorithm yields a very slow solution process.

3.3 Extended formulation of correlation clustering

Now, let us present an extended formulation of correlation clustering that has a much tighter LP relaxation than the compact formulation. Consider the power set of $\mathcal{D}$ denoted $\mathcal{G}$. A set $g \in \mathcal{G}$ is described using $G \in \{0,1\}^{|\mathcal{D}|\times|\mathcal{G}|}$, where $G_{dg} = 1$ if and only if node $d \in \mathcal{D}$ is in g. Correlation clustering corresponds to selecting a non-overlapping subset of $\mathcal{G}$. Each selected member of $\mathcal{G}$ is a cluster in our solution to correlation clustering and each node that is not part of a selected cluster is in a cluster by itself.

The cost of cluster $g \in \mathcal{G}$ is denoted Γ_g and is defined as the sum of the weights of all edges between the nodes it contains, i.e.,

$$\Gamma_g = \sum_{(d_1,d_2)\in\mathcal{E}} G_{d_1g} G_{d_2g} \theta_{d_1d_2}. \tag{7}$$

We represent a selection of sets in $\mathcal{G}$ using variables $\gamma \in \{0,1\}^{|\mathcal{G}|}$, i.e., we set $\gamma_g = 1$ if cluster $g \in \mathcal{G}$ is selected and $\gamma_g = 0$ otherwise. Below we frame correlation clustering as selecting the least cost non-overlapping subset of $\mathcal{G}$, or equivalently, as the MWSP problem

$$\min_{\gamma\geq 0} \quad \sum_{g\in\mathcal{G}} \Gamma_g \gamma_g \tag{8}$$

$$\text{s.t.:} \quad \sum_{g\in\mathcal{G}} G_{dg}\gamma_g \leq 1 \qquad \forall d \in \mathcal{D} \tag{9}$$

$$\gamma_g \in \{0,1\} \qquad \forall g \in \mathcal{G}. \tag{10}$$

Objective function (8) consists of minimizing the sum of the costs of the selected clusters. Constraints (9) impose that every node is assigned to at most one cluster. If the solution γ does not select a cluster that includes $d \in \mathcal{D}$, then d is in a cluster by itself. Constraints (10) enforce that γ is binary. Formulation (9)–(10) is referred to as the integer master problem (IMP). In addition, its LP relaxation is called the MP.

We provide here an illustrative example in which the MP provides a tighter relaxation than the compact relaxation. Let $\mathcal{D} = \{d_1, d_2, d_3\}$ and θ be defined as $\theta_{d_1 d_2} = -1, \theta_{d_1 d_3} = -2, \theta_{d_2 d_3} = 10$. The optimal integer solution groups d_1 and d_3 together while d_2 lies by itself. The MP optimal value is equal to -2, which is the cost of the optimal integer solution. However, the compact relaxation has the optimal solution described in Section 3.2, thus achieving a lower bound of -3.

3.4 Solving extended formulations via column generation

The most apparent difficulty in solving extended formulations is the potentially massive size of the set of variables, which is $\mathcal{G}$ in (8)–(10). For the computer vision problems considered, it is not feasible to enumerate, much less to consider in optimization all possible variables. Now, let us discuss CG in the context of correlation clustering, which is used identically for our applications.

CG circumvents the problem of considering the entire set $\mathcal{G}$ by constructing a subset of $\mathcal{G}$ denoted $\hat{\mathcal{G}}$. CG constructs $\hat{\mathcal{G}}$ so that solving the MP over $\hat{\mathcal{G}}$ provides the same optimal value as solving it over $\mathcal{G}$. Subset $\hat{\mathcal{G}}$ is constructed iteratively. At each iteration, the MP restricted to the current subset $\hat{\mathcal{G}}$, i.e.,

$$\min_{\gamma \geq 0} \quad \sum_{g \in \hat{\mathcal{G}}} \Gamma_g \gamma_g \tag{11}$$

$$\text{s.t.:} \quad \sum_{g \in \hat{\mathcal{G}}} G_{dg} \gamma_g \leq 1 \qquad \forall d \in \mathcal{D}, \tag{12}$$

is solved using a linear programming solver. This problem is called the restricted master problem (RMP). Let $\lambda_d \leq 0$, $d \in \mathcal{D}$, be the dual variables associated with constraints (12).

Solving the RMP yields a primal solution γ_g, $g \in \hat{\mathcal{G}}$, and a dual solution λ_d, $d \in \mathcal{D}$. This primal solution, augmented with $\gamma_g = 0$ for all $g \in \mathcal{G} \setminus \hat{\mathcal{G}}$, is also optimal for the MP if no non-generated variable γ_g, $g \in \mathcal{G} \setminus \hat{\mathcal{G}}$, has a negative reduced cost. This condition is verified by an oracle that must provide a negative reduced cost variable if one exists. The task of finding a least reduced cost variable γ_g is referred to as pricing and is modeled as

$$\min_{g \in \mathcal{G}} \quad \Gamma_g - \sum_{d \in \mathcal{D}} G_{dg} \lambda_d. \tag{13}$$

The pricing problem (13) is not solved by explicitly considering each set $g \in \mathcal{G}$, but rather as an integer program, or very commonly a dynamic program. The CG process stops when no more negative reduced cost variables exist.

A pseudo-code of the proposed CG algorithm is presented in Algorithm 1. The algorithm begins with an empty set $\hat{\mathcal{G}}$. It then iterates between solving the RMP (11)–(12) and adding to $\hat{\mathcal{G}}$ sets found by solving the pricing problem (13). When no negative reduced cost variables exist, CG terminates. For practical problems in our applications, the computed MP optimal solution is nearly always integral. When this is not the case, an approximate integer solution is produced by solving the MWSP problem over the set $\hat{\mathcal{G}}$ instead of $\mathcal{G}$ using an ILP solver (Step 10).

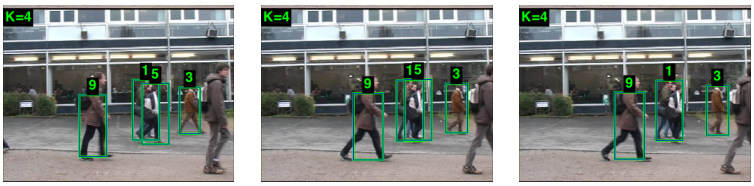
Solving the MWSP problem restricted to $\hat{\mathcal{G}}$ may not produce an optimal integer solution because there is no guarantee that $\hat{\mathcal{G}}$ contains all sets that are part of an optimal solution. Nevertheless, the tight bound yielded by the MP usually leads to high-quality approximate solutions. Note that the MP can still be tightened using valid inequalities such as the subset-row inequalities (Jepsen et al., 2008; Wang et al., 2017b) which are described in Section 6.

Algorithm 1 Basic column generation

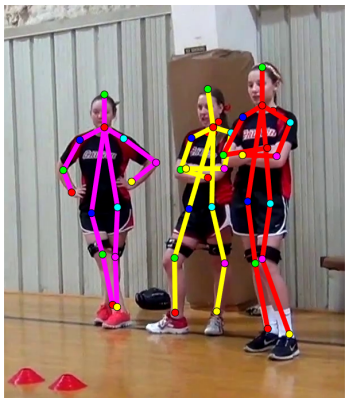
$\hat{\mathcal{G}} \leftarrow \emptyset$
repeat
  $\lambda, \gamma \leftarrow$ Solve the RMP (11)–(12)
  $g^* \leftarrow$ Solve the pricing problem (13)
  if $\Gamma_{g^*} - \sum_{d \in \mathcal{D}} G_{dg^*} \lambda_d < 0$ **then**
    $\hat{\mathcal{G}} \leftarrow \hat{\mathcal{G}} \cup \{g^*\}$
  end if
until $\Gamma_{g^*} - \sum_{d \in \mathcal{D}} G_{dg^*} \lambda_d \geq 0$
if γ is not integer **then**
  $\gamma \leftarrow$ Solve MWSP (8)–(10) over $\hat{\mathcal{G}}$ instead of $\mathcal{G}$
end if
Return γ

4 MWSP formulations of problems in computer vision

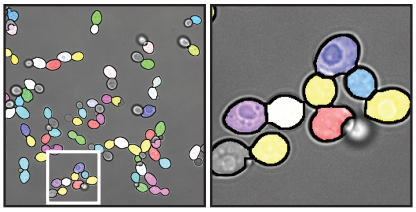
In this section we provide a high-level discussion of some MWSP formulations of problems in computer vision. In Sections 4.1, 4.2 and 4.3, we consider MWSP formulations of multi-person tracking, multi-person pose estimation, and multi-cell segmentation, respectively. These applications are visualized in Figure 1. In Section 4.4, we describe how MWSP complements supervised learning.



(a) Multi-person tracking



(b) Multi-person pose estimation



(c) Multi-cell segmentation

Figure 1: 1(a): Observations correspond to detections of people and hypotheses to tracks of people moving across time. Numbers denote the bounding boxes for a common person across frames. (Picture from Wang et al., 2017b). 1(b): Observations correspond to detections of body parts and hypotheses to people. Lines of a common color associate a person to the average position of each body part. There is a surjection of body parts (head, neck, etc.) to colors for dots that indicate the body part. 1(c): Observations correspond to super-pixels and hypotheses to complete biological cells. Cells are color-coded arbitrarily, with each cell being provided a single color. (Picture from Zhang et al., 2017).

4.1 Multi-person tracking

Multi-person tracking is the task of identifying and tracking each unique person in video. This task is motivated by security applications and autonomous vehicle applications. In multi-person tracking, the specific identities of the people in the image are unspecified, as well as the number of people present.

Combinatorial optimization has been applied to multi-person tracking in the form of minimum-cost network flow techniques (Zhang et al., 2008; Butt and Collins, 2013) and a MWSP-based approach (Wang et al., 2017b). We focus on multi-person tracking using MWSP, and describe the corresponding workflow below.

1. First, a classifier (such as an ANN) identifies all candidate detections of people in each frame of the video. Some of these detections are false detections.
2. Second, a classifier associates each group of K detections (for a single user-defined parameter K that trades off modeling power and computation requirements) ordered in time, each on a separate frame, with a real valued cost. This cost describes how plausible it is for those K detections to follow each other directly in the track of a single person. These groups are called subtracks. The set of subtracks is pruned by relying on the fact that most subsets of K detections are non-sensible since the detections are not sufficiently visually similar to correspond to a common person. Similarly, subtracks that do not follow the known statistics of human motion are removed; e.g., humans cannot teleport across space within a few frames of video.
3. Finally, the packing of detections into sequences of subtracks forming complete tracks is formulated as a MWSP problem, where the observations are the detections and the hypotheses are the complete tracks.

The MWSP formulation proposed by Wang et al. (2017b) relies on the classic approach of using a Markov model for scoring the quality of a track (Zhang et al., 2008; Butt and Collins, 2013). This model incorporates scores corresponding to the statistical support for the subtracks within a track. Let $\mathcal{D}$ be the set of detections of people in the video frames and $\mathcal{S}$ the set of subtracks. For a given subtrack $s \in \mathcal{S}$, let s_k indicate the k^{th} detection in the sequence $s = \{s_1, ..., s_K\}$ ordered by time from earliest to latest. Note that the detections that compose a subtrack need not be consecutive in time, thus permitting a person to disappear and reappear in video. The set of potential tracks is denoted $\mathcal{G}$, where a track is a sequence of subtracks ordered in time. In any track, the latest $K-1$ detections in time of any subtrack s^1 in the sequence are the earliest $K-1$ detections of the subtrack s^2 that immediately succeeds s^1. Observe that a track can be equivalently described as a sequence of detections ordered in time or a sequence of subtracks ordered in time. The mapping of subtracks to tracks is described using $T \in \{0,1\}^{|\mathcal{S}|\times|\mathcal{G}|}$, where $T_{sg} = 1$ indicates that track g contains subtrack s as a subsequence. Subtracks are illustrated in Figure 2.

Track costs Γ are written in terms of the subtrack costs $\theta \in \mathbb{R}^{|\mathcal{S}|}$, where each subtrack s is associated with a cost θ_s. Here, positive/negative values of θ_s discourage/encourage the use of the subtrack s. We model a Bayesian prior belief on the number of people (tracks) in an image using θ^0, which is the cost for instancing a track. Positive/negative values of θ^0 discourage/encourage the presence of more tracks in the packing. Using θ, the cost of a track $g \in \mathcal{G}$, denoted Γ_g, is defined as

$$\Gamma_g = \theta^0 + \sum_{s\in\mathcal{S}} T_{sg}\theta_s. \tag{14}$$

To permit the construction of tracks that have fewer detections than K, the set of subtracks is augmented with subtracks padded with empty detections. Such subtracks have no possible predecessors or successors.

4.2 Multi-person pose estimation

Multi-person pose estimation is the task of identifying each unique person in an image, and annotating their body parts. As in tracking, one does not know the specific identities of the people in the image, and the number of people present is unspecified. Multi-person pose estimation is relevant in multiple domains including but not limited to personal robot assistant, rehabilitation, and security. We now consider the workflow of the MWSP formulation of multi-person pose estimation (Wang et al., 2018), which built off the seminal work of Pishchulin et al. (2016); Insafutdinov et al. (2016).

