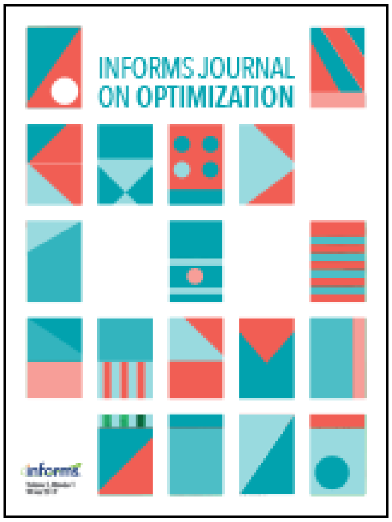


INFORMS Journal on Optimization

Data Association via Set Packing for Computer Vision Applications

Julian Yarkony, Yossiri Adulyasak, Maneesh Singh, Guy Desaulniers

informs
http://pubsonline.informs.org/journal/ijoo

INFORMS JOURNAL ON OPTIMIZATION
Vol. 2, No. 3, Summer 2020, pp. 167–191
ISSN 2575-1484 (print), ISSN 2575-1492 (online)

Data Association via Set Packing for Computer Vision Applications

Julian Yarkony,[a] Yossiri Adulyasak,[b,c] Maneesh Singh,[a] Guy Desaulniers[c,d]

[a] Computational and Human Intelligence Laboratory, Verisk AI, Jersey City, New Jersey 07310; [b] HEC Montréal, Montreal, Quebec H3T 2A7, Canada; [c] Group for Research in Decision Analysis (GERAD), Montreal, Quebec H3T 1J4, Canada; [d] Polytechnique Montréal, Montreal, Quebec H3T 1J4, Canada

Contact: julian.yarkony@verisk.com (JY); yossiri.adulyasak@hec.ca, https://orcid.org/0000-0002-6996-0742 (YA); msingh@verisk.com (MS); guy.desaulniers@gerad.ca, https://orcid.org/0000-0003-4469-9813 (GD)

Received: June 13, 2019
Revised: December 7, 2019
Accepted: January 31, 2020
Published Online in Articles in Advance: July 22, 2020

https://doi.org/10.1287/ijoo.2019.0030

Abstract. Significant progress has been made in the field of computer vision because of 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 are 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 review paper, we present a unified framework based on column generation for some computer vision applications, namely multiperson tracking, multiperson pose estimation, and multicell segmentation, which can be formulated as set packing problems with a massive number of variables. To solve them, 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 over benchmark approaches. The common framework presented allows us to leverage operations research methodologies to efficiently tackle computer vision problems.

Keywords: data association • computer vision • set packing • column generation

1. Introduction

Artificial neural networks (ANNs; 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 (e.g., dog, horse, car, etc.) using large amounts of labeled training data. An ANN learns a function, which generalizes beyond the training data set, so as to produce the correct label as output on test data that are 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 14 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 to as *data association*, is challenging because the hypotheses need not be mutually consistent. For example, multiple predicted hypotheses may share a common pixel, but, clearly, multiple objects cannot be associated with the same pixel in the ground truth. Heuristics called *nonmax suppression* (Dalal and Triggs 2005) are often used to remove conflicts between predicted hypotheses. Nonmax suppression removes from consideration all but one of each set of “similar” and/or overlapping predictions. Combinatorial optimization provides a principled alternative to nonmax 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 a hypothesis. We motivate the use of data association with the examples of multiperson detection, 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.

From 2000 to 2016, the use of combinatorial optimization in computer vision/machine learning has focused on network flows (Boykov and Kolmogorov 2004, Zhang et al. 2008, Butt and Collins 2013), 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). Unfortunately, this body of research has developed largely without influence from the operations research community, which may have led to less efficient solvers than desirable. However, and perhaps more important, 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, Desaulniers et al. 2005) and (nested) Benders decomposition ((N)BD; Birge 1985, Benders 1962) have been introduced to the machine learning and computer vision communities (Yarkony et al. 2012; Leal-Taixe et al. 2012; Yarkony and Fowlkes 2015; Wang et al. 2017a, b, c, 2018; 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 paper 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 thus be thought of as noise.

Given the huge number of possible hypotheses, the MWSP is usually solved by a CG algorithm that starts with a small subset of hypotheses and enlarges it as needed. New hypotheses are identified by solving one or several pricing problems, which are specific to each application. CG is often enhanced by generating cutting planes, such as the subset-row inequalities (Jepsen et al. 2008), for the MWSP. Furthermore, to fight the degeneracy that is inherent to the MWSP when solved by CG, dual-optimal inequalities (DOIs; Ben Amor et al. 2006) can be applied.

This review paper makes the following contributions to research in combinatorial optimization for computer vision. First, we provide a comprehensive review of the literature on MWSP formulations for computer vision problems, with a particular focus on the recent works of Zhang et al. (2017) and Wang et al. (2017b, 2018). This review targets an operations research audience. Second, through this review, we propose a unified CG framework and motivating examples to allow the combinatorial optimization community to apply their methodologies to these problems. Third, our paper extends the work of Wang et al. (2017b) to produce a general approach for applying subset-row inequalities so as to exploit the initial structure of the pricing problem. Fourth, we introduce novel DOIs 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 paper as follows. In Section 2, we present a generic MWSP formulation for data association and a CG algorithm that can be applied for solving it. In Section 3, three computer vision applications are described, and the MWSP formulation is specialized for each of them. In Section 4, CG pricing is presented for each application. In Section 5, subset-row inequalities used to tighten the LP relaxation of the MWSP formulation are exposed. In Section 6, the use of DOIs to bound the dual variables and accelerate optimization is discussed. In Section 7, computational results regarding computation time, tightness of bounds, and accuracy relative to the ground truth are reported for these three problems. In Section 8, we conclude and discuss extensions.

We wish to emphasize that most of the material presented afterward is based on the works of Wang et al. (2017b) for multiperson tracking, Wang et al. (2017c) for multiperson pose estimation, and Zhang et al. (2017) for multicell segmentation. They all used a MWSP formulation and a similar CG algorithm.

2. MWSP Formulation and CG

A basic data-association problem in computer vision can be stated as a correlation clustering problem (Bansal et al. 2004). Let $\mathcal{D}$ and $\mathcal{E}$ be the node and edge sets of an undirected graph, respectively, where edge $\langle d_1, d_2\rangle \in \mathcal{E}$ has a weight $\theta_{d_1 d_2} \in \mathbb{R}$, which might be positive, negative, or equal to zero. Correlation clustering partitions the nodes in $\mathcal{D}$ into clusters so as to minimize the sum of the weights of the intracluster edges, that is, those linking nodes in a same cluster. Correlation clustering, which is known to be nondeterministic polynomial-time hard, can be formulated as a compact integer linear program (ILP) with an easily enumerated set of variables and constraints (Kappes et al. 2016). However, this ILP has, in general, a very weak LP relaxation, yielding a very slow branch-and-bound solution process.

In Section 2.1, we briefly review how the cost terms (weights) $\theta_{d_1 d_2}$ can be computed in correlation clustering problems, that is, basic data-association problems. Then, in Section 2.2, we present an extended MWSP formulation (with an exponential number of variables) for correlation clustering, which yields a tighter LP relaxation than the compact formulation. Such a formulation was first proposed by Yarkony et al. (2012) and Leal-Taixe et al. (2012) for a computer vision application and adopted later by Wang et al. (2018) and Zhang et al. (2017). Finally, in Section 2.3, we discuss how CG can be applied to solve this extended formulation.

2.1. Learning Cost Terms

In correlation clustering, the cost terms $\theta_{d_1 d_2}$ are often produced by training a standard linear classifier to determine the probability that a given pair of nodes is associated with the same cluster in the ground truth (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 because it does not consider use of the optimization model.

To take into account this model, structured support vector machines (SVMs) are employed (Tsochantaridis et al. 2005). A structured SVM learns a mechanism to produce cost terms such that the optimal solution to the optimization model is similar to the ground truth. The structured SVM takes as input features that are 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 is solved using a cutting-plane approach. For detailed studies on the calculation of cost terms, we refer readers to Insafutdinov et al. (2016), Pishchulin et al. (2016), and Wang and Fowlkes (2015).

2.2. MWSP Formulation for Correlation Clustering

Consider the power set of $\mathcal{D}$ denoted by $\mathcal{G}$. The sets $g \in \mathcal{G}$ are described using a matrix $G \in \{0,1\}^{|\mathcal{D}|\times|\mathcal{G}|}$, where $G_{dg} = 1$ if and only if node $d \in \mathcal{D}$ is in g. Correlation clustering corresponds to selecting a nonoverlapping subset of $\mathcal{G}$. Each selected member of $\mathcal{G}$ is a cluster in a 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 by Γ_g and is defined as the sum of the weights of all edges between the nodes it contains; that is,

$$\Gamma_g = \sum_{\langle d_1,d_2\rangle\in\mathcal{E}} G_{d_1 g} G_{d_2 g} \theta_{d_1 d_2}. \tag{1}$$

A selection of sets in $\mathcal{G}$ is represented using variables $\gamma \in \{0,1\}^{|\mathcal{G}|}$; that is, $\gamma_g = 1$ if cluster $g \in \mathcal{G}$ is selected and $\gamma_g = 0$ otherwise. In what follows, correlation clustering is framed as selecting the least-cost nonoverlapping subset of $\mathcal{G}$ or, equivalently, as the MWSP problem

$$\min_{\gamma\geq 0} \quad \sum_{g\in\mathcal{G}} \Gamma_g \gamma_g \tag{2}$$

$$\text{subject to (s.t.)} \quad \sum_{g\in\mathcal{G}} G_{dg}\gamma_g \leq 1, \quad \forall d \in \mathcal{D} \tag{3}$$

$$\gamma_g \in \{0,1\}, \qquad \forall g \in \mathcal{G}. \tag{4}$$

Objective function (2) consists of minimizing the sum of the costs of the selected clusters. Constraints (3) 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 (4) enforce that γ is binary. In a CG context, formulations (3) and (4) are referred to as the *integer master problem* (IMP), and its LP relaxation is called the *master problem* (MP).

Note that in a MWSP formulation, $\mathcal{G}$ can be a subset of the power set of $\mathcal{D}$, which contains only the feasible hypotheses. Thus, the definition of $\mathcal{G}$ depends on the application considered. In Section 3, we discuss the MWSP formulation of three computer vision applications, more specifically, their definitions of sets $\mathcal{D}$ and $\mathcal{G}$, and of the cost Γ_g of a cluster (hypothesis) $g \in \mathcal{G}$.

2.3. Solving the MWSP Formulation via CG

The most apparent difficulty in solving the MWSP formulation ((2)–(4)) is the potentially massive size of its set of variables. For the computer vision problems considered, it is not feasible to enumerate all possible variables. CG, which is used identically for our applications, circumvents the problem of considering the entire set $\mathcal{G}$ by constructing a subset of $\mathcal{G}$ denoted by $\hat{\mathcal{G}}$. It 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 is restricted to the current subset $\hat{\mathcal{G}}$; that is,

$$\min_{\gamma \geq 0} \quad \sum_{g \in \hat{\mathcal{G}}} \Gamma_g \gamma_g \tag{5}$$

$$\text{s.t.} \quad \sum_{g \in \hat{\mathcal{G}}} G_{dg} \gamma_g \leq 1, \quad \forall d \in \mathcal{D}, \tag{6}$$

is solved using a LP 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 (6).

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 nongenerated 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

$$\min_{g \in \mathcal{G}} \quad \Gamma_g - \sum_{d \in \mathcal{D}} G_{dg} \lambda_d. \tag{7}$$

In practice, the pricing problem (7) is not solved by explicitly considering each set $g \in \mathcal{G}$ but instead is solved as an integer program or, very commonly, a dynamic program. It is thus specific to each application. In Section 4, we present the pricing problem used for each computer vision application considered in this paper. The CG process stops when no more negative-reduced-cost variables exist.

A pseudocode 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 ((5) and (6)) and adding to $\hat{\mathcal{G}}$ sets found by solving the pricing problem (7). When no negative-reduced-cost variables exist, CG terminates. For practical computer vision instances, the computed MP optimal solution is nearly always integral. When this is not the case, a heuristic 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). Such a heuristic is called a *restricted master heuristic* (see Joncour et al. 2010).

Algorithm 1 (Basic Column Generation)

1: $\hat{\mathcal{G}} \leftarrow \emptyset$
2: **repeat**
3: $\quad \lambda, \gamma \leftarrow$ Solve the RMP ((5) and (6)).
4: $\quad g^* \leftarrow$ Solve the pricing problem (7).
5: $\quad$ **if** $\Gamma_{g^*} - \sum_{d \in \mathcal{D}} G_{dg^*} \lambda_d < 0$, **then**
6: $\quad\quad \hat{\mathcal{G}} \leftarrow \hat{\mathcal{G}} \cup \{g^*\}$
7: $\quad$ **end if**
8: **until** $\Gamma_{g^*} - \sum_{d \in \mathcal{D}} G_{dg^*} \lambda_d \geq 0$
9: **if** γ is not an integer, **then**
10: $\quad \gamma \leftarrow$ Solve the MWSP problem ((2)–(4)) over $\hat{\mathcal{G}}$ instead of $\mathcal{G}$.
11: **end if**
12: Return γ.

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 heuristic 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 5.

3. 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. More specifically, we define the sets $\mathcal{D}$ and $\mathcal{G}$ of observations and feasible hypotheses and how the costs Γ_g of the hypotheses $g \in \mathcal{G}$ are computed. In Sections 3.1–3.3, we consider MWSP formulations of multiperson tracking, multiperson pose estimation, and multicell segmentation, respectively. These applications are visualized in Figure 1. In Section 3.4, we describe how MWSP complements supervised learning.