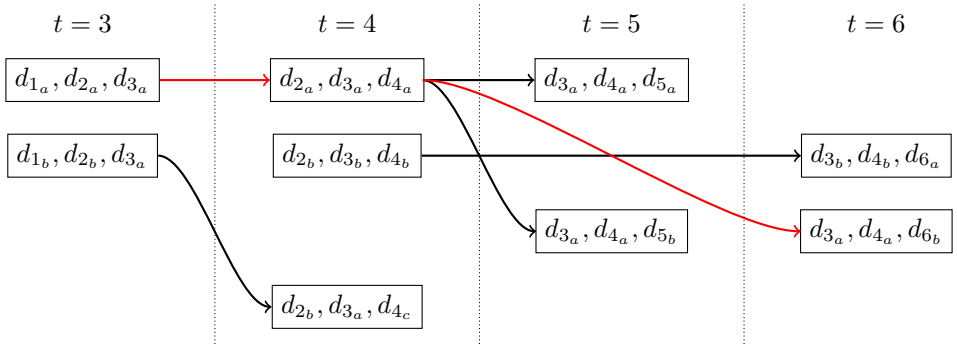


Figure 2: Possible tracks and subtracks (boxes), where directed arrows indicate the valid successors of a given subtrack. The subtracks are ordered by the time of their final detection. Note that a subtrack may skip some time steps, e.g., $[d_{3a}, d_{4a}, d_{6b}]$ skips time 5. Red lines highlight a single track containing detections $d_{1_a}, d_{2_a}, d_{3_a}, d_{4_a}, d_{6_b}$.

1. First, an ANN identifies all instances of each of fourteen human body parts (head, neck, and left/right of the following: shoulder, elbow, wrist, hip, knee, ankle). Some of these detections are false detections. Some sets of detections correspond to the same body part, but are separated in pixel space.
2. Second, a classifier (such as an ANN) computes for each pair of detections the cost incurred if they are associated with the same person. This cost is derived from the probability that the two detections belong to a common person. Similarly, a cost associating each detection with a person is computed. These classifiers take as input local statistics of the pixel values around the detections, and or spatial, angular statistics concerning the relative location of the pairs of detections. We refer to the cost terms over pairs of detections as pairwise, and those over a single detection as unary (for details on the cost term generation, see Insafutdinov et al. (2016) and Pishchulin et al. (2016)). Person detection in computer vision relies traditionally on tree (pictorial) structured models (Felzenszwalb et al., 2008), which describe the feasibility of poses of the human body. Feasibility is modeled according to a cost function defined on a graph (typically, a tree), where nodes correspond to body parts and edges indicate adjacency. Thus, pairwise cost terms may be non-zero only between detections corresponding to the same body part, or adjacent body parts in the tree model. Such a model, augmented with additional connections, is drawn in Figure 3. The use of connections outside the tree structure increases modeling power and computational performance (as reported in Section 8.2) in the next step of this pipeline.
3. Finally, the body part detections (observations) are aggregated to form people (hypotheses) using a MWSP formulation.

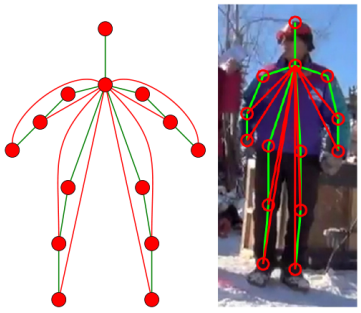


Figure 3: (Left) Augmented-tree model of a person as a stick figure. Each red node represents a body part, green edges indicate connections in traditional pictorial structure, and red edges indicate augmented connections. (Right) Augmented-tree model superimposed on an image of a person. (Picture from Wang et al., 2018).

The MWSP formulation of multi-person pose estimation proposed by Wang et al. (2018) denotes by $\mathcal{D}$ the set of body part detections. For each detection $d \in \mathcal{D}$, parameter R_d indicates the body part associated with d. The set of potential people $\mathcal{G}$ is defined as the power set of $\mathcal{D}$. Observe that a person can contain more than one detection of any given body part. This is a modeling choice and a consequence of the body part detector firing at multiple places in close proximity, corresponding to the same ground truth body part. Similarly, since human body parts can be occluded in real images, it is possible for a hypothesis to contain zero detections of some body parts.

The cost of a person is defined using unary and pairwise terms $\theta^1 \in \mathbb{R}^{|\mathcal{D}|}$ and $\theta^2 \in \mathbb{R}^{|\mathcal{D}|\times|\mathcal{D}|}$. For each detection $d \in \mathcal{D}$, θ^1_d denotes the cost of including d in a person. Similarly, for each pair of detections $d_1, d_2 \in \mathcal{D}$, $\theta^2_{d_1 d_2}$ is the cost of including d_1 and d_2 in a common person. Here positive/negative values of θ^1_d discourage/encourage the use of detection d in a person. Similarly, positive/negative values of $\theta^2_{d_1 d_2}$ discourage/encourage the presence of d_1 and d_2 jointly in a single person. Note that θ^2 respects the augmented tree structure used to model a person. Thus, $\theta^2_{d_1 d_2}$ can only be non-zero if $R_{d_1} = R_{d_2}$ or if R_{d_2} is adjacent to R_{d_1} in this tree. Furthermore, for the sake of notational conciseness, we assume that both $\theta^2_{d_1 d_2}$ and $\theta^2_{d_2 d_1}$ are defined and equal for each pair of distinct detections $d_1, d_2 \in \mathcal{D}$, and that $\theta^2_{dd} = 0$ for each detection $d \in \mathcal{D}$. Finally, as for multi-person tracking, a constant cost θ^0 is associated with instancing a person for which a positive/negative value discourages/encourages the presence of more people in the packing. Given these cost terms, the cost of a person $g \in \mathcal{G}$ is defined as

$$\Gamma_g = \theta^0 + \sum_{d \in \mathcal{D}} \theta^1_d G_{dg} + \sum_{\substack{d_1 \in \mathcal{D} \\ d_2 \in \mathcal{D}}} \theta^2_{d_1 d_2} G_{d_1 g} G_{d_2 g}. \tag{15}$$

4.3 Multi-cell segmentation

Multi-cell segmentation is the task of identifying each unique biological cell in an image, and identifying the pixels associated with each cell. The number of cells present in a image is unspecified. Multi-cell segmentation is useful in domains such as image microscopy, where characterizing the movements and activities of cells is important, but the capacity of human annotators is limited. Multi-cell segmentation can be cast as a correlation clustering problem (Zhang et al., 2014a) and formulated as a MWSP problem (Zhang et al., 2017). The pipeline used by Zhang et al. (2017) is described below.

1. First, given a biological image, we apply dimensionality reduction by partitioning set of pixels into subsets called super-pixels (Ren and Malik, 2003). This is done by aggregating pixels for which a classifier is extremely confident that they correspond to the same cell or they are in the background. This classifier uses local spatial and color statistics. This conversion reduces the space of millions of pixels to thousands of super-pixels, but rarely meaningfully compromises the boundaries of any cell in the ground truth.
2. Second, for each pair of adjacent super-pixels, a classifier is used to provide a cost for the pair to be associated with a common cell. Similarly, a classifier generates a cost for each super-pixel to be part of a cell. As above, we refer to these costs as pairwise and unary, respectively.
3. Third, we compute the maximum radius and area (volume in 3D images) of cells on the annotated data.
4. Finally, identifying each cell in the image is formulated as a MWSP problem, where observations are super-pixels and hypotheses are cells.

In their MWSP formulation of multi-cell segmentation, Zhang et al. (2017) denotes by $\mathcal{D}$ the set of super-pixels and by $\mathcal{G}$ the set of potential biological cells. The quality of a cell is defined in terms of obeying the known structural properties of a cell, which in this case describe the radius, area (volume in 3D for supervoxels), and agreement with the local image statistics.

The constraint on the radius of a cell is defined as follows. For any cell $g \in \mathcal{G}$, there exists a super-pixel d^*, which we refer to as an anchor, such that all super-pixels in cell g are within a user-defined

distance $R_{\max}$ of d^*. Let $S_{d_1 d_2}$ be the distance between the centers of super-pixels d_1 and d_2. We use [...] to denote the binary indicator function, which takes value one if the statement inside is true and zero otherwise. The radius constraint is satisfied for a given cell $g \in \mathcal{G}$ if

$$\exists d^* \in \mathcal{D} \quad \text{such that} \quad [G_{dg} = 1] \Rightarrow [S_{d^* d} \leq R_{\max}] \quad \forall d \in \mathcal{D}. \tag{16}$$

Optionally, the anchor can be required to be present in the cell, which we do in our experiments.

To define the constraint on the area of a cell, denote by $A_{\max}$ the upper bound on the area of a cell and by A_d the area of a super-pixel $d \in \mathcal{D}$. A cell $g \in \mathcal{G}$ satisfies the area constraint if

$$\sum_{d \in \mathcal{D}} A_d G_{dg} \leq A_{\max}. \tag{17}$$

The cost of a cell (reflecting the within-image evidence of its quality) is again defined using unary and pairwise term costs $\theta^1 \in \mathbb{R}^{|\mathcal{D}|}$ and $\theta^2 \in \mathbb{R}^{|\mathcal{D}| \times |\mathcal{D}|}$. For each super-pixel $d \in \mathcal{D}$, θ^1_d denotes the cost for super-pixel d to be part of any cell. Similarly, for each super-pixel pair $d_1, d_2 \in \mathcal{D}$, $\theta^2_{d_1 d_2}$ is the cost for d_1 and d_2 to belong to a common cell (as in the previous section, both $\theta^2_{d_1 d_2}$ and $\theta^2_{d_2 d_1}$ are defined and equal for each pair $d_1, d_2 \in \mathcal{D}$, and $\theta^2_{dd} = 0$ for each $d \in \mathcal{D}$). Here, positive/negative values of θ^1_d discourage/encourage the use of the super-pixel d in a cell, whereas positive/negative values of $\theta^2_{d_1 d_2}$ discourage/encourage the presence of d_1 and d_2 jointly in a common cell. Furthermore, as in the previous applications, a constant offset θ^0 is added to the cost of a cell to penalize/reward having additional cells in the image. Given these terms, the cost Γ_g of a potential cell g that satisfies the maximum radius and area constraints is expressed as

$$\Gamma_g = \theta^0 + \sum_{d \in \mathcal{D}} \theta^1_d G_{dg} + \sum_{\substack{d_1 \in \mathcal{D} \\ d_2 \in \mathcal{D}}} \theta^2_{d_1 d_2} G_{d_1 g} G_{d_2 g}. \tag{18}$$

Otherwise, we assume that $\Gamma_g = \infty$.

4.4 Complementarity of MWSP and classifiers

MWSP is a natural complement to supervised learning methods such as ANNs. This is a consequence of the following two properties of classifiers trained to indicate if two observations are associated with a common hypothesis. (1) Classification may produce predictions that are inconsistent with regards to transitivity. (2) Classification may produce predictions that are inconsistent with regards to domain knowledge as illustrated in the following example.

Consider ten observations and that all hypotheses consist of exactly nine observations. Assume that each observation and each pair of observations is equally likely to be part of the single hypothesis which is known to exist. Thus, each observation is in the hypothesis with probability $\frac{9}{10}$, and each pair with probability $\frac{4}{5}$. If we rely solely on these probabilities alone ignoring the fact that a hypothesis contains exactly nine observations then the maximum likelihood solution predicts a single hypothesis consisting of all ten observations. Below we list some real-world cases, where domain knowledge provides structural requirements to hypotheses.

- **Multi-person pose estimation**: A person can contain no more than two legs and two arms;
- **Neuron tracing**: Each neuron is connected across space;
- **Cell detection**: Certain types of biological cells are convex.

5 Pricing problems for our applications

For the MWSP formulations of the applications considered in the previous section, we define the CG pricing problem and, when necessary, briefly discuss how to solve it.

5.1 Pricing for multi-person tracking

The task of identifying the least reduced cost track can be formulated as the following dynamic program. A subtrack $s \in \mathcal{S}$ may be preceded by another subtrack $\hat{s}$ if and only if the first $K-1$ detections in s correspond to the last $K-1$ detections in $\hat{s}$. We denote by Π_s the set of valid subtracks that may precede a subtrack s. For each subtrack $s \in \mathcal{S}$, let ℓ_s be the reduced cost of the least reduced cost track that terminates at subtrack s. Ordering subtracks by the time of its last detection allows efficient computation of the reduced costs ℓ_s, $s \in \mathcal{S}$, using the dynamic program

$$\ell_s = \theta_s - \lambda_{s_K} + \min\{\min_{\hat{s}\in\Pi_s} \ell_{\hat{s}}, \theta^0 - \sum_{k=1}^{K-1} \lambda_{s_k}\}. \tag{19}$$

We may choose to add not only the least reduced cost track to $\hat{\mathcal{G}}$, but other distinct negative reduced cost tracks. This strategy can easily be implemented since the dynamic program produces a least reduced cost track terminating at each subtrack. The strategy employed in Wang et al. (2017b) adds to $\hat{\mathcal{G}}$ the least reduced cost track terminating at each detection (excluding those with non-negative reduced cost).