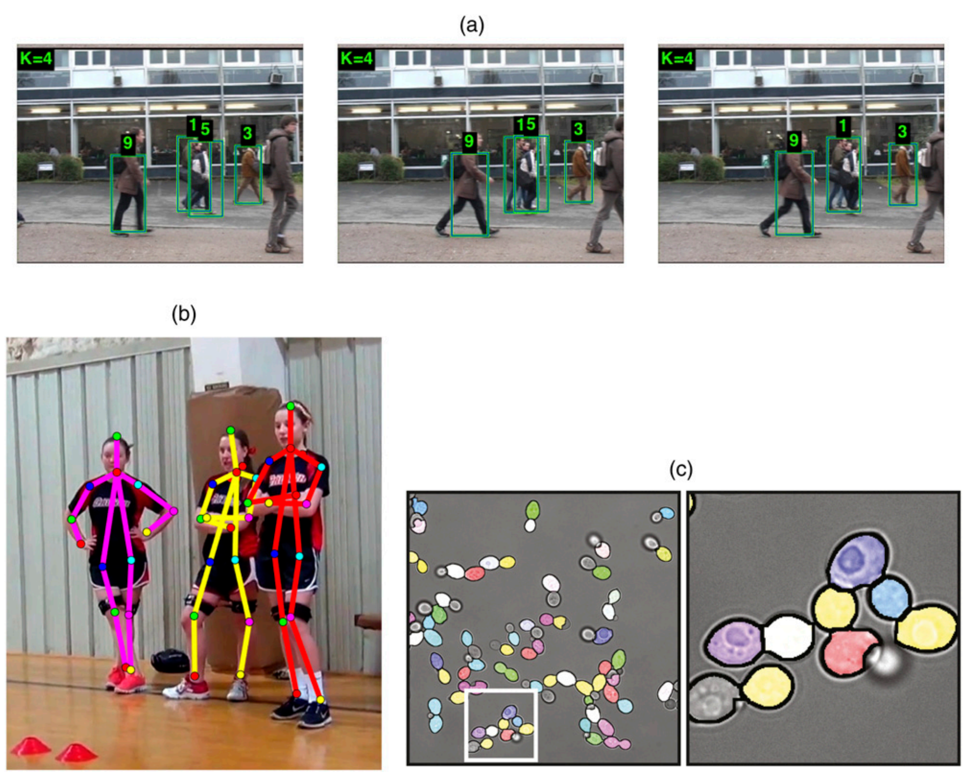
3.1. Multiperson Tracking

Multiperson tracking is the task of identifying and tracking each unique person in a video. This task is motivated by security applications and autonomous vehicle applications. In multiperson tracking, the specific identities of the people in the image, as well as the number of people present, are unspecified. Combinatorial optimization has been applied to multiperson 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 multiperson tracking using MWSP and describe the corresponding workflow as follows:

1. 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. 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

Figure 1. (Color online) (a) Multiperson Tracking; (b) Multiperson Pose Estimation; (c) Multicell Segmentation



Sources. (a) Wang et al. (2017b); (b) created using the code developed by Wang et al. (2018); (c) Zhang et al. (2017).
Notes. (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. (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. (c) Observations correspond to superpixels and hypotheses to complete biological cells. Cells are color-coded arbitrarily, with each cell being provided a single color.

subsets of K detections are nonsensible because 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; for example, humans cannot teleport across space within a few frames of video.

3. The packing of detections into sequences of subtracks forming complete tracks is formulated as an 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 $\mathscr{D}$ be the set of detections of people in the video frames, and let $\mathscr{S}$ be the set of subtracks. For a given subtrack $s \in \mathscr{S}$, let s_k indicate the kth detection in the sequence $s = \{s_1, \ldots, s_K\}$ 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 by $\mathscr{G}$, 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 $T \in \{0,1\}^{|\mathscr{S}| \times |\mathscr{G}|}$, 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 $\theta \in \mathbb{R}^{|\mathscr{S}|}$, 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. A Bayesian prior belief on the number of people (tracks) in an image is modeled 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 $g \in \mathscr{G}$, denoted by Γ_g, is defined as

$$\Gamma_g = \theta^0 + \sum_{s \in \mathscr{S}} T_{sg}\theta_s. \tag{8}$$

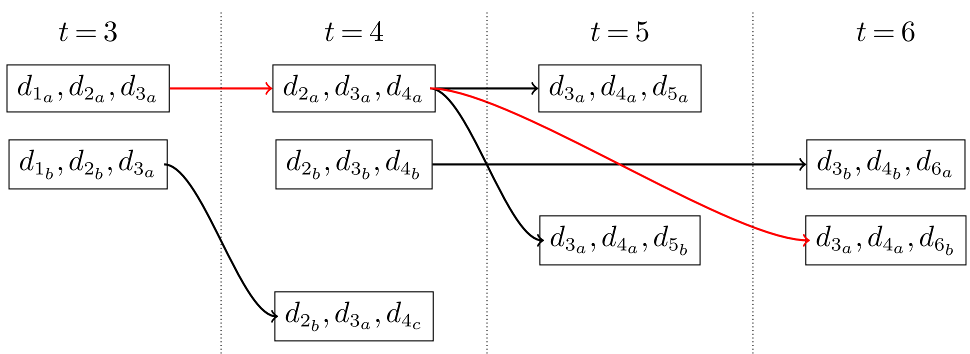
To permit the construction of tracks that have fewer detections than K, the set of subtracks is augmented with subtracks padded with empty detections. Such subtracks have no possible predecessors or successors.

3.2. Multiperson Pose Estimation

Multiperson 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. Multiperson pose estimation is relevant in multiple domains, including, but not limited to, personal robot assistants, rehabilitation, and security. Wang et al. (2018) consider the following workflow of the MWSP formulation of multiperson pose estimation, which is built off the seminal work of Pishchulin et al. (2016) and Insafutdinov et al. (2016).

1. An ANN identifies all instances of each of 14 human body parts (head, neck, and left/right of the following: shoulder, elbow, wrist, hip, knee, and ankle). Some of these detections are false detections. Some sets of detections correspond to the same body part but are separated in pixel space.

Figure 2. (Color online) Possible Tracks and Subtracks (Boxes), Where Directed Arrows Indicate the Valid Successors of a Given Subtrack



Notes. 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). Observe that the top square at time 3, followed by the top square at time 4, followed by the bottom square at time 6 corresponds to a single track containing detections $d_{1_a}, d_{2_a}, d_{3_a}, d_{4_a}, d_{6_b}$.

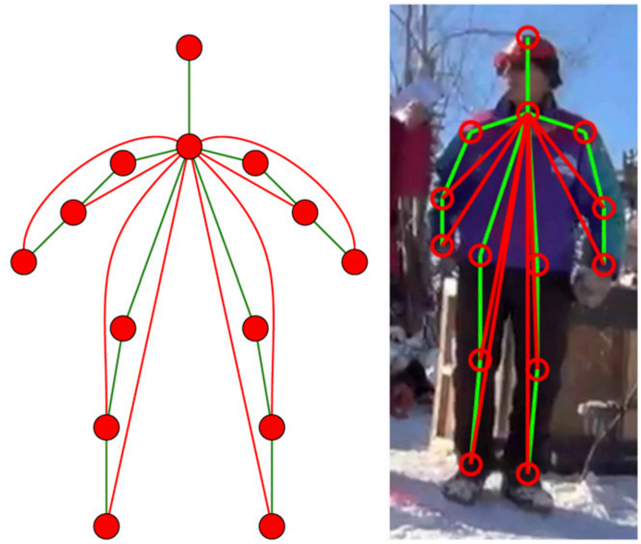
2. A classifier (such as an ANN) computes for each pair of detections the cost incurred if they are associated with the same person. This cost is derived from the probability that the two detections belong to a common person. Similarly, a cost associating each detection with a person is computed. These 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. The cost terms over pairs of detections are referred to 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 nonzero only between detections corresponding to the same body part or to adjacent body parts in the tree model. Such a model, augmented with additional connections, is shown in Figure 3. The use of connections outside the tree structure increases modeling power and computational performance (as reported in Section 7.2) in the next step of this pipeline.

3. The body part detections (observations) are aggregated to form people (hypotheses) using a MWSP formulation.

The MWSP formulation of multiperson 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, because 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 nonzero 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, 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 multiperson tracking, a constant cost θ^0 is associated with instancing a person for which a positive/negative value

Figure 3. (Color online) (Left) Augmented Tree Model of a Person as a Stick Figure; (Right) Augmented Tree Model Superimposed on an Image of a Person



Source. Wang et al. (2018).

Notes. Each circular node represents a body part, light gray (green online) edges indicate connections in traditional pictorial structure, and dark gray (red online) edges indicate augmented connections.

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

$$\Gamma_g = \theta^0 + \sum_{d \in \mathcal{D}} \theta^1_d G_{dg} + \sum_{\substack{d_1 \in \mathcal{D} \\ d_2 \in \mathcal{D}}} \theta^2_{d_1 d_2} G_{d_1 g} G_{d_2 g}. \tag{9}$$

3.3. Multicell Segmentation

Multicell 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 an image is unspecified. Multicell 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. Multicell segmentation can be cast as a correlation clustering problem (Zhang et al. 2014b) and formulated as a MWSP problem (Zhang et al. 2017). The pipeline used by Zhang et al. (2017) is described as follows:

1. Given a biological image, dimensionality reduction is applied by partitioning a set of pixels into subsets called *superpixels* (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 are in the background. This classifier uses local spatial and color statistics. This conversion reduces the space of millions of pixels to thousands of superpixels but rarely meaningfully compromises the boundaries of any cell in the ground truth.
2. For each pair of adjacent superpixels, 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 superpixel to be part of a cell. As earlier, these costs are referred to as *pairwise* and *unary*, respectively.
3. The maximum radius and area (volume in three-dimensional (3D) images) of cells on the annotated data are computed.
4. Identifying each cell in the image is formulated as a MWSP problem, where observations are superpixels and hypotheses are cells.

In their MWSP formulation of multicell segmentation, Zhang et al. (2017) denote by $\mathcal{D}$ the set of superpixels 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 superpixel d^*, which is referred to as an *anchor*, such that all superpixels 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 superpixels d_1 and d_2. We use $[\ldots]$ 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

$$\exists d^* \in \mathcal{D} \quad \text{such that} \quad \left[G_{dg} = 1\right] \Rightarrow [S_{d^* d} \le R_{\max}], \quad \forall d \in \mathcal{D}. \tag{10}$$

Optionally, the anchor can be required to be present in the cell.

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 superpixel $d \in \mathcal{D}$. A cell $g \in \mathcal{G}$ satisfies the area constraint if

$$\sum_{d \in \mathcal{D}} A_d G_{dg} \le A_{\max}. \tag{11}$$

The cost of a cell (reflecting the within-image evidence of its quality) is again defined using unary and pairwise cost terms $\theta^1 \in \mathbb{R}^{|\mathcal{D}|}$ and $\theta^2 \in \mathbb{R}^{|\mathcal{D}| \times |\mathcal{D}|}$. For each superpixel $d \in \mathcal{D}$, θ^1_d denotes the cost for superpixel d to be part of any cell. Similarly, for each superpixel 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 preceding 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 superpixel 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 preceding 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

$$\Gamma_g = \theta^0 + \sum_{d \in \mathcal{D}} \theta^1_d G_{dg} + \sum_{\substack{d_1 \in \mathcal{D} \\ d_2 \in \mathcal{D}}} \theta^2_{d_1 d_2} G_{d_1 g} G_{d_2 g}. \tag{12}$$

Otherwise, it is assumed that $\Gamma_g = \infty$.

3.4. 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 regard to transitivity, and (2) classification may produce predictions that are inconsistent with regard to domain knowledge, as illustrated in the following example.

Consider 10 observations and that all hypotheses consist of exactly 9 observations. Assume that each observation and each pair of observations are equally likely to be part of the single hypothesis that 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 9 observations, then the maximum-likelihood solution predicts a single hypothesis consisting of all 10 observations. Below we list some real-world cases where domain knowledge provides structural requirements to hypotheses.

- *Multiperson 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.

4. Pricing Problems for the Applications Considered

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

4.1. Pricing for Multiperson Tracking

Wang et al. (2017b) formulate the task of identifying the least-reduced-cost track 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 $\hat{s}$. 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

$$\ell_s = \theta_s - \lambda_{s_K} + \min\left\{\min_{\hat{s}\in\Pi_s} \ell_{\hat{s}}, \theta^0 - \sum_{k=1}^{K-1} \lambda_{s_k}\right\}. \tag{13}$$

One 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 be easily implemented because 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 nonnegative reduced cost).

4.2. Pricing for Multiperson Pose Estimation

As proposed by Wang et al. (2018), identifying the least-reduced-cost person can be formulated 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.

The pricing algorithm iterates through the power set of neck detections and computes the least-reduced-cost person containing exactly those neck detections. Let $\breve{\mathcal{D}}$ be a member of the power set of neck detections. We use notation $[g \leftrightarrow \breve{\mathcal{D}}] = 1$ 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

$$\min_{\substack{g\in\mathcal{G} \\ [g\leftrightarrow\breve{\mathcal{D}}]=1}} \Gamma_g - \sum_{d\in\mathcal{D}} G_{dg}\lambda_d. \tag{14}$$

To solve model (14) as a dynamic program, assume that the power set of detections corresponding to pairs of adjacent parts in the augmented tree can be enumerated.

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 those detections. Set $\mathcal{S}^r$ is described 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 zero and, thus, denote the power set of neck detections by $\mathcal{S}^0$.

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. This tree is made 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. Denote by $\mu^r_{\hat{s}}$ 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

$$\mu^r_{\hat{s}} = \min_{s \in \mathcal{S}^r} \quad 2 \sum_{\substack{\hat{d} \in \mathcal{D}^{\hat{r}} \\ d \in \mathcal{D}^r}} S^{\hat{r}}_{\hat{d}\hat{s}} S^r_{ds} \theta^2_{\hat{d}d} + \nu^r_s, \tag{15}$$

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

$$\nu^r_s = \sum_{d \in \mathcal{D}^r} (\theta^1_d - \lambda_d) S^r_{ds} + \sum_{\substack{d_1 \in \mathcal{D}^r \\ d_2 \in \mathcal{D}^r}} \theta^2_{d_1 d_2} S^r_{d_1 s} S^r_{d_2 s} + 2 \sum_{\substack{d_1 \in \mathcal{D}^0 \\ d_2 \in \mathcal{D}^r}} \theta^2_{d_1 d_2} S^0_{d_1 s_0} S^r_{d_2 s} + \sum_{\bar{r} \in \Upsilon^r} \mu^{\bar{r}}_s. \tag{16}$$

Observe that (15) and (16) describe a dynamic program.

To compute $\mu^r_{\hat{s}}$ for each $\hat{s} \in \mathcal{S}^{\hat{r}}$, one needs 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 1 billion configurations would have to be enumerated, which is prohibitively expensive. This has motivated the use of nested Benders decomposition 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.