5.2 Pricing for multi-person pose estimation

Let us formulate the task of identifying the least reduced cost person as a set of dynamic programs. Consider the subgraph of the graph in Figure 3 in which the neck is removed. It has a tree structure, more precisely, it is composed of multiple disconnected trees where the nodes correspond to human body parts and the edges indicate adjacency. Connecting the disconnected component trees with zero valued pairwise terms produces a single tree.

During the pricing step, we iterate through the power set of neck detections, and compute the least reduced cost person containing exactly those neck detections. We index the power set of neck detections with $\breve{\mathcal{D}}$ and use $[g \leftrightarrow \breve{\mathcal{D}}] = 1$ to indicate that the neck detections in a hypothesis $g \in \mathcal{G}$ are exactly those in $\breve{\mathcal{D}}$. The pricing problem for an arbitrary neck detection subset $\breve{\mathcal{D}}$ is

$$\min_{\substack{g\in\mathcal{G} \\ [g\leftrightarrow\breve{\mathcal{D}}]=1}} \Gamma_g - \sum_{d\in\mathcal{D}} G_{dg}\lambda_d. \tag{20}$$

To solve model (20) as a dynamic program, we assume that we can enumerate the power set of detections corresponding to pairs of adjacent parts in the augmented tree.

Let $\mathcal{R}$ be the set of human body parts. For each part $r \in \mathcal{R}$, denote by $\mathcal{D}^r$ the set of its detections and by $\mathcal{S}^r$ the power set of these detections. We describe $\mathcal{S}^r$ using $S^r \in \{0,1\}^{|\mathcal{D}^r|\times|\mathcal{S}^r|}$, where $S^r_{ds} = 1$ indicates that detection $d \in \mathcal{D}^r$ is in set $s \in \mathcal{S}^r$. For convenience, we explicitly define the neck as part 0 and thus the power set of neck detections is denoted $\mathcal{S}^0$.

Note that, when conditioned on a specific set s_0 of neck detections, the pairwise costs from these neck detections to all other detections can be added to unary costs of the other detections. Thus, the augmented-tree structure becomes a typical tree structure, and exact inference can be done via dynamic programming. We make this tree directed by choosing an arbitrary node to be the root, and orienting the edges in the graph away from the root.

Let Υ^r be the set of children of a body part $r \in \mathcal{R}$ in the tree graph. Below we define $\mu^r_{\hat{s}}$ as the reduced cost of the least reduced cost subtree rooted at r, given that its parent $\hat{r}$ takes on state $\hat{s} \in \mathcal{S}^{\hat{r}}$. This reduced cost $\mu^r_{\hat{s}}$ includes the cost of the pairwise terms between the detections of parts $\hat{r}$ and r, and is written as

$$\mu^r_{\hat{s}} = \min_{s\in\mathcal{S}^r} \sum_{\substack{\hat{d}\in\mathcal{D}^{\hat{r}} \\ d\in\mathcal{D}^r}} S^{\hat{r}}_{\hat{d}\hat{s}} S^r_{ds} \theta^2_{\hat{d}d} + \nu^r_s, \tag{21}$$

where ν_s^r is the cost of the subtree rooted at part r with state $s \in \mathcal{S}^r$:

$$\nu_s^r = \sum_{d \in \mathcal{D}^r} (\theta_d^1 - \lambda_d) S_{ds}^r + \sum_{\substack{d_1 \in \mathcal{D}^r \\ d_2 \in \mathcal{D}^r}} \theta_{d_1 d_2}^2 S_{d_1 s}^r S_{d_2 s}^r + 2 \sum_{\substack{d_1 \in \mathcal{D}^0 \\ d_2 \in \mathcal{D}^r}} \theta_{d_1 d_2}^2 S_{d_1 s_0}^0 S_{d_2 s}^r + \sum_{\bar{r} \in \Upsilon^r} \mu_s^{\bar{r}}. \tag{22}$$

Observe that (21)–(22) describe a dynamic program.

To compute $\mu_{\hat{s}}^r$ for each $\hat{s} \in \mathcal{S}^{\hat{r}}$, we need to iterate over all $s \in \mathcal{S}^r$. For most, though not all problem instances in Wang et al. (2018), this is feasible. However, observe that, if $|D^r| = |D^{\hat{r}}| = 15$, then the joint space of more than one billion configurations would have to be enumerated, which is prohibitively expensive. This has motivated the use of NBD in Wang et al. (2018), which is able to solve the dynamic program exactly with a practical computational complexity of $O(|\mathcal{D}^r|)$, not $O(|\mathcal{D}^r| \times |\mathcal{D}^{\hat{r}}|)$.

Given that pricing is computationally expensive with respect to solving the RMP in each CG iteration, all negative reduced cost hypotheses found when solving the dynamic programs are added to $\hat{\mathcal{G}}$ in each iteration.

5.3 Pricing for multi-cell segmentation

To find negative reduced cost cells, we exploit the fact that cells are small and compact. Recall that every cell with a finite cost is associated with an anchor d^* in close proximity to all other super-pixels that compose the cell. Let $\mathcal{D}_{d^*} \subseteq \mathcal{D}$ be the subset of detections that may be in a cell with anchor $d^* \in \mathcal{D}$, i.e.,

$$\mathcal{D}_{d^*} = \{d \in \mathcal{D} \mid S_{d^* d} \leq R_{\max}\}. \tag{23}$$

We attack pricing by conditioning on the choice of the anchor d^* and by finding for each possible anchor $d^* \in \mathcal{D}$ a least reduced cost cell g_{d^*}, that is,

$$g_{d^*} \in \arg \min_{\substack{g \in \mathcal{G} \\ G_{dg} = 0, \forall d \notin \mathcal{D}_{d^*}}} \theta^0 + \sum_{d \in \mathcal{D}} (\theta_d^1 - \lambda_d) G_{dg} + \sum_{d_1, d_2 \in \mathcal{D}} \theta_{d_1 d_2}^2 G_{d_1 g} G_{d_2 g}. \tag{24}$$

This problem can be formulated as an ILP using the decision variables $x \in \{0,1\}^{|\mathcal{D}|}$ and $y \in \{0,1\}^{|\mathcal{D}| \times |\mathcal{D}|}$, where $x_d = 1$ if detection $d \in \mathcal{D}$ is selected to be part of the cell g_{d^*} and $y_{d_1,d_2} = 1$ if both detections $d_1 \in \mathcal{D}$ and $d_2 \in \mathcal{D}$ are part of it. The ILP is:

$$\min_{\substack{x \in \{0,1\}^{|\mathcal{D}|} \\ y \geq 0}} \theta^0 + \sum_{d \in \mathcal{D}} (\theta_d^1 - \lambda_d) x_d + \sum_{d_1, d_2 \in \mathcal{D}} \theta_{d_1 d_2}^2 y_{d_1 d_2} \tag{25}$$

$$\text{s.t.:} \quad y_{d_1 d_2} \leq x_{d_1} \qquad \forall d_1, d_2 \in \mathcal{D} \tag{26}$$

$$y_{d_1 d_2} \leq x_{d_2} \qquad \forall d_1, d_2 \in \mathcal{D} \tag{27}$$

$$-y_{d_1 d_2} + x_{d_1} + x_{d_2} \leq 1 \qquad \forall d_1, d_2 \in \mathcal{D} \tag{28}$$

$$\sum_{d \in \mathcal{D}} x_d \geq 1 \tag{29}$$

$$\sum_{d \in \mathcal{D}} A_d x_d \leq A_{\max} \tag{30}$$

$$x_d = 0 \qquad \forall d \in \mathcal{D} \setminus \mathcal{D}_{d^*}. \tag{31}$$

The binary requirements on y are enforced through (26)–(28). Indeed, constraints (26)–(27) state that $y_{d_1 d_2}$ cannot be set to one unless both detections d_1 and d_2 are included in the cell g_{d^*}, whereas (28) ensures that $y_{d_1 d_2}$ is set to one if both d_1 and d_2 are included in g_{d^*}. Consequently, there is no need to explicitly require y to be binary. Constraint (29) imposes that at least one super-pixel be included in the cell. Constraints (30) enforces that the area of the cell does not exceed the maximum area.

Constraints (31) ensure that all detections not respecting the maximum radius constraint from d^* are not selected in the cell. Recall from Section 4.3 that the anchor may be required to be included in the cell. This condition is imposed by setting x_{d^*} to one.

Because ILP (24) is solved for different anchors d^*, many distinct hypotheses with a negative reduced cost are often generated at each CG iteration. In this case, all these hypotheses are added to the nascent set $\hat{\mathcal{G}}$.

6 Subset-row inequalities

In this section, we tighten the LP relaxation of the MWSP formulation (8)–(10). To motivate this, we provide a case from Wang et al. (2017b) where the LP relaxation is loose. Consider four hypothesis $\mathcal{G} = \{g_1, g_2, g_3, g_4\}$ over three observations $\mathcal{D} = \{d_1, d_2, d_3\}$, where the first three hypotheses each contain two of the three observations $\{d_1, d_2\}$, $\{d_1, d_3\}$, $\{d_2, d_3\}$, respectively, and the fourth hypothesis contains all three $\{d_1, d_2, d_3\}$. Suppose that the costs of the hypotheses are given by $\Gamma_{g_1} = \Gamma_{g_2} = \Gamma_{g_3} = -4$ and $\Gamma_{g_4} = -5$. The optimal integer solution sets $\gamma_{g_4} = 1$, and has a cost of -5. However, the optimal LP solution sets $\gamma_{g_1} = \gamma_{g_2} = \gamma_{g_3} = 0.5$ and $\gamma_{g_4} = 0$, and has a cost of -6. Hence, the LP relaxation of (8)–(10) is loose in this case.

The LP relaxation of MWSP (8)–(10), or equivalently, the MP in the CG algorithm can be tightened by employing the subset-row inequalities introduced by (Jepsen et al., 2008). They can be parameterized by two integers m_1 and m_2, each greater than or equal to two, and a subset $\hat{\mathcal{D}} \subseteq \mathcal{D}$ of cardinality $m_1 m_2 - 1$. A subset-row inequality requires that the number of selected hypotheses containing m_1 or more members of $\hat{\mathcal{D}}$ must not exceed $m_2 - 1$. It writes (in a more generalized form) as

$$\sum_{g \in \mathcal{G}} \gamma_g \left\lfloor \frac{\sum_{d \in \hat{\mathcal{D}}} G_{dg}}{m_1} \right\rfloor \leq m_2 - 1. \tag{32}$$

These inequalities are added to the MP but their dual variables need to be handled by the pricing problem to compute the exact reduced costs of the hypotheses. Handling these dual values often destroys the structure of the pricing problem. For the multi-person tracking problem, Wang et al. (2017b) propose an elegant way to deal with them which relies on the original structure of the pricing problem.

In this section, we focus on the subset-row inequalities with $m_1 = m_2 = 2$, which are referred to as 3-SRIs because $|\hat{\mathcal{D}}| = m_1 m_2 - 1 = 3$. For any given subset $\hat{\mathcal{D}}$ of three observations, the corresponding 3-SRI enforces that the number of selected hypotheses, that include two or more of those observations, can be no larger than one. Note, however, that all content of this section can be generalized to other subset-row inequalities.

In Section 6.1, we present a MWSP formulation tightened using 3-SRIs and discuss how the CG algorithm is modified to account for them. In Section 6.2, we discuss the application of the subset-row inequalities to the multi-cell segmentation problem which preserves the structure of the pricing problem. Finally, in Section 6.3, we present the procedure of Wang et al. (2017b) to solve the pricing problem when the trivial use of subset-row inequalities destroys the structure of the pricing problem as in the multi-person tracking and multi-person pose estimation applications.

6.1 Tightened MWSP formulation

Let $\mathcal{C}$ be the set of the subsets of three distinct observations in $\mathcal{D}$. For a subset $c \in \mathcal{C}$, denote by $\mathcal{D}_c$ the set of observations in c. We describe the mapping of 3-SRIs to hypotheses using matrix $C \in \{0,1\}^{|\mathcal{C}| \times |\mathcal{G}|}$, where $C_{cg} = \left\lfloor \frac{\sum_{d \in \mathcal{D}_c} G_{dg}}{2} \right\rfloor$ for each pair of set $c \in \mathcal{C}$ and hypothesis $g \in \mathcal{G}$, i.e., $C_{cg} = 1$ if and only if hypothesis g contains at least two observations in c. With this notation, the

MWSP MP tightened using 3-SRIs writes as

$$\min_{\gamma \geq 0} \quad \sum_{g \in \mathcal{G}} \Gamma_g \gamma_g \tag{33}$$

$$\text{s.t.:} \quad \sum_{g \in \mathcal{G}} G_{dg} \gamma_g \leq 1 \qquad \forall d \in \mathcal{D} \tag{34}$$

$$\sum_{g \in \mathcal{G}} C_{cg} \gamma_g \leq 1 \qquad \forall c \in \mathcal{C}. \tag{35}$$

Given the relatively large size of $\mathcal{C}$, the 3-SRIs are generated only as needed, i.e., the MP is solved using a column/row generation (CRG) algorithm. It starts with an empty subset $\hat{\mathcal{C}}$ of 3-SRIs and solve MP (33)–(34) by CG, generating a subset $\hat{\mathcal{G}}$ of columns. If the computed MP solution is not integer, the CRG algorithm iterates over all sets $c \in \mathcal{C}$ to identify the set c^* that maximizes $\sum_{g \in \mathcal{G}} \gamma_g C_{cg}$, given the current fractional solution γ. If $\sum_{g \in \mathcal{G}} \gamma_g C_{c^* g} > 1$, then the corresponding 3-SRI is violated and c^* is added to set $\hat{\mathcal{C}}$. CG is then re-started to solve the MP (33)–(35), where the constraint set (35) is restricted to the subset $\hat{\mathcal{C}}$ of the generated 3-SRIs. Once solved, we search for a violated 3-SRI again and, if one is found, it is added to $\hat{\mathcal{C}}$. This process alternating between solving the MP and searching for a violated 3-SRI is repeated until no violated 3-SRI is found. Note that more than one 3-SRI can be added at once. Furthermore, note that the search for a violated 3-SRI can be limited to sets $c \in \mathcal{C}$ for which each detection $d \in \mathcal{D}_c$ is included in a hypothesis g associated with a fractional-valued variable γ_g (Wang et al., 2017b).