4.3. Pricing for Multicell Segmentation

To find negative-reduced-cost cells, Zhang et al. (2017) 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 superpixels that compose the cell. Let $\mathcal{D}_{d^*} \subseteq \mathcal{D}$ be the subset of detections that may be in a cell with anchor $d^* \in \mathcal{D}$; that is,

$$\mathcal{D}_{d^*} = \{d \in \mathcal{D} \mid S_{d^*d} \leq R_{\max}\}. \tag{17}$$

Pricing is attacked 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,

$$g_{d^*} \in \underset{\substack{g \in \mathcal{G} \\ G_{dg}=0, \forall d \notin \mathcal{D}_{d^*}}}{\arg\min} \; \theta^0 + \sum_{d \in \mathcal{D}} (\theta^1_d - \lambda_d) G_{dg} + \sum_{d_1, d_2 \in \mathcal{D}} \theta^2_{d_1 d_2} G_{d_1 g} G_{d_2 g}. \tag{18}$$

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

$$\min_{\substack{x \in \{0,1\}^{|\mathcal{D}|} \\ y \geq 0}} \theta^0 + \sum_{d \in \mathcal{D}} (\theta^1_d - \lambda_d) x_d + \sum_{d_1, d_2 \in \mathcal{D}} \theta^2_{d_1 d_2} y_{d_1 d_2} \tag{19}$$

$$\text{s.t.} \quad y_{d_1 d_2} \leq x_{d_1}, \quad \forall d_1, d_2 \in \mathcal{D}, \tag{20}$$

$$y_{d_1 d_2} \leq x_{d_2}, \quad \forall d_1, d_2 \in \mathcal{D}, \tag{21}$$

$$-y_{d_1 d_2} + x_{d_1} + x_{d_2} \leq 1, \quad \forall d_1, d_2 \in \mathcal{D}, \tag{22}$$

$$\sum_{d \in \mathcal{D}} x_d \geq 1, \tag{23}$$

$$\sum_{d \in \mathcal{D}} A_d x_d \leq A_{\max}, \tag{24}$$

$$x_d = 0, \quad \forall d \in \mathcal{D} \setminus \mathcal{D}_{d^*}. \tag{25}$$

The binary requirements on y are enforced via (20)–(22). Indeed, constraints (20) and (21) 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 (22) 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 (23) imposes that at least one superpixel be included in the cell. Constraints (24) enforce that the area of the cell does not exceed the maximum area. Constraints (25) ensure that all detections not respecting the maximum radius constraint from d^* are not selected in the cell. Recall from Section 3.3 that the anchor may be required to be included in the cell. This condition is imposed by setting x_{d^*} to one.

Because the ILP (18) 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 $\hat{\mathcal{G}}$.

5. Subset-Row Inequalities

In this section, we discuss how the LP relaxation of the MWSP formulation ((2)–(4)) can be tightened by adding subset-row inequalities. To motivate this, we provide a case from Wang et al. (2017b) where the LP relaxation is loose. Consider four hypotheses $\mathcal{G} = \{g_1, g_2, g_3, g_4\}$ over three observations $\mathcal{D} = \{d_1, d_2, d_3\}$, where the first three hypotheses each contain two of the three observations $\{d_1, d_2\}$, $\{d_1, d_3\}$, and $\{d_2, d_3\}$, 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 (2)–(4) is loose in this case.

The LP relaxation of MWSP (2)–(4) or, equivalently, the MP in the CG algorithm can be tightened by employing the subset-row inequalities introduced by Jepsen et al. (2008) and also used by (Wang et al. 2017b) and Zhang et al. (2017) for computer vision applications. These inequalities are Chvátal–Gomory inequalities of rank one and 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 $m_1 m_2 - 1$. Applying the Chvátal–Gomory procedure using the multiplier $1/m_1$ for each constraint (3) associated with an observation $d \in \hat{\mathcal{D}}$ results in the following subset-row inequality:

$$\sum_{g \in \mathcal{G}} \gamma_g \left\lfloor \frac{\sum_{d \in \hat{\mathcal{D}}} G_{dg}}{m_1} \right\rfloor \leq m_2 - 1. \tag{26}$$

Observe that this inequality imposes, in particular, that the number of selected hypotheses containing m_1 or more members of $\hat{\mathcal{D}}$ must not exceed $m_2 - 1$.

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. Wang et al. (2017b) propose an elegant way to deal with the multiperson tracking problem that relies on the original structure of the pricing problem.

This section focuses on the subset-row inequalities (SRIs) 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, which 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 5.1, we present a MWSP formulation tightened using 3-SRIs and discuss how the CG algorithm is modified to account for them. In Section 5.2, we discuss the application of the subset-row inequalities to the multicell segmentation problem, which preserves the structure of the pricing problem. Finally, in Section 5.3, 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 multiperson tracking and multiperson pose estimation applications.

5.1. 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. The mapping of 3-SRIs to hypotheses is described 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 $g \in \mathcal{G}$; that is, $C_{cg} = 1$ if and only if hypothesis g contains at least two observations in c. With this notation, the MWSP MP tightened using 3-SRIs can be written as

$$\min_{\gamma \geq 0} \quad \sum_{g \in \mathcal{G}} \Gamma_g \gamma_g \tag{27}$$

$$\text{s.t.} \quad \sum_{g \in \mathcal{G}} G_{dg} \gamma_g \leq 1, \quad \forall d \in \mathcal{D}, \tag{28}$$

$$\sum_{g \in \mathcal{G}} C_{cg} \gamma_g \leq 1, \quad \forall c \in \mathcal{C}. \tag{29}$$

Given the relatively large size of $\mathscr{C}$, the 3-SRIs are generated only as needed; that is, the MP is solved using a column/row-generation (CRG) algorithm. It starts with an empty subset $\hat{\mathscr{C}}$ of 3-SRIs and solves the MP ((27) and (28)) by CG, generating a subset $\hat{\mathscr{G}}$ of columns. If the computed MP solution is not integer, the CRG algorithm iterates over all sets $c \in \mathscr{C}$ to identify the set c^* that maximizes $\sum_{g\in\mathscr{G}} \gamma_g C_{cg}$, given the current fractional solution γ. If $\sum_{g\in\mathscr{G}} \gamma_g C_{c^*g} > 1$, then the corresponding 3-SRI is violated, and c^* is added to set $\hat{\mathscr{C}}$. CG is then restarted to solve the MP ((27)–(29)), where the constraint set (29) is restricted to the subset $\hat{\mathscr{C}}$ of the generated 3-SRIs. Once solved, the search for a violated 3-SRI is performed again, and if one is found, it is added to $\hat{\mathscr{C}}$. 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, the search for a violated 3-SRI can be limited to sets $c \in \mathscr{C}$, for which each detection $d \in \mathscr{D}_c$ is included in a hypothesis g associated with a fractional-valued variable γ_g (Wang et al. 2017b).

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

$$\min_{g\in\mathscr{G}} \quad \Gamma_g - \sum_{d\in\mathscr{D}} G_{dg}\lambda_d - \sum_{c\in\hat{\mathscr{C}}} C_{cg}\psi_c. \tag{30}$$

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

5.2. 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 multicell segmentation, where the pricing problem remains an ILP. Let $z_c \in \{0,1\}$, $c \in \hat{\mathscr{C}}$, be a binary variable that is equal to one if and only if two or more detections in $\mathscr{D}_c$ are included in the cell. The ILP is given by

$$\begin{aligned}
\min_{\substack{x\ge0\\y\ge0\\z\ge0}} \quad & \theta^0 + \sum_{d\in\mathscr{D}}(\theta^1_d - \lambda_d)x_d + \sum_{d_1,d_2\in\mathscr{D}} \theta^2_{d_1d_2} y_{d_1d_2} - \sum_{c\in\hat{\mathscr{C}}} \psi_c z_z && (31)\\
\text{s.t.} \quad & y_{d_1d_2} \le x_{d_1}, \quad \forall d_1, d_2 \in \mathscr{D}, && (32)\\
& y_{d_1d_2} \le x_{d_2}, \quad \forall d_1, d_2 \in \mathscr{D}, && (33)\\
& x_{d_1} + x_{d_2} - y_{d_1d_2} \le 1, \quad \forall d_1, d_2 \in \mathscr{D}, && (34)\\
& x_d \in \{0,1\}, \quad \forall d \in \mathscr{D}, && (35)\\
& \sum_{d\in\mathscr{D}} x_d \ge 1, && (36)\\
& \sum_{d\in\mathscr{D}} A_d x_d \le A_{\max}, && (37)\\
& x_d = 0, \quad \forall d \in \mathscr{D} \setminus \mathscr{D}_{d^*}, && (38)\\
& x_{d_3} + x_{d_4} - z_c \le 1, \quad \forall c \in \hat{\mathscr{C}}, d_3, d_4 \in \mathscr{D}_c \mid d_3 \ne d_4. && (39)
\end{aligned}$$

This pricing model is identical to model (19)–(25), 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 (39) are required to set the values of the z variables. Observe that for every $c \in \hat{\mathscr{C}}$, z_c is set to its smallest possible value at optimality because ψ_c is nonpositive. Thus, z_c is not explicitly required to be integer because its integrality is assured when x is integral. Finally, note that all constraints (39) for which $\psi_c = 0$ can be ignored.

5.3. 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 multiperson tracking and multiperson pose estimation that rely on dynamic programming as a specialized pricing solver. Given that $\psi \le 0$, 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 a slightly improved version of this algorithm in which 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 5.3.1 and 5.3.2, we specify the bounding and branching operations, respectively.

5.3.1. Bounding. For every node $b \in \mathcal{B}$, let $V_b^* = \min_{g \in \mathcal{G}_b^{\pm}} >(\Gamma_g - \sum_{d \in \mathcal{D}} \lambda_d G_{dg} - \sum_{c \in \hat{\mathcal{C}}} \psi_c C_{cg})$ 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^* that exploits the knowledge stored in the subset $\mathcal{D}_b^+$ of the observations that must be part of any feasible hypothesis at node b. Recall that D_c denotes the subset of observations involved in SRI $c \in \mathcal{C}$.

Proposition 1. (From Wang et al. 2017b). *Let $b \in \mathcal{B}$ be a branching node. Then*

$$V_b^{lb} = \underline{V}_b^* - \sum_{c \in \hat{\mathcal{C}}} \psi_c \left\lfloor \frac{|\mathcal{D}_c \cap \mathcal{D}_b^+|}{2} \right\rfloor \tag{40}$$

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

For multiperson tracking and multiperson pose estimation, the value of $\underline{V}_b^*$ can be computed by solving the dynamic program presented in Sections 4.1 and 4.2, respectively. This program is, however, modified to enforce that $g \in \mathcal{G}_b^{\pm}$.

5.3.1.1. Multiperson Tracking. To discard the detections in $\mathcal{D}_b^-$, all subtracks that include a detection in D_b^- are removed from $\mathcal{S}$. To impose that all detections in $\mathcal{D}_b^+$ be part of the track, all subtracks that include a detection $d' \notin \mathcal{D}_b^+$ or an occlusion co-occurring in time with any $d \in \mathcal{D}_b^+$ are first removed. Then, starting a track after the occurrence of the first member of $\mathcal{D}_b^+$ in time is not allowed. Finally, after completing the dynamic program algorithm, any generated track that terminates prior to the point in time of the last member of $\mathcal{D}_b^+$ is omitted. Note that Wang et al. (2017b) did not require the hypotheses to be consistent with the decisions in $\mathcal{D}_b^+$, as we propose here.

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

5.3.2. 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 be easily 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); that is,

$$V_b^{ub} = \Gamma_{g_b} - \sum_{d \in \mathcal{D}} \lambda_d G_{dg_b} - \sum_{c \in \hat{\mathcal{C}}} \psi_c C_{cg_b}. \tag{41}$$

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 $\hat{\mathcal{C}}_b = \{c \in \hat{\mathcal{C}} >| >C_{cg_b} - \lfloor \frac{|\mathcal{D}_c \cap \mathcal{D}_b^+|}{2} \rfloor = 1\}$ 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}, branching is performed 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 because 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 the $\mathcal{D}^+$ 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

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

Note. 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 ($D^+_{b_8} = D^+_b \cup \{d_3\}$).

6. Dual-Optimal Inequalities

The MWSP formulation is known to be subject to high degeneracy, especially when there is a large number of set packing constraints (3) and the generated hypotheses contain, on average, a relatively large number of detections (say 10 or more). Degeneracy induces multiple dual solutions for each primal solution and slow convergence of the CG algorithm because the dual variables may strongly oscillate from one CG iteration to the next, yielding the generation of many useless columns. One way to reduce degeneracy and stabilize the dual variables is to incorporate DOIs in the formulation (Valério de Carvalho 2005, Ben Amor et al. 2006). A DOI is an inequality defined on the dual variables that cuts away a part of the dual space while preserving at least one optimal dual solution. It is included in the formulation by adding an additional variable in the MP with appropriate coefficients. DOIs have been successfully applied to various problems, including stock cutting and packing (Valério de Carvalho 2005, Ben Amor et al. 2006, Gschwind and Irnich 2016), image segmentation (Yarkony et al. 2012, Yarkony and Fowlkes 2015), and multiperson pose estimation (Wang et al. 2018).

In this section, we present for the general MWSP formulation DOIs that do not vary with $\hat{\mathcal{G}}$, the current set of columns in the RMP (Section 6.1), and DOIs that vary as $\hat{\mathcal{G}}$ changes (Section 6.2). The former were introduced by Wang et al. (2018), and the latter are new. In both cases, each DOI is defined as a lower bound on a dual variable. Finally, in Section 6.3, we describe how these DOIs can be defined for two of the considered applications.

6.1. Basic Dual-Optimal Inequalities

One way to introduce DOIs in the MWSP formulation is to allow the inclusion of an observation in multiple selected hypotheses at a sufficiently large cost. To do so, a surplus variable $\xi_d \geq 0$ with a cost $\Xi_d \geq 0$ is added to each constraint (3), resulting in the following modified MP:

$$\min_{\substack{\gamma \geq 0 \\ \xi \geq 0}} \quad \sum_{g \in \mathcal{G}} \Gamma_g \gamma_g + \sum_{d \in \mathcal{D}} \Xi_d \xi_d \tag{42}$$

$$\text{s.t.} \quad \sum_{g \in \mathcal{G}} G_{dg} \gamma_g - \xi_d \leq 1, \quad \forall d \in \mathcal{D}. \tag{43}$$

Each surplus variable ξ_d, $d \in \mathcal{D}$, induces a lower bound $-\Xi_d$ on the dual variable λ_d, which writes as the DOI $\lambda_d \geq -\Xi_d$ in the dual of the MP (recall that $\lambda_d \leq 0$).

Obviously, if the costs Ξ_d are very large, then the modified MP ((42) and (43)) is equivalent to the original one ((2) and (3)). However, in this case, the DOIs do not reduce the dual space much and have no impact on the CG process. To devise useful DOIs, one must set the values of Ξ as small as possible but ensure that the equivalence between (42) and (43) and (2) and (3) is preserved.