Let $\psi_c \leq 0$, $c \in \hat{\mathcal{C}}$, be the dual variables associated with the generated 3-SRIs. To take these dual variables into account, the CG pricing problem is redefined as

$$\min_{g \in \mathcal{G}} \quad \Gamma_g - \sum_{d \in \mathcal{D}} G_{dg} \lambda_d - \sum_{c \in \hat{\mathcal{C}}} C_{cg} \psi_c. \tag{36}$$

Given that $C_{cg} = \left\lfloor \frac{\sum_{d \in \mathcal{D}_c} G_{dg}}{2} \right\rfloor$ for each pair of set $c \in \hat{\mathcal{C}}$ and $g \in \mathcal{G}$, solving this pricing problem is often more complex than solving the pricing problem without the ψ_c dual values. Below, we discuss how this can be done for the three applications considered.

6.2 Pricing without modifying the structure of the pricing problem

In some cases, the pricing problem can preserve the same structure and computational complexity while dealing with the 3-SRI dual variables. One such example is multi-cell segmentation, where the pricing problem remains an ILP. Let $z_c \in \{0,1\}$, $c \in \hat{\mathcal{C}}$, be a binary variable that is equal to one if and only if two or more detections in $\mathcal{D}_c$ are included in the cell. The ILP is given by

$$\min_{\substack{x \geq 0 \\ y \geq 0 \\ z \geq 0}} \quad \theta^0 + \sum_{d \in \mathcal{D}} (\theta^1_d - \lambda_d) x_d + \sum_{d_1, d_2 \in \mathcal{D}} \theta^2_{d_1 d_2} y_{d_1 d_2} - \sum_{c \in \hat{\mathcal{C}}} \psi_c z_z \tag{37}$$

$$\text{s.t.:} \quad y_{d_1 d_2} \leq x_{d_1} \qquad \forall d_1, d_2 \in \mathcal{D} \tag{38}$$

$$y_{d_1 d_2} \leq x_{d_2} \qquad \forall d_1, d_2 \in \mathcal{D} \tag{39}$$

$$x_{d_1} + x_{d_2} - y_{d_1 d_2} \leq 1 \qquad \forall d_1, d_2 \in \mathcal{D} \tag{40}$$

$$x_d \in \{0,1\} \qquad \forall d \in \mathcal{D} \tag{41}$$

$$\sum_{d \in \mathcal{D}} x_d \geq 1 \tag{42}$$

$$\sum_{d \in \mathcal{D}} A_d x_d \leq A_{\max} \tag{43}$$

$$x_d = 0 \qquad \forall d \in \mathcal{D} \setminus \mathcal{D}_{d^*} \tag{44}$$

$$x_{d_3} + x_{d_4} - z_c \leq 1 \qquad \forall c \in \hat{\mathcal{C}}, d_3, d_4 \in \mathcal{D}_c \mid d_3 \neq d_4 \tag{45}$$

This pricing model is identical to model (25)–(31) except that the last term of the objective function is added to take into account the dual variables ψ in the cell reduced cost, and the constraints (45) are required to set the values of the z variables. Observe that, for every $c \in \hat{\mathcal{C}}$, z_c is set to its smallest possible value at optimality because ψ_c is non-positive. Thus, z_c is not explicitly required to be integer, since its integrality is assured when x is integral. Finally, note that all constraints (45) for which $\psi_c = 0$ can be ignored.

6.3 Pricing while modifying the structure of the pricing problem

Let us consider the problem of finding negative reduced cost variables when the dual variables ψ cannot be handled directly by the specialized solver used for pricing without them. This case often emerges in problems such as multi-person tracking and multi-person pose estimation that rely on dynamic programming as a specialized pricing solver. Given that $\psi \leq 0$, solving a so-called ψ-independent pricing problem where these dual variables are ignored provides a lower bound on the optimal value of the pricing problem. Based on this observation, Wang et al. (2017b) introduce a branch-and-bound algorithm for solving the pricing problem, where the lower bound at each node of the search tree is computed using the specialized solver. Let us describe this algorithm, where each branching decision imposes that an observation $d \in \mathcal{D}$ is included or not in the hypothesis.

Let $\mathcal{B}$ be the set of nodes in the search tree. For each node $b \in \mathcal{B}$, denote by $\mathcal{D}_b^+$ and $\mathcal{D}_b^-$ the subsets of observations that must be included in the hypothesis according to the branching decisions and that must be excluded from it, respectively. The set of all hypotheses that are consistent with both sets $\mathcal{D}_b^+$ and $\mathcal{D}_b^-$ is denoted as $\mathcal{G}_b^{\pm}$. At the root node b_0 of the search tree, we have $\mathcal{D}_{b_0}^+ = \mathcal{D}_{b_0}^- = \emptyset$ and $\mathcal{G}_{b_0}^{\pm} = \mathcal{G}$.

In Sections 6.3.1 and 6.3.2, we specify the bounding and branching operations, respectively.

6.3.1 Bounding

For every $b \in \mathcal{B}$, let $V_b^* = \min_{g \in \mathcal{G}_b^{\pm}} \left(\Gamma_g - \sum_{d \in \mathcal{D}} \lambda_d G_{dg} - \sum_{c \in \hat{\mathcal{C}}} \psi_c C_{cg}\right)$ be the optimal value of the pricing problem over the hypotheses in $\mathcal{G}_b^{\pm}$. Furthermore, denote by $\underline{V}_b^*$ the optimal value of the corresponding ψ-independent pricing problem (i.e., restricted to $\mathcal{G}_b^{\pm}$). Obviously, $\underline{V}_b^* \leq V_b^*$. Wang et al. (2017b) introduce the following stronger lower bound V_b^{lb} on V_b^*, which exploits the knowledge stored in the subset $\mathcal{D}_b^+$.

Proposition 1 *(from Wang et al., 2017b) Let $b \in \mathcal{B}$. Then,*

$$V_b^{lb} = \underline{V}_b^* - \sum_{c \in \hat{\mathcal{C}}} \psi_c \left\lfloor \frac{|\mathcal{D}_c \cap \mathcal{D}_b^+|}{2} \right\rfloor \tag{46}$$

is a lower bound on V_b^*, i.e., $V_b^{lb} \leq V_b^*$.

For multi-person tracking and multi-person pose estimation, the value of $\underline{V}_b^*$ can be computed by solving the dynamic program presented in Sections 5.1 and 5.2, respectively. This program is, however, modified as described below to enforce that $g \in \mathcal{G}_b^{\pm}$.

Multi-person tracking: To discard the detections in $\mathcal{D}_b^-$, we remove from $\mathcal{S}$ all subtracks that include a detection in D_b^-. To impose that all detections in $\mathcal{D}_b^+$ be part of the track, we first remove all substracks that includes a detection $d' \notin \mathcal{D}_b^+$ or an occlusion co-occurring in time with any $d \in \mathcal{D}_b^+$. Furthermore, we do not consider starting a track after the occurrence of the first member of $\mathcal{D}_b^+$ in time. Finally, after completing the dynamic program algorithm, we remove any generated track that terminates prior to the point in time of the last member of $\mathcal{D}_b^+$.

Multi-person pose estimation: In contrast to multi-person tracking, enforcing that $g \in \mathcal{G}_b^{\pm}$ in multi-person pose estimation is simple. We force detections in $\mathcal{D}_b^+$ and $\mathcal{D}_b^-$ to be active/inactive, respectively, when generating a person. Specifically, for any given body part $r \in \mathcal{R}$, we only consider detection subsets $s \in \mathcal{S}^r$ that are consistent with node b when solving the dynamic program. Thus, a subset s such that $S^r_{ds} = 1$ for any $d \in \mathcal{D}_b^-$ or $S^r_{ds} = 0$ for any $d \in \mathcal{D}_b^+$ is ignored. Finally, in $\mathcal{S}_0$, we only consider subsets that are consistent with b when iterating over the power set of neck detections during pricing.

6.3.2 Branching

Let V^* be the optimal value of the pricing problem. After computing a lower bound V_b^{lb} at a node $b \in \mathcal{B}$, an upper bound V_b^{ub} on V^* can easily be computed from the optimal hypothesis g_b obtained by solving the ψ-independent pricing problem at node b. Indeed, it suffices to compute its reduced cost (taking into account the ψ values), i.e.,

$$V_b^{ub} = \Gamma_{g_b} - \sum_{d \in \mathcal{D}} \lambda_d G_{dg_b} - \sum_{c \in \hat{\mathcal{C}}} \psi_c C_{cg_b}. \tag{47}$$

When V_b^{lb} is less than the best upper bound found so far and $V_b^{lb} < V_b^{ub}$, branching is performed as follows. First, observe that $V_b^{ub} - V_b^{lb} = -\sum_{c \in \hat{\mathcal{C}}_b} \psi_c$, where $\hat{\mathcal{C}}_b = \{c \in \hat{\mathcal{C}} \mid C_{cg_b} - \lfloor \frac{|\mathcal{D}_c \cap \mathcal{D}_b^+|}{2} \rfloor = 1\}$ is the index set of the 3-SRIs whose duals were not considered in the computation of V_b^{lb}. Therefore, to close the gap between V_b^{lb} and V_b^{ub}, we propose to branch on the observations in a subset $\mathcal{D}_{c_b}$, where $c_b \in \arg\min_{c \in \hat{\mathcal{C}}_b} \psi_c$. Eight child nodes, denoted b_1 to b_8, are created, one for each way of splitting the observations in $\mathcal{D}_{c_b}$ between the include ($\mathcal{D}^+$) and exclude ($\mathcal{D}^-$) sets. Table 1 enumerates the splits for a triplet of observations $c_b = \{d_1, d_2, d_3\}$.

Note that not all child nodes need to be created as some are guaranteed to be infeasible if some observations in c_b already belong to D_b^- or D_b^+. For instance, if $c_b = \{d_1, d_2, d_3\}$ and $d_1 \in D_b^+$, then the nodes b_2, b_4, b_6, and b_8 will all be infeasible because d_1 belongs to both $\mathcal{D}^+$ and $\mathcal{D}^-$ sets.

Table 1: Eight different ways of splitting d_1, d_2, d_3 between the $\mathcal{D}^+$ and $\mathcal{D}^-$ sets, yielding eight child nodes b_1 to b_8 for node b. For example, node b_8 excludes d_1 and d_2 ($D^-_{b_8} = D_b^- \cup \{d_1, d_2\}$) but includes d_3 ($D^+_{b_8} = D_b^+ \cup \{d_3\}$).

Child	$\mathcal{D}^-$	$\mathcal{D}^+$	Child	$\mathcal{D}^-$	$\mathcal{D}^+$	Child	$\mathcal{D}^-$	$\mathcal{D}^+$	Child	$\mathcal{D}^-$	$\mathcal{D}^+$
b_1	d_3	d_1, d_2	b_2	d_1, d_3	d_2	b_3	d_2, d_3	d_1	b_4	d_1, d_2, d_3	$\emptyset$
b_5	$\emptyset$	d_1, d_2, d_3	b_6	d_1	d_2, d_3	b_7	d_2	d_1, d_3	b_8	d_1, d_2	d_3

7 Dual optimal inequalities

In this section, we introduce DOIs (Ben Amor et al., 2006) that are expressed in the form of lower bounds on the dual variables λ and do not remove all dual optimal solutions. DOIs decrease the dual search space that CG (with or without 3-SRIs) needs to explore and, typically, the total number of CG iterations. They have been successfully applied to various applications including stock cutting (Ben Amor et al., 2006) and image segmentation (Yarkony et al., 2012; Yarkony and Fowlkes, 2015).

In this section, we introduce for the general MWSP formulation DOIs that do not vary with $\hat{\mathcal{G}}$, the current set of columns in the RMP, (Subsection 7.1) and DOIs that vary as $\hat{\mathcal{G}}$ changes (Subsection 7.2). Finally, in Subsection 7.3, we describe how these DOIs can be defined for two of our applications.

7.1 Basic dual optimal inequalities

Recall that $\lambda_d \leq 0$ for $d \in \mathcal{D}$. For any observation $d \in \mathcal{D}$, imposing a lower bound $-\Xi_d$ on λ_d (with $\Xi_d \geq 0$) writes as $\lambda_d \geq -\Xi_d$ in the dual of the MP. In the MP (8)–(9), it corresponds to introducing

a surplus variable $\xi_d \geq 0$ in the corresponding constraint (9) that has a cost coefficient of Ξ_d in the objective function (8). The modified MP is:

$$\min_{\substack{\gamma \geq 0 \\ \xi \geq 0}} \quad \sum_{g \in \mathcal{G}} \Gamma_g \gamma_g + \sum_{d \in \mathcal{D}} \Xi_d \xi_d \tag{48}$$

$$\text{s.t.:} \quad \sum_{g \in \mathcal{G}} G_{dg} \gamma_g - \xi_d \leq 1 \quad \forall d \in \mathcal{D}. \tag{49}$$

This MP allows the inclusion of an observation in multiple selected hypotheses, which may facilitate its solution by CG. However, to ensure that this MP is equivalent to the original one (8)–(9) and that each inequality $\lambda_d \geq -\Xi_d$ is a DOI, the cost reduction induced by "over-including" an observation d should be at least compensated by Ξ_d. Notice that, in (48)–(49), no upper bounds are imposed on the variables.

For each observation $d \in \mathcal{D}$, we propose to compute the value of Ξ_d as an upper bound on the cost increase realized by removing d from a hypothesis that contains it. It must satisfy

$$\Xi_d \geq \epsilon + \max_{g \in \mathcal{G}_d} \ \max\{0, \Delta_{gd} - \Gamma_g\}, \tag{50}$$

where $\mathcal{G}_d = \{g \in \mathcal{G} \mid G_{dg} = 1\}$ is the subset of hypotheses including observation d, ϵ is a tiny positive constant which ensures that the corresponding DOI is not active at termination of CG, and

$$\Delta_{gd} \quad = \min_{\gamma \geq 0} \sum_{h \in \mathcal{G} \setminus \mathcal{G}_d} \Gamma_h \gamma_h \tag{51}$$

$$\text{s.t.:} \sum_{h \in \mathcal{G} \setminus \mathcal{G}_d} G_{\bar{d}h} \gamma_h \leq G_{\bar{d}g} \qquad \forall \bar{d} \in \mathcal{D} \setminus \{d\} \tag{52}$$

$$\gamma_h \in \{0, 1\} \qquad \forall h \in \mathcal{G} \setminus \mathcal{G}_d. \tag{53}$$

This ILP allows the computation of a least-cost (possibly empty) subset of non-overlapping hypotheses, denoted $\mathcal{H}_{gd}$, that can feasibly replace hypothesis g in a solution but without including observation d. Note that computing the value of the right-hand side of (50) may be computationally expensive and, thus, we rather compute an upper bound on its value as discussed in Section 7.3.

The validity of these DOIs, that we call invariant DOIs, is stated in the following proposition.

Proposition 2 *Let ζ^* and ζ^*_{DOI} be the optimal value of the MP (8)–(9) and the modified MP (48)–(49), respectively. If Ξ satisfies (50), then $\zeta^*_{DOI} = \zeta^*$.*

Proof. Let (γ^*, ξ^*) be an optimal solution to (48)–(49). If $\xi^*_d = 0$ for all observations $d \in \mathcal{D}$, then $\zeta^*_{DOI} = \zeta^*$ because γ^* is feasible and optimal for (8)–(9). Otherwise, there exists an observation $d \in \mathcal{D}$ such that $\xi^*_d > 0$ and, therefore, a hypothesis $g \in \mathcal{G}_d$ such that $\gamma^*_g > 0$. Let $\alpha = \min\{\gamma^*_g, \xi^*_d\}$. Given that $\Xi_d > \Delta_{gd} - \Gamma_g$ according to (50) and $\alpha > 0$, the solution obtained from (γ^*, ξ^*) by decreasing γ_g and ξ_d by α and increasing γ_h by α for all $h \in \mathcal{H}_{gd}$ is feasible for (48)–(49) and has a cost that is less than ζ^*_{DOI}. This contradicts the optimality of (γ^*, ξ^*) and proves that there is no observation $d \in \mathcal{D}$ such that $\xi^*_d > 0$. □

From this proposition, we can deduce that any RMP encountered during the solution of the modified MP (48)–(49) by CG is bounded.

7.2 Dual optimal inequalities that vary with $\hat{\mathcal{G}}$

For a given CG iteration, let $\hat{\mathcal{G}}$ be the set of generated hypotheses, i.e., those considered in the RMP. In this section, we define DOIs that are not looser than the above invariant DOIs and are functions of $\hat{\mathcal{G}}$.

Let $\hat{\mathcal{G}}^* = \{g \in \mathcal{G} \,|\, \exists h \in \hat{\mathcal{G}} \text{ such that } G_{gd} \leq G_{hd}, \forall d \in \mathcal{D}\}$ be the set of hypotheses that contain a subset of the observations of a hypothesis $g \in \hat{\mathcal{G}}$. Note that $\hat{\mathcal{G}}^* \supseteq \hat{\mathcal{G}}$. For $d \in \mathcal{D}$, the parameter Ξ_d in the proposed DOI $\lambda_d \geq -\Xi_d$ must satisfy

$$\Xi_d \geq \epsilon + \max_{g \in \hat{\mathcal{G}}^*_d} \max\{0, \Delta_{gd} - \Gamma_g\}, \tag{54}$$

where ϵ and Δ_{gd} are defined as in the previous section, and $\hat{\mathcal{G}}^*_d = \{g \in \hat{\mathcal{G}}^* \,|\, G_{dg} = 1\}$ is the subset of hypotheses in $\hat{\mathcal{G}}^*$ that include observation d. Observe that the right-hand side of (54) may increase when hypotheses are added to $\hat{\mathcal{G}}$ but it is never greater than that of (50). As for (50), we do not necessarily compute the right-hand side of (54), but rather an upper bound on its value (see Section 7.3).

With these DOIs that we call the varying DOIs, the modified RMP writes as

$$\min_{\substack{\gamma \geq 0 \\ \xi \geq 0}} \quad \sum_{g \in \hat{\mathcal{G}}} \Gamma_g \gamma_g + \sum_{d \in \mathcal{D}} \Xi_d \xi_d \tag{55}$$

$$\text{s.t.:} \quad \sum_{g \in \hat{\mathcal{G}}} G_{dg} \gamma_g - \xi_d \leq 1 \qquad \forall d \in \mathcal{D}. \tag{56}$$

The next proposition proves that it is always bounded. Its proof relies on the following lemma.

Lemma 1 *If Ξ satisfies (54) and there exists a hypothesis $g \in \hat{\mathcal{G}}$ with $\Gamma_g + \sum_{d \in \mathcal{D}} \Xi_d G_{dg} < 0$, then there exists a hypothesis $g^* \in \hat{\mathcal{G}}^*$ that includes an observation $d^* \in \mathcal{D}$ such that $\Gamma_{g^*} + \sum_{d \in \mathcal{D}} \Xi_d G_{dg^*} < 0$ and $\sum_{h \in \mathcal{H}_{g^* d^*}} \left(\Gamma_h + \sum_{d \in \mathcal{D}} \Xi_d G_{dh}\right) \geq 0$.*

Proof. By construction. Let $g \in \hat{\mathcal{G}}$ be a hypothesis such that $\Gamma_g + \sum_{d \in \mathcal{D}} \Xi_d G_{dg} < 0$. If g includes an observation $\bar{d} \in \mathcal{D}$ such that $\sum_{h \in \mathcal{H}_{g\bar{d}}} \left(\Gamma_h + \sum_{d \in \mathcal{D}} \Xi_d G_{dh}\right) \geq 0$, then $g^* = g$ and $d^* = \bar{d}$. Otherwise, for any observation $\bar{d}$ included in g, $\mathcal{H}_{g\bar{d}} \neq \emptyset$ and $\sum_{h \in \mathcal{H}_{g\bar{d}}} \left(\Gamma_h + \sum_{d \in \mathcal{D}} \Xi_d G_{dh}\right) < 0$. In this case, there exists a hypothesis $\bar{h} \in \mathcal{H}_{g\bar{d}}$ (that contains less observations than g and is, thus, in $\hat{\mathcal{G}}^*$) such that $\Gamma_{\bar{h}} + \sum_{d \in \mathcal{D}} \Xi_d G_{d\bar{h}} < 0$. We can thus repeat the above process with $g = \bar{h}$. This iterative process is finite because the number of observations in the hypothesis g selected at each iteration is strictly decreasing, ensuring that, at some point, there must exist an observation $\bar{d}$ included in g such that $\mathcal{H}_{g\bar{d}} = \emptyset$ and, therefore, $\sum_{h \in \mathcal{H}_{g\bar{d}}} \left(\Gamma_h + \sum_{d \in \mathcal{D}} \Xi_d G_{dh}\right) = 0$. □

Proposition 3 *If Ξ satisfies (54), then the modified RMP (55)–(56) is bounded.*

Proof. Assume that (55)–(56) is unbounded. In this case, there must exist at least one hypothesis $g \in \hat{\mathcal{G}}$ such that $\Gamma_g + \sum_{d \in \mathcal{D}} \Xi_d G_{dg} < 0$. According to Lemma 1, there also exists a hypothesis $g^* \in \hat{\mathcal{G}}^*$ that includes an observation $d^* \in \mathcal{D}$ such that $\Gamma_{g^*} + \sum_{d \in \mathcal{D}} \Xi_d G_{dg^*} < 0$ and $\sum_{h \in \mathcal{H}_{g^* d^*}} \left(\Gamma_h + \sum_{d \in \mathcal{D}} \Xi_d G_{dh}\right) \geq 0$. These inequalities imply that

$$\sum_{h \in \mathcal{H}_{g^* d^*}} \left(\Gamma_h + \sum_{d \in \mathcal{D}} \Xi_d G_{dh}\right) - \left(\Gamma_{g^*} + \sum_{d \in \mathcal{D}} \Xi_d G_{dg^*}\right) > 0, \tag{57}$$

which rewrites as

$$\sum_{h \in \mathcal{H}_{g^* d^*}} \Gamma_h - \Gamma_{g^*} > \sum_{d \in \mathcal{D}_{g^* d^*}} \Xi_d, \tag{58}$$

where $\mathcal{D}_{g^* d^*} \supseteq \{d^*\}$ denotes the subset of observations that are included in g^* but not in any hypothesis of $\mathcal{H}_{g^* d^*}$.

Because $\sum_{d \in \mathcal{D}_{g^* d^*}} \Xi_d > \Xi_{d^*} \geq \epsilon + \Delta_{g^* d^*} - \Gamma_{g^*}$ and $\Delta_{g^* d^*} = \sum_{h \in \mathcal{H}_{g^* d^*}} \Gamma_h$, relation (58) implies

$$\sum_{h \in \mathcal{H}_{g^* d^*}} \Gamma_h - \Gamma_{g^*} > \epsilon + \sum_{h \in \mathcal{H}_{g^* d^*}} \Gamma_h - \Gamma_{g^*}, \tag{59}$$

or equivalently, $\epsilon < 0$. This contradicts the definition of ϵ and proves that (55)–(56) is never unbounded. □

The following proposition ensures the validity of the varying DOIs.

Proposition 4 *Let ζ^* and $\hat{\zeta}^*_{DOI}$ be the optimal value of the MP (8)–(9) and the modified RMP (55)–(56), respectively. If Ξ satisfies (54), then $\hat{\zeta}^*_{DOI} = \zeta^*$ or there exists a (non-generated) hypothesis with a negative reduced cost.*

Proof. Let $(\hat{\gamma}^*, \hat{\xi}^*)$ and $\hat{\lambda}^*$ be optimal primal and dual solutions to the modified RMP (55)–(56), respectively. The proof consists of showing that, if the reduced cost of each hypothesis $g \in \mathcal{G}$ is non-negative, i.e., $\Gamma_g - \sum_{d\in\mathcal{D}} G_{dg}\hat{\lambda}^*_d \geq 0$, then $\hat{\zeta}^*_{DOI} = \zeta^*$. Assume that $\Gamma_g - \sum_{d\in\mathcal{D}} G_{dg}\hat{\lambda}^*_d \geq 0$ for all $g \in \mathcal{G}$. If $\hat{\xi}^*_d = 0$ for all $d \in \mathcal{D}$, then $\zeta^*_{DOI} = \zeta^*$ because $\hat{\gamma}^*$ (augmented with $\lambda_g = 0$ for $g \in \mathcal{G} \setminus \hat{\mathcal{G}}$) is feasible and optimal for (8)–(9). Otherwise, there exists an observation $d \in \mathcal{D}$ such that $\hat{\xi}^*_d > 0$ and a hypothesis $g \in \hat{\mathcal{G}}_d$ such that $\hat{\gamma}^*_g > 0$. In this case, the reduced costs of ξ_d and γ_g are both equal to 0, i.e., $\hat{\lambda}^*_d + \Xi_d = 0$ and $\Gamma_g - \sum_{e\in\mathcal{D}} G_{eg}\hat{\lambda}^*_e = 0$. Given that $\Gamma_h - \sum_{e\in\mathcal{D}} G_{eh}\hat{\lambda}^*_e \geq 0$ for all $h \in \mathcal{H}_{gd}$, $\hat{\lambda}^*_e \leq 0$ for all $e \in \mathcal{D}$ and $\sum_{h\in\mathcal{H}_{gd}} \Gamma_h = \Delta_{gd}$, we find that

$$\sum_{h\in\mathcal{H}_{gd}} \big(\Gamma_h - \sum_{e\in\mathcal{D}} G_{eh}\hat{\lambda}^*_e\big) - \big(\Gamma_g - \sum_{e\in\mathcal{D}} G_{eg}\hat{\lambda}^*_e\big) \geq 0 \tag{60}$$

$$\Rightarrow \sum_{h\in\mathcal{H}_{gd}} \Gamma_h - \Gamma_g \geq -\sum_{e\in\mathcal{D}_{gd}} \hat{\lambda}^*_e \geq -\hat{\lambda}^*_d \tag{61}$$

$$\Rightarrow \Delta_{gd} - \Gamma_g \geq \Xi_d. \tag{62}$$

This contradicts the definition (54) of Ξ_d which guarantees that $\Xi_d > \Delta_{gd} - \Gamma_g$. Consequently, this case is not possible. □

With these varying DOIs, the values of Ξ need to be re-computed at each CG iteration before solving the modified RMP (55)–(56). Indeed, for an observation $d \in \mathcal{D}$, the left-hand side of (54) depends on $\hat{\mathcal{G}}$ and may, thus, vary from one iteration to another.