With this in mind, Wang et al. (2018) propose to compute the values of Ξ as follows. For each observation $d \in \mathcal{D}$, let $\mathcal{G}_d = \{g \in \mathcal{G} >| > G_{dg} = 1\}$ be the subset of hypotheses including d. Then, for each hypothesis $g \in \mathcal{G}_d$, let Δ_{gd} be the cost of a least-cost (possibly empty) subset of nonoverlapping hypotheses that can feasibly replace hypothesis g in a solution but without including observation d. This cost can be computed by solving the following ILP:

$$\Delta_{gd} \quad = \quad \min_{\gamma \geq 0} \quad \sum_{h \in \mathcal{G} \setminus \mathcal{G}_d} \Gamma_h \gamma_h \tag{44}$$

$$\text{s.t.} \quad \sum_{h \in \mathcal{G} \setminus \mathcal{G}_d} G_{\bar{d}h} \gamma_h \leq G_{\bar{d}g}, \quad \forall \bar{d} \in \mathcal{D} \setminus \{d\}, \tag{45}$$

$$\gamma_h \in \{0, 1\}, \quad \forall h \in \mathcal{G} \setminus \mathcal{G}_d, \tag{46}$$

where the objective function (44) aims at minimizing the cost of the selected hypotheses and constraints (45) require that these hypotheses do not overlap or include any observation not included in hypothesis g. The set of the selected hypotheses is denoted by $\mathcal{H}_{gd}$. Thus, the best possible option to replace hypothesis $g \in \mathcal{G}_d$ in a solution is to use the hypotheses in $\mathcal{H}_{gd}$. This yields a cost difference of $\Delta_{gd} - \Gamma_g$. Consequently, setting $\Xi_d \geq \epsilon + \max\{0, \Delta_{gd} - \Gamma_g\}$, where ϵ is a tiny positive constant, ensures that hypothesis g cannot be part of a solution in which observation d is included in more than one hypothesis. Therefore, to exclude this possibility for all hypotheses $g \in \mathcal{G}_d$, cost Ξ_d must satisfy

$$\Xi_d \geq \epsilon + \max_{g \in \mathcal{G}_d} \max\{0, \Delta_{gd} - \Gamma_g\}. \tag{47}$$

Note that computing the value of the right-hand side of (47) may be computationally expensive, and thus, an upper bound on its value is rather computed, as discussed in Section 6.3.

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

Proposition 2. *Let ζ^* and ζ^*_{DOI} be the optimal value of the MP* (2) *and* (3) *and the modified MP* (42) *and* (43), *respectively. If Ξ satisfies* (47), *then $\zeta^*_{DOI} = \zeta^*$.*

Proof. Let (γ^*, ξ^*) be an optimal solution to (42) and (43). If $\xi^*_d = 0$ for all observations $d \in \mathcal{D}$, then $\zeta^*_{DOI} = \zeta^*$ because γ^* is feasible and optimal for (2) and (3). 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 (47) 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 (42) and (43) 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 ((42) and (43)) by CG is bounded.

6.2. Dual-Optimal Inequalities That Vary with $\hat{\mathcal{G}}$

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

Let $\hat{\mathcal{G}}^* = \{g \in \mathcal{G}\ >|\ > \exists h \in \hat{\mathcal{G}}$ 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 $h \in \hat{\mathcal{G}}$. 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

$$\Xi_d \geq \epsilon + \max_{g \in \hat{\mathcal{G}}^*_d} \max\{0, \Delta_{gd} - \Gamma_g\}, \tag{48}$$

where ϵ and Δ_{gd} are defined as in the preceding 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 (48) may increase when hypotheses are added to $\hat{\mathcal{G}}$, but it is never greater than that of (47). As for (47), we do not necessarily compute the right-hand side of (48) but rather an upper bound on its value (see Section 6.3).

With these DOIs, which we call the *varying DOIs*, the modified RMP can be written as

$$\min_{\substack{\gamma \geq 0 \\ \xi \geq 0}} \quad \sum_{g \in \hat{\mathcal{G}}} \Gamma_g \gamma_g + \sum_{d \in \mathcal{D}} \Xi_d \xi_d \tag{49}$$

$$\text{s.t.} \quad \sum_{g \in \hat{\mathcal{G}}} G_{dg} \gamma_g - \xi_d \leq 1, \quad \forall d \in \mathcal{D}. \tag{50}$$

It is identical to the one obtained using invariant DOIs, except that the values of Ξ need to be recomputed at each CG iteration because they depend on $\mathcal{G}$. The next proposition proves that it is always bounded. Its proof relies on the following lemma.

Lemma 1. *If Ξ satisfies* (48) *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 $\sum_{h \in \mathcal{H}_{g^*d^*}} (\Gamma_h + \sum_{d \in \mathcal{D}} \Xi_d G_{dh}) \geq 0$.*

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 $\sum_{h \in \mathcal{H}_{g\bar{d}}} (\Gamma_h + \sum_{d \in \mathcal{D}} \Xi_d G_{dh}) \geq 0$, then $g^* = g$ and $d^* = \bar{d}$. Otherwise, for any observation $\bar{d}$ included in g, $\mathcal{H}_{g\bar{d}} \neq \emptyset$ and $\sum_{h \in \mathcal{H}_{g\bar{d}}} (\Gamma_h + \sum_{d \in \mathcal{D}} \Xi_d G_{dh}) < 0$. In this case, there exists a hypothesis $\bar{h} \in \mathcal{H}_{g\bar{d}}$ (which contains fewer 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 this 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, $\sum_{h \in \mathcal{H}_{g\bar{d}}} (\Gamma_h + \sum_{d \in \mathcal{D}} \Xi_d G_{dh}) = 0$. □

Proposition 3. *If Ξ satisfies (48), then the modified RMP (49) and (50) is bounded.*

Proof. Assume that (49) and (50) are 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 $\sum_{h\in\mathcal{H}_{g^*d^*}} (\Gamma_h + \sum_{d\in\mathcal{D}} \Xi_d G_{dh}) \geq 0$. These inequalities imply that

$$\sum_{h\in\mathcal{H}_{g^*d^*}} \left(\Gamma_h + \sum_{d\in\mathcal{D}} \Xi_d G_{dh}\right) - \left(\Gamma_{g^*} + \sum_{d\in\mathcal{D}} \Xi_d G_{dg^*}\right) > 0, \tag{51}$$

which can be rewritten as

$$\sum_{h\in\mathcal{H}_{g^*d^*}} \Gamma_h - \Gamma_{g^*} > \sum_{d\in\mathcal{D}_{g^*d^*}} \Xi_d, \tag{52}$$

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 (52) implies that

$$\sum_{h\in\mathcal{H}_{g^*d^*}} \Gamma_h - \Gamma_{g^*} > \epsilon + \sum_{h\in\mathcal{H}_{g^*d^*}} \Gamma_h - \Gamma_{g^*} \tag{53}$$

or, equivalently, $\epsilon < 0$. This contradicts the definition of ϵ and proves that the model (49)–(50) 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 (2) and (3) and the modified RMP (49) and (50), respectively. If Ξ satisfies (48), then $\hat{\zeta}^*_{DOI} = \zeta^*$ or there exists a (nongenerated) 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 (49) and (50), respectively. The proof consists of showing that if the reduced cost of each hypothesis $g \in \mathcal{G}$ is nonnegative, that is, $\Gamma_g - \sum_{d\in\mathcal{D}} G_{dg}\hat{\lambda}^*_d \geq 0$, 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 (2) and (3). 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 zero; that is, $\hat{\lambda}^*_d + \Xi_d = 0$ 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

$$\sum_{h\in\mathcal{H}_{gd}} \left(\Gamma_h - \sum_{e\in\mathcal{D}} G_{eh}\hat{\lambda}^*_e\right) - \left(\Gamma_g - \sum_{e\in\mathcal{D}} G_{eg}\hat{\lambda}^*_e\right) \geq 0 \tag{54}$$

$$\Rightarrow \quad \sum_{h\in\mathcal{H}_{gd}} \Gamma_h - \Gamma_g \geq -\sum_{e\in\mathcal{D}_{gd}} \hat{\lambda}^*_e \geq -\hat{\lambda}^*_d \tag{55}$$

$$\Rightarrow \quad \Delta_{gd} - \Gamma_g \geq \Xi_d. \tag{56}$$

This contradicts the definition (48) of Ξ_d, which guarantees that $\Xi_d > \Delta_{gd} - \Gamma_g$. Consequently, this case is not possible. □

6.3. Application-Specific Dual-Optimal Inequalities

As mentioned earlier, computing the right-hand side values of (47) and (48) can be computationally expensive. Nevertheless, setting Ξ to greater or equal values still yields valid DOIs. In this section, we present ways to compute such values for the multiperson pose estimation application and the multicell segmentation application when the anchor is not required to lie in the cell. In both cases (which are treated simultaneously later), removing any observation from any hypothesis always yields a feasible hypothesis. This is not the case for the multicell segmentation application when the anchor must be in the cell or for the multiperson tracking application when the set of feasible hypotheses (tracks) is restricted by the subset of subtracks considered.