7.3 Application-specific dual optimal inequalities

In this section, we present ways to compute Ξ values that yield valid DOIs for the multi-person pose estimation application and the multi-cell segmentation application when the anchor is not required to lie in the cell. In both cases (which are treated simultaneously below), removing any observation from any hypothesis always yields a feasible hypothesis. This is not the case for the multi-cell segmentation application when the anchor must be in the cell or for the multi-person tracking application where the set of feasible hypotheses (tracks) is restricted by the subset of subtracks considered.

For the invariant DOIs based on (50), observe first that the right-hand side of (50)

$$\epsilon + \max_{g\in\mathcal{G}_d} \max\{0, \Delta_{gd} - \Gamma_g\} \leq \epsilon + \max_{g\in\mathcal{G}_d} \max\{0, \Gamma_{\bar{g}(g,d)} - \Gamma_g\},$$

where $\bar{g}(g,d)$ is the (possibly empty) hypothesis obtained by removing observation d from hypothesis g. To find an upper bound on this expression, Wang et al. (2018) propose to evaluate the worst-case cost difference $\Gamma_{\bar{g}(g,d)} - \Gamma_g$ over all hypotheses $g \in \mathcal{G}$ and all observations $d \in \mathcal{D}$ contained in the given g as follows. The removal of an observation $d \in \mathcal{D}$ from an arbitrary hypothesis $g \in \mathcal{G}$ removes from its cost Γ_g defined by (15) or (18) the associated cost θ^1_d and any active pairwise costs $\theta^2_{d\bar{d}}$ and $\theta^2_{\bar{d}d}$, $\bar{d} \in \mathcal{D}$. Moreover, if d is the only observation in g, then θ^0 is also removed. Therefore, $\Gamma_{\bar{g}(g,d)} - \Gamma_g$ is upper bounded by the negative of the sum of these terms by considering only the negative-valued $\theta^2_{d\bar{d}}$, $\theta^2_{\bar{d}d}$ and θ^0 terms. Consequently, Ξ_d can be set to

$$\Xi_d = \epsilon + \max\ \Big\{0,\ -\big(\min\{0, \theta^0\} + \theta^1_d + \sum_{\bar{d}\in\mathcal{D}} \min\{0, \theta^2_{d\bar{d}} + \theta^2_{\bar{d}d}\}\big)\Big\}. \tag{63}$$

For the varying DOIs based on (54), the values of Ξ are produced by using the same approach as in (63), except that only the pairwise terms that could be removed when replacing members of $\hat{\mathcal{G}}^*$ with other members of $\hat{\mathcal{G}}^*$ are considered. More precisely, Ξ_d is set to

$$\Xi_d = \epsilon + \max\ \Big\{0,\ -\big(\min\{0,\theta^0\} + \theta^1_d + \min_{g\in\hat{\mathcal{G}}_d} \sum_{\bar{d}\in\mathcal{D}} G_{\bar{d}g} \min\{0,\theta^2_{d\bar{d}} + \theta^2_{\bar{d}d}\}\big)\Big\}. \tag{64}$$

8 Computational results

In this section we provide computational results on the multi-person tracking, multi-person pose estimation and multi-cell segmentation applications in Sections 8.1, 8.2, and 8.3, respectively.

8.1 Multi-person tracking

We use a part of the MOT 2015 training set (Leal-Taixé et al., 2015) to train and evaluate multi-person tracking in video. The structured SVM based learning approach of Wang and Fowlkes (2015) is used to produce the cost terms, given the raw detector outputs provided by the MOT dataset generates the set of detections $\mathcal{D}$. The models are trained with varying subtrack lengths ($K = 2, 3, 4$), and allow for occlusions of up to three frames. The experiments in this section employ the 3-SRIs but not the DOIs. To assess CG convergence, lower bounds are computed throughout the solution process. At a given CG iteration, the computed lower bound is given by

$$\sum_{d\in\mathcal{D}} \lambda_d + \sum_{d\in\mathcal{D}} \min\{0, \min_{\substack{s\in\mathcal{S}\\ s_K=d}} \ell_s\},$$

where the first term is equal to the optimal value of the current RMP and the second is the sum over all detections $d \in \mathcal{D}$ of the least reduced cost of the tracks ending with detection d if negative (for computational efficiency, the ψ terms are ignored in this computation). This bound is valid because, in a feasible solution, at most one track can end with each detection.

In the problem instance that we use for testing, there are 71 frames and 322 detections in the video. The numbers of subtracks considered are 1,068, 3,633 and 13,090 for $K = 2, 3, 4$, respectively. For $K = 2$, we observe 48.5% "Multiple Object Tracking Accuracy" (Bernardin and Stiefelhagen, 2008), 11 identity switches, and 9 track fragments, which we write in short hand as (48.5,11,9). However, when setting $K = 3, 4$, the performance is (49,10,7), and (49.9,9,7), respectively. Thus, increasing subtrack length provides improvement on all metrics. The importance of this improvement is demonstrated visually in Figure 4.

In Figure 5, we compare the timing/cost performance of our CG algorithm with the baseline dual decomposition (DD) approach of Butt and Collins (2013) for $K = 3, 4$. These problem instances are associated with a loose lower bound, which is tightened in the CG algorithm using 3-SRIs. For both instances, CG achieves tight upper and lower bounds at termination while DD does not. Furthermore, CG terminates much more rapidly than DD.

8.2 Multi-person pose estimation

We present the experimental results from Wang et al. (2018) on the MPII-multi-person dataset (Andriluka et al., 2014), which consists of 418 images. Initially, the cost terms and problem structure are provided by Insafutdinov et al. (2016). We modify them to improve modeling power and optimization speed. Notably, we require that each selected person contains exactly one neck detection. Thus, during pricing, we need only to iterate over selections of the neck detections containing one detection. We also restrict the cardinality of $\mathcal{S}^r$ for each body part $r \in \mathcal{R}$ other than the neck to be no greater than 50,000. We detail all modifications below.

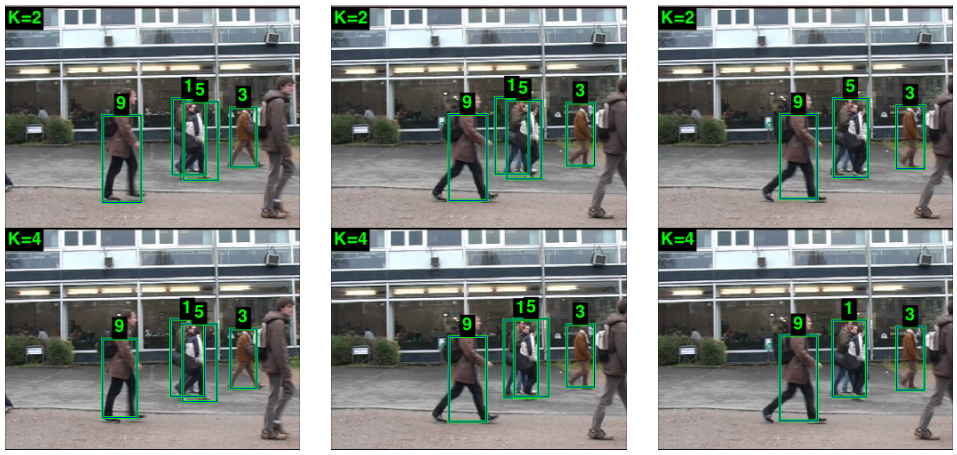


Figure 4: Qualitative example of improvement as a result of increasing subtrack length. The first and second rows describe tracks outputted when $K = 2$ and $K = 4$, respectively. Notice that, for $K = 2$, track 1 changes identity to track 5, while with $K = 4$, the identity of track 1 does not change. (Picture from Wang et al., 2017b).

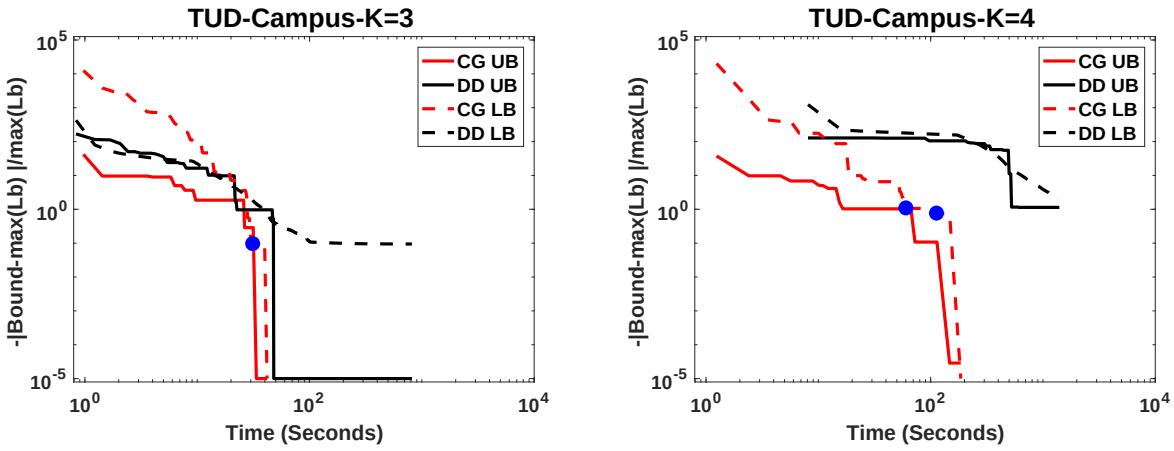


Figure 5: Convergence of upper/lower bounds as a function of time, illustrated using plots of the absolute gap between the bounds and the final lower bound. All plotted values are normalized by dividing each of them by the value of the maximum lower bound multiplied by -1. The addition of a 3-SRI is indicated with a blue dot on the lower bound plot.

1. Pairwise cost terms between detections corresponding to different body parts that are not connected in the augmented tree are ignored (set to zero).
2. Sets $\mathcal{D}^r$, $r \in \mathcal{R}$, are constructed as follows. Insafutdinov et al. (2016) provide a probability m_{dr} that each detection $d \in \mathcal{D}$ is associated with each body part $r \in \mathcal{R}$. Each detection r is assigned to a single set $\mathcal{D}^r$, namely, that with the largest probability.
3. The size of $\mathcal{S}^r$ is limited to 50,000 for each $r \in \mathcal{R}$ other than the neck. To construct $\mathcal{S}^r$, we iterate over integer k from 1 to $|\mathcal{D}^r|$ and, at each iteration, add to $\mathcal{S}^r$ the group of configurations containing exactly k detections in $\mathcal{D}^r$. If adding a group would make $|\mathcal{S}^r|$ exceeds 50,000, then the group is not added and the construction of $\mathcal{S}^r$ stops.
4. For each pair of neck detections $d_1, d_2 \in \mathcal{D}$, cost $\theta^2_{d_1 d_2} = \infty$ to enforce a single neck detection in each person.
5. To model the prior on the number of people in the image, we assume that all persons have a neck detection. This prior is modeled as follows.

(a) Let $\hat{\theta}^0$ be a desired value for θ^0. Since Insafutdinov et al. (2016) does not model a prior on the number of people, we hand set $\hat{\theta}^0$ to a single value for the entire dataset.

(b) Set $\theta^0 = \hat{\theta}^0 - \Omega$, where Ω is such that subtracting it from the cost of any person makes the cost of that person positive. Setting Ω to be less (by one) than the sum of all negative valued cost terms achieves this, i.e.,

$$\Omega = -1 + \min\{0, \hat{\theta}^0\} + \sum_{d \in \mathcal{D}} \min\{0, \theta^1_d\} + \sum_{\substack{d_1 \in \mathcal{D} \\ d_2 \in \mathcal{D}}} \min\{0, \theta^2_{d_1 d_2}\}. \tag{65}$$

(c) For each neck detection $d \in \mathcal{D}$, add Ω to θ^1_d. Observe that the Ω terms cancel out in the cost of a person if a single neck detection is included. However, if no neck detection is included, then the cost of the person is positive. Similarly, if two or more neck detections are included in a given person, then the cost of that person is infinite. Thus, no optimal solution to the MWSP formulation selects a person containing a number of neck detections not equal to one.

For the experiments in this section, the CG algorithm does not apply 3-SRIs. Furthermore, DOIs are not applied for the results of Wang et al. (2018) that are reviewed in Table 2. However, we study their impact of using them in Section 8.2.1.

The results of the experiments of Wang et al. (2018) show that the MWSP LP relaxation is tight for more than 99% of tested instances, and in the remaining cases, the gap between the lower and upper bounds is less than 1.5%. In Table 2, we compare the CG algorithm against the greedy heuristic optimization procedure of Levinkov et al. (2017) in terms of the accuracy on standard computer vision benchmarks. The approach of Levinkov et al. (2017) considers a distinct but related objective function to our MWSP objective. The CG algorithm outperforms the heuristic of Levinkov et al. (2017) on hard-to-localize body parts, such as wrists and ankles, but fails for body parts closer to the head. This could be a side effect of the fact that costs from Insafutdinov et al. (2016) are trained on the power set of all detections including neck, thus pose associated with multiple neck detections could be a better choice for certain cases. In a more robust model, one could make a reliable head/neck detector, restricting each person to have only one head/neck. Some sample outputs of our algorithm can be visualized in Figure 6. Note that the dynamic programming pricing problems are solved using the NBD technique of Wang et al. (2018), which provides a speedup factor of up to 500 times for these instances.

Table 2: Average precision (in %) of the CG algorithm versus (Levinkov et al., 2017). Columns mAP and mAP(Ubody) indicate the mean average precision across all body parts and across all upper body parts (excluding hips, knees and ankles), respectively. Running times are measured on an Intel i7-6700k quad-core CPU.

	Shoulder	Elbow	Head	Wrist	Hip	Knee	Ankle	mAP (UBody)	mAP	time (s/frame)
Levinkov et al. (2017)	**88.2**	78.2	**93.0**	68.4	**78.9**	70.0	64.3	**81.9**	**77.6**	0.136
CG	87.3	**79.5**	90.6	**70.1**	78.5	**70.5**	**64.8**	81.8	**77.6**	1.95

8.2.1 Value of dual optimal inequalities

In this section we aim at assessing the speedup that can be obtained by using DOIs. However, this speedup varies with the "system" used to solve the RMP at each CG iteration, where a system is defined by an LP toolbox (linprog, CPLEX, Gurobi), the toolbox options such as the algorithm used (interior point, primal simplex, etc.), and the computer used. Indeed, different systems provide different dual optimal solutions that might impact differently the number of CG iterations required to achieve optimality. Using dual optimal solutions that are well centered in the optimal dual face is known to yield faster CG convergence (Desrosiers and Lübbecke, 2005). Therefore, to establish the value of the DOIs, we need to decouple the DOI value from that obtained by varying the "system" used to solve the RMP.

Figure 6: Output examples. For each person, the locations of the detections of each body part are averaged to produce the corresponding colored dot, denoting the part position. (Picture from Wang et al., 2018).

For our test instances, the time spent solving the pricing problems vastly exceeds that for solving the RMP. Thus, using high performance ILP solvers (such as CPLEX or Gurobi) to solve the RMP adds little value if the resulting dual solution is not well centered. The systems we compared are as follows.

System One: MATLAB 2016 linprog solver with default settings, on a 2014 Macbook Pro.

System Two: MATLAB 2017 with the interior point solver on a powerful workstation with Intel(R) Core(TM) i7-6850K CPU @ 3.60GHz.

We also performed the experiments using Gurobi and CPLEX solvers. Interestingly, the Gurobi and CPLEX solvers with default options perform worse on the workstation than the built-in MATLAB solver when DOIs are not used. Thus we did not include those systems in the comparisons.

Table 3 compares the speedups obtained by the DOIs using the two systems, whereas Figure 7 provides a scatter plot of the time consumed using the DOIs for each system. Our experiments show that varying DOIs outperform invariant DOIs. The use of DOIs provides a large speedup for System Two (over ten times speedup), but limited speedup for System One (only 1.4-1.6 times speedup). Interestingly, compared to System Two, System One is an older computer running an older version of MATLAB, but the running times with System One are better than those with System Two when no DOIs are used. This is a consequence of System One producing well centered dual solutions, and System Two not doing so. The use of DOIs makes System Two perform better than System One, demonstrating the value of DOIs when the system is poorly selected.

Table 3: Total computing times in seconds and comparative speedups (over no DOI) using DOIs on different systems. For each system, the first three columns indicate the total time (summing over all problem instances) needed to solve the MP for CG using no DOI, invariant DOIs, and varying DOIs, respectively. The last two columns display the speedup factors achieved by using invariant and varying DOIs over not using DOIs.

System	No DOI	Invariant DOIs	Varying DOIs	Speed Up Invariant	Speed Up Varying
One	2092.6	1450.5	1290.5	1.44	1.62
Two	9821.2	937.6	860.2	10.47	11.42

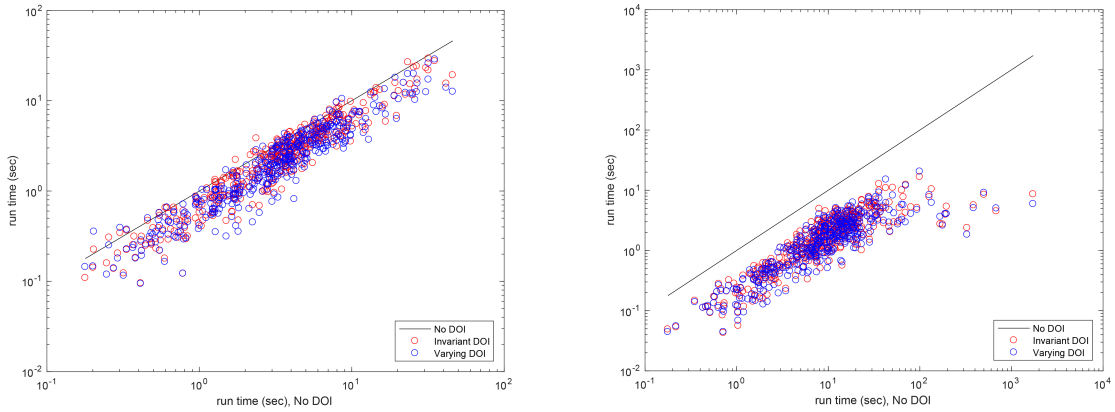


Figure 7: Relative computational time when using DOIs on System One (left) and System Two (right). Each data point specifies the computing time to solve the MP with DOIs (vertical axis) relative to the computing time when not using DOI (horizontal axis).

8.3 Multi-cell segmentation

In this section, CG is applied for multi-cell segmentation on three different datasets containing between 10 and 15 images each. These problem instances include challenging properties such as densely packed and touching cells, out-of-focus artifacts, and variations in the shape/size of cells. To generate cost terms, we use the open source toolbox of Sommer et al. (2011) to train a random forest classifier to discriminate between: (1) boundaries of in-focus cells; (2) in-focus cells; (3) out-of-focus cells; and (4) background. For training, we use less than 1% of the pixels per dataset with generic features, e.g., Gaussian, Laplacian, and structured tensor. The output of this random forest classifier is also used to generate super-pixels.

For the experiments in this section, the CG algorithm is enhanced with the 3-SRIs, but not with the DOIs. Note that, during pricing, we enforce the inclusion of the anchor in the cell produced.

8.3.1 Segmentation quality

In Figure 8, we display the output obtained by the CG algorithm for three images, one in each dataset. For dataset two (the middle column), we observe that our approach successfully segments the cells in a problem instance where there are large variations of cell shape/size.

Next, we compare the performance of our algorithm with state-of-the-art methods (Arteta et al., 2012, 2016; Funke et al., 2015; Hilsenbeck et al., 2017; Dimopoulos et al., 2014; Ronneberger et al., 2015; Zhang et al., 2014b) in terms of detection (precision, recall and F-score) and segmentation (Dice coefficient and Jaccard index), which are common measures in bio-image analysis. The average results of our experiments over all instances of each dataset are summarized in Table 4. In general, the CG algorithm achieves or exceeds state-of-the-art performance. Additionally, it requires very little training data compared to other methods, including those of Arteta et al. (2012, 2016) and Funke et al. (2015).

8.3.2 Optimization performance

To determine the best values of the parameters θ^0, R_{max} and A_{max}, a large number of problem instances were generated by varying the values of these parameters. In this section, we report average results over all these instances. For each instance, we computed the relative optimality gap between the upper and lower bounds obtained at termination of CG. For the three datasets, the proportions of instances that achieve an optimality gap of 10% or less are 99.3 %, 80 % and 100 % on datasets one, two, and three, respectively, showing the high quality of the computed solutions.

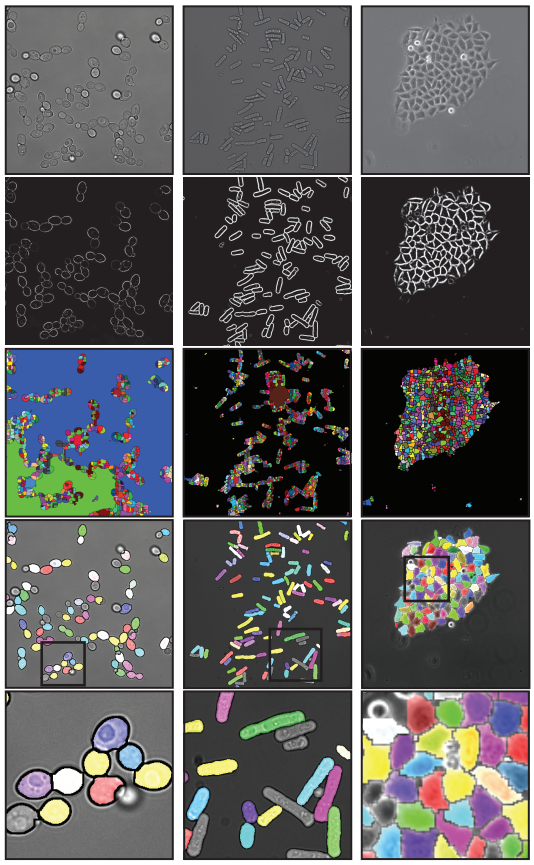


Figure 8: Cell segmentation examples for datasets one to three (left to right). Rows are (top to bottom): original image, cell boundary classifier prediction image, super-pixels, color map of segmentation, and enlarged views of the inset (black square). (Picture from Zhang et al., 2017).

To illustrate the time required by the CG algorithm to solve these instances, we show in Figure 9 the proportion of instances from dataset one for which the computational time exceeds a given amount of time. We observe, for example, that around 50% of the instances are solved in less than 200 seconds. Given that the computational time is dominated by pricing and there are many pricing problems, parallelization can drastically accelerate CG.

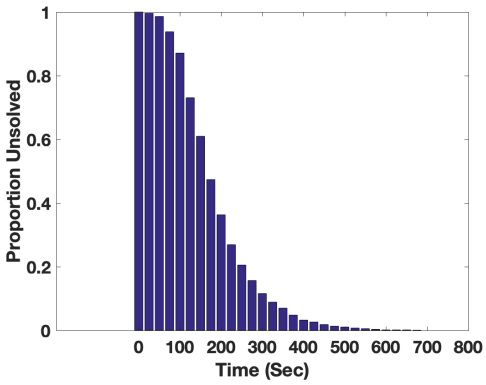


Figure 9: Proportion of instances from dataset one that take a longer computational time than a given amount of time.

Table 4: Comparative average results for datasets one to three on precision (P), recall (R), F-score (F), dice coefficient (D) and Jaccard index (J) obtained by the CG algorithm and state-of-the-art methods. PCC and NPCC denote the planar correlation clustering and non-planar correlation clustering algorithms of Zhang et al. (2014b). A '-' indicates that the result is unreported in the cited paper.

Dataset	**1**					**2**					**3**				
Metric	P	R	F	D	J	P	R	F	D	J	P	R	F	D	J
Arteta et al. (2012)	-	-	-	-	-	-	-	-	-	-	0.89	0.86	0.87	-	-
(Arteta et al., 2016)	-	-	-	-	-	-	-	-	-	-	0.99	0.96	0.97	-	-
Funke et al. (2015)	0.93	0.89	0.91	0.90	0.82	0.99	0.90	0.94	0.90	0.83	0.95	0.98	0.97	**0.84**	0.73
Hilsenbeck et al. (2017)	-	-	-	-	-	-	-	-	-	-	-	-	0.97	-	**0.75**
Dimopoulos et al. (2014)	-	-	-	-	0.87	-	-	-	-	-	-	-	-	-	-
Ronneberger et al. (2015)	-	-	-	-	-	-	-	-	-	-	-	-	0.97	-	0.74
PCC Zhang et al. (2014b)	0.95	0.86	0.90	0.87	0.84	0.80	0.75	0.76	**0.91**	**0.85**	0.92	0.92	0.92	0.79	0.72
NPCC Zhang et al. (2014b)	0.71	0.96	0.82	0.86	0.89	0.75	0.83	0.78	0.91	0.84	0.85	**0.97**	0.90	0.80	0.70
CG	**0.99**	**0.97**	**0.98**	**0.91**	**0.90**	**1.00**	**0.94**	**0.97**	0.90	0.83	**1.00**	**0.97**	**0.99**	0.82	0.71

9 Conclusion

In this paper, we introduced the problem of data association for computer vision to the operations research community. Our aim is to create a platform from which operations research scientists, with an expertise in integer programming, can apply their methodologies to machine learning and computer vision problems. This paper is designed to be accessible to those familiar with the theory and use of CG in typical operations research contexts such as vehicle routing, crew scheduling, and stock cutting.