For the invariant DOIs based on (47), observe first that the right-hand side of (47)

$$\epsilon + \max_{g\in\mathcal{G}_d} \max\{0, \Delta_{gd} - \Gamma_g\} \leq \epsilon + \max_{g\in\mathcal{G}_d} \max\left\{0, \Gamma_{\bar{g}(g,d)} - \Gamma_g\right\}, \tag{57}$$

where $\bar{g}(g,d)$ is the (possibly empty) hypothesis obtained by removing observation d from hypothesis g. This relationship ensues from the fact that $\Gamma_{\bar{g}(g,d)} \geq \Delta_{gd}$ because the solution obtained by setting $\gamma_{\bar{g}(g,d)} = 1$ and $\gamma_h = 0$, $\forall h \in \mathcal{G} \setminus \mathcal{G}_d$, $h \neq \bar{g}(g,d)$, is feasible for the ILP ((44)–(46)) defining Δ_{gd}.

To find an upper bound on the right-hand side of (57), 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 (9) or (12), 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

$$\Xi_d = \epsilon + \max\left\{0,\ -\left(\min\left\{0, \theta^0\right\} + \theta^1_d + \sum_{\bar{d} \in \mathcal{D}} \min\left\{0, \theta^2_{d\bar{d}} + \theta^2_{\bar{d}d}\right\}\right)\right\}. \tag{58}$$

For the varying DOIs based on (48), the values of Ξ are produced by using the same approach as in (58), 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

$$\Xi_d = \epsilon + \max\left\{0,\ -\left(\min\left\{0, \theta^0\right\} + \theta^1_d + \min_{g \in \hat{\mathcal{G}}_d} \sum_{\bar{d} \in \mathcal{D}} G_{\bar{d}g} \min\left\{0, \theta^2_{d\bar{d}} + \theta^2_{\bar{d}d}\right\}\right)\right\}. \tag{59}$$

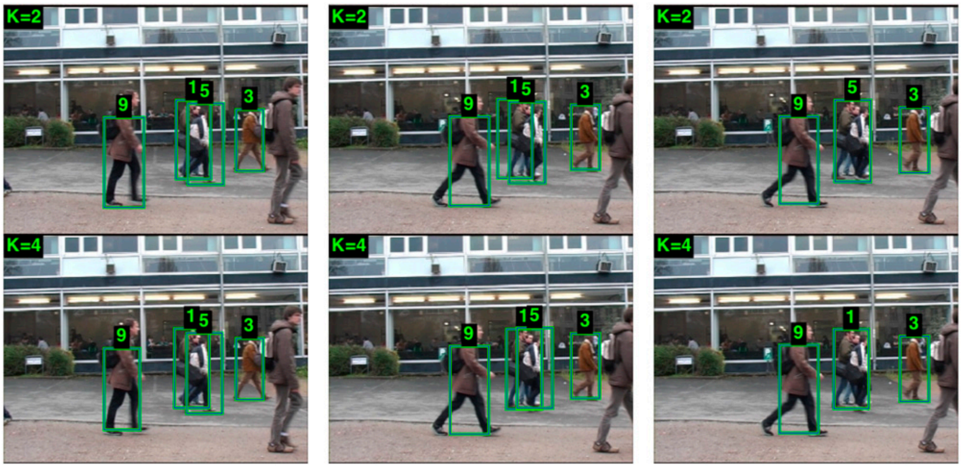
7. Computational Results

In this section, we provide computational results on the multiperson tracking, multiperson pose estimation, and multicell segmentation applications in Sections 7.1–7.3, respectively. Note that most of these results are taken from the works of Wang et al. (2017b, 2018) and Zhang et al. (2017).

7.1. Multiperson Tracking

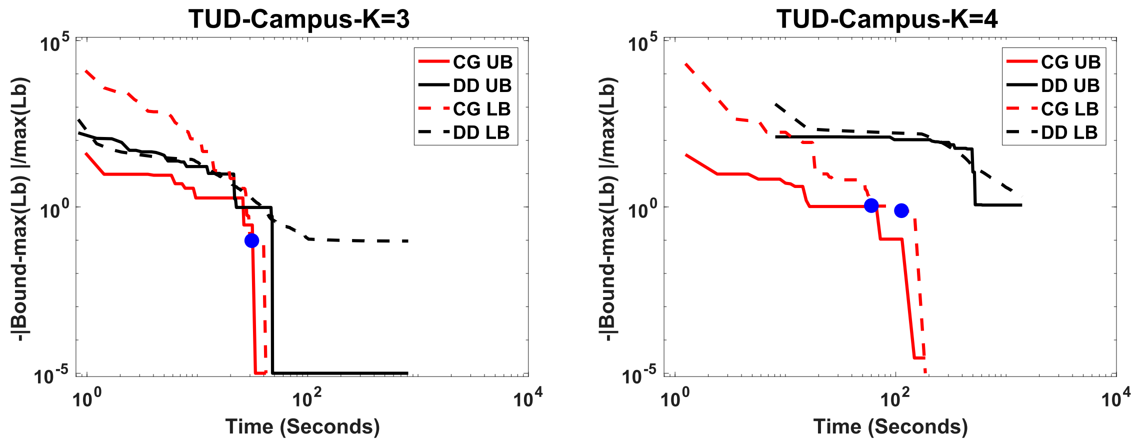
A part of the Multiple Objective Tracking (MOT) 2015 training set (Leal-Taixé et al. 2015) is used to train and evaluate multiperson tracking in video. The structured SVM-based learning approach of Wang and Fowlkes (2015) produces the cost terms, given that the raw detector outputs provided by the MOT data set generate 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.

Figure 4. (Color online) Qualitative Example of Improvement as a Result of Increasing Subtrack Length



Source. Wang et al. (2017b).
Notes. 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, whereas with $K = 4$, the identity of track 1 does not change.

Figure 5. (Color online) 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



Notes. 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 dot on the lower-bound plot.

To assess CG convergence, lower bounds are computed throughout the solution process. At a given CG iteration prior to adding any 3-SRIs, the computed lower bound is given by

$$\sum_{d\in\mathcal{D}} \lambda_d + \sum_{d\in\mathcal{D}} \min \left\{0, \min_{\substack{s\in\mathcal{S} \\ s_K=d}} \ell_s \right\}, \tag{60}$$

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 (see Wang et al. 2017b for lower-bound computation once 3-SRIs are added). This bound is valid because, in a feasible solution, at most one track can end with each detection.

In the instance used 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 shorthand 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, the timing/cost performance of the CG algorithm is compared with that of 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, whereas DD does not. Furthermore, CG terminates much more rapidly than DD.

7.2. Multiperson Pose Estimation

We present the experimental results from Wang et al. (