This paper is devoted to the MWSP formulation of data association, where an instance is parameterized by a set of observations and a set of possible hypotheses that are each defined by a subset of observations. The MWSP formulation searches for the least total cost set of hypotheses such that no two selected hypotheses share a common observation. We modeled as a MWSP problem the well-studied computer vision applications of multi-person tracking, multi-person pose estimation, and multi-cell segmentation. Because the set of hypotheses in the MWSP model grows exponentially with the size of the set of observations, we employed a CG algorithm to solve it. In our applications, the CG pricing problems correspond to tractable optimization problems, that are either dynamic programs or small-scale binary integer programs. To tighten the MP, we considered SRIs and demonstrated that they can be used for our applications without destroying the structure of the pricing problem or while exploiting its initial structure within a branch-and-bound framework. To accelerate CG, we introduced new DOIs that depend on the set of columns in the current RMP. The computational results reported showed the effectiveness of the proposed CG algorithms for our applications.

In future work, we recommend that the cost of hypotheses be dependent on a higher-level variable. For example, in the context of multi-person pose estimation, the cost of a person could be conditioned on the action (the higher-level variable) being taken by the person, e.g., standing, sitting, dancing, etc. Each action would, then, be associated with a different tree structure and cost terms. A similar approach can be applied for multi-cell segmentation where the higher-level variable could correspond to the cell species or cell orientation (with regards to rotation). The higher level variable can be a latent variable and the corresponding learning problem attacked using a latent structured SVM (Yu and Joachims, 2009).

References

Abrams Z, Mendelevitch O, Tomlin J (2007) Optimal delivery of sponsored search advertisements subject to budget constraints. Proc. 8th ACM Conference on Electronic Commerce, 272–278 (San Diego, California).

Andres B, Kappes JH, Beier T, Kothe U, Hamprecht FA (2011) Probabilistic image segmentation with closedness constraints. Proc. 13th International Conference on Computer Vision, 2611–2618 (Barcelona, Spain).

Andres B, Kroger T, Briggman KL, Denk W, Korogod N, Knott G, Kothe U, Hamprecht FA (2012) Globally optimal closed-surface segmentation for connectomics. Proc. 12th International Conference on Computer Vision (Florence, Italy).

Andriluka M, Pishchulin L, Gehler P, Schiele B (2014) 2D human pose estimation: New benchmark and state of the art analysis. Proc. 27th Conference on Computer Vision and Pattern Recognition, 3686–3693 (Columbus, Ohio).

Arteta C, Lempitsky V, Noble J, Zisserman A (2012) Learning to detect cells using non-overlapping extremal regions. Proc. 15th International Conference on Medical Image Computing and Computer-Assisted Intervention, 348–356 (Nice,France).

Arteta C, Lempitsky V, Noble J, Zisserman A (2016) Detecting overlapping instances in microscopy images using extremal region trees. Medical Image Analysis 27:3–16.

Bansal N, Blum A, Chawla S (2004) Correlation clustering. Journal of Machine Learning 56(1–3):89–113.

Barnhart C, Johnson EL, Nemhauser GL, Savelsbergh MWP, Vance PH (1996) Branch-and-price: Column generation for solving huge integer programs. Operations Research 46:316–329.

Ben Amor H, Desrosiers J, Valério de Carvalho JM (2006) Dual-optimal inequalities for stabilized column generation. Operations Research 54(3):454–463.

Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. Numerische Mathematik 4(1):238–252.

Bernardin K, Stiefelhagen R (2008) Evaluating multiple object tracking performance: the clear mot metrics. Journal on Image and Video Processing 2008:1.

Birge JR (1985) Decomposition and partitioning methods for multistage stochastic linear programs. Operations Research 33(5):989–1007.

Boykov Y, Kolmogorov V (2004) 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.

Butt A, Collins R (2013) Multi-target tracking by lagrangian relaxation to min-cost network flow. Proc. 26th Conference on Computer Vision and Pattern Recognition, 1846–1853 (Portland, Oregon).

Costa L, Contardo C, Desaulniers G (2019) Exact branch-price-and-cut algorithms for vehicle routing. Transportation Science Forthcoming.

Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. Proc. 18th Conference on Computer Vision and Pattern Recognition, volume 1, 886–893 (San Diego, California).

Dantzig GB, Wolfe P (1960) Decomposition principle for linear programs. Operations Research 8(1):101–111.

Delorme M, Iori M, Martello S (2016) Bin packing and cutting stock problems: Mathematical models and exact algorithms. European Journal of Operational Research 255(1):1–20.

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. Proc. 22nd Conference on Computer Vision and Pattern Recognition, 248–255 (Miami, Florida).

Desai C, Ramanan D, Fowlkes CC (2011) Discriminative models for multi-class object layout. International Journal of Computer Vision 95(1):1–12.

Desaulniers G, Desrosiers J, Solomon MM, eds. (2005) Column Generation (Springer, New York), 1st edition.

Desaulniers G, Desrosiers J, Solomon MM, Soumis F, Villeneuve D, et al. (1998) A unified framework for deterministic time constrained vehicle routing and crew scheduling problems. Crainic TG, Laporte G, eds., Fleet Management and Logistics, 57–93 (Boston, MA: Springer).

Desrochers M, Desrosiers J, Solomon M (1992) A new optimization algorithm for the vehicle routing problem with time windows. Operations Research 40(2):342–354.

Desrosiers J, Lübbecke ME (2005) A primer in column generation. Desaulniers G, Desrosiers J, Solomon MM, eds., Column Generation, 1–32 (New York, NY: Springer).

Dimopoulos S, Mayer C, Rudolf F, Stelling J (2014) Accurate cell segmentation in microscopy images using membrane patterns. Bioinformatics 30(18):2644–2651.

Felzenszwalb P, McAllester D, Ramanan D (2008) A discriminatively trained, multiscale, deformable part model. Proc. 30th Conference on Computer Vision and Pattern Recognition, 1–8 (Anchorage, Alaska).

Funke J, Hamprecht F, Zhang C (2015) Learning to segment: Training hierarchical segmentation under a topological loss. Proc. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, 268–275 (Munich, Germany).

Gamache M, Soumis F, Marquis G, Desrosiers J (1999) A Column Generation Approach for Large-scale Aircrew Rostering Problems. Operations Research 47(2):247–263.

Gilmore P, Gomory R (1961) A linear programming approach to the cutting-stock problem. Operations Research 9(6):849–859.

Hilsenbeck O, Schwarzfischer M, Loeffler D, Dimopoulos S, Hastreiter S, Marr C, Theis F, Schroeder T (2017) fastER : a user-friendly tool for ultrafast and robust cell segmentation in large-scale microscopy. Bioinformatics 33(13):2020–2028.

Insafutdinov E, Pishchulin L, Andres B, Andriluka M, Schiele B (2016) Deepercut: A deeper, stronger, and faster multi-person pose estimation model. Proc. 14th European Conference on Computer Vision, 34–50 (Amsterdam, The Netherlands).

Jepsen M, Petersen B, Spoorendonk S, Pisinger D (2008) Subset-row inequalities applied to the vehicle-routing problem with time windows. Operations Research 56(2):497–511.

Karp RM (1972) Reducibility among combinatorial problems. Proc. Symposium on the Complexity of Computer Computations, 85–103 (Yorktown Heights, New York).

Kasirzadeh A, Saddoune M, Soumis F (2017) Airline Crew Scheduling: Models, Algorithms, and Data Sets. EURO Journal on Transportation and Logistics 6(2):111–137.

Kolmogorov V (2006) Convergent tree-reweighted message passing for energy minimization. IEEE Rransactions on Pattern Analysis and Machine Intelligence 28(10):1568–1583.

Komodakis N, Paragios N, Tziritas G (2007) Mrf optimization via dual decomposition: Message-passing revisited. Proc. 11th International Conference on Computer Vision, 1–8 (Rio de Janeiro, Brazil).

Leal-Taixé L, Milan A, Reid I, Roth S, Schindler K (2015) MOTChallenge 2015: Towards a benchmark for multi-target tracking. arXiv preprint arXiv:1504.01942 .

Leal-Taixe L, Pons-Moll G, Rosenhahn B (2012) Branch-and-price global optimization for multi-view multi-target tracking. Proc. 25th Conference on Computer Vision and Pattern Recognition, 1987–1994 (Providence, Rhode Island).

Levinkov E, Uhrig J, Tang S, Omran M, Insafutdinov E, Kirillov A, Rother C, Brox T, Schiele B, Andres B (2017) Joint graph decomposition and node labeling: Problem, algorithms, applications. Proc. 30th Conference on Computer Vision and Pattern Recognition, 6012–6020 (Honolulu, Hawaii).

Magnanti TL, Wong RT (1981) Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria. Operations Research 29(3):464–484.

Pishchulin L, Insafutdinov E, Tang S, Andres B, Andriluka M, Gehler PV, Schiele B (2016) Deepcut: Joint subset partition and labeling for multi person pose estimation. Proc. 22nd Conference on Computer Vision and Pattern Recognition, 4929–4937 (Las Vegas, Nevada).

Ren X, Malik J (2003) Learning a classification model for segmentation. Proc. 16th International Conference on Computer Vision and Pattern Recognition, 10–17 (Madison, Wisconsin).

Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. Proc. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, 234–241 (Munich, Germany).

Rumelhart DE, Hinton GE, Williams RJ (1985) Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science.

Shih WK, Wu S, Kuo Y (1990) Unifying maximum cut and minimum cut of a planar graph. IEEE Transactions on Computers 39(5):694–697.

Silberman N, Sontag D, Fergus R (2014) Instance segmentation of indoor scenes using a coverage loss. Proc 14th European Conference on Computer Vision, 616–631 (Zurich, Switzerland).

Sommer C, Straehle C, Koethe U, Hamprecht FA (2011) Ilastik: Interactive learning and segmentation toolkit. Proc. 8th International Symposium on Biomedical Imaging, 230–233 (Beijing, China).

Sontag D, Meltzer T, Globerson A, Jaakkola T, Weiss Y (2008) Tightening LP relaxations for MAP using message passing. Proc. 24th, Conference on Uncertainty in Artificial Intelligence, 503–510 (Helsinki, Finland).

Tsochantaridis I, Joachims T, Hofmann T, Altun Y (2005) Large margin methods for structured and interdependent output variables. Journal Machine Learning Research 6:1453–1484.

Wang S, Fowlkes C (2015) Learning optimal parameters for multi-target tracking. Proc. 26th British Machine Vision Conference, 484–501 (Swansea, England).

Wang S, Ihler A, Kording K, Yarkony J (2018) Accelerating dynamic programs via nested benders decomposition with application to multi-person pose estimation. Proc. 15th European Conference on Computer Vision, 652–666 (Munich, Germany).

Wang S, Kording K, Yarkony J (2017a) Exploiting skeletal structure in computer vision annotation with Benders decomposition. arXiv preprint arXiv:1709.04411 .

Wang S, Wolf S, Fowlkes C, Yarkony J (2017b) Tracking objects with higher order interactions via delayed column generation. Proc. 20th International Conference on Artificial Intelligence and Statistics, 1132–1140 (Fort Lauderdale, Florida).

Wang S, Zhang C, Gonzalez-Ballester MA, Ihler A, Yarkony J (2017c) Multi-person pose estimation via column generation. arXiv preprint arXiv:1709.05982 .

Yarkony J, Fowlkes C (2015) Planar ultrametrics for image segmentation. Proc. 28th Advances in Neural Information Processing Systems, 64–72 (Montreal, Quebec).

Yarkony J, Ihler A, Fowlkes C (2012) Fast planar correlation clustering for image segmentation. Proc. 12th European Conference on Computer Vision, 1169–1176 (Florence, Italy).

Yu CN, Joachims T (2009) Learning structural SVMs with latent variables. Proc. 26th International Conference on Machine Learning, 1169–1176 (Montreal, Quebec).

Zhang C, Huber F, Knop M, Hamprecht FA (2014a) Yeast cell detection and segmentation in bright field microscopy. Proc. 11th International Symposium on Biomedical Imaging, 1267–1270 (Beijing, China).

Zhang C, Wang S, Gonzalez-Ballester MA, Yarkony J (2017) Efficient column generation for cell detection and segmentation. arXiv preprint arXiv:1709.07337 .

Zhang C, Yarkony J, Hamprecht FA (2014b) Cell detection and segmentation using correlation clustering. Proc. 17th International Conference on Medical Image Computing and Computer-Assisted Intervention, 9–16 (Boston, Massachusetts).

Zhang L, Li Y, Nevatia R (2008) Global data association for multi-object tracking using network flows. Proc. 21st Conference on Computer Vision and Pattern Recognition, 1–8 (Anchorage, Alaska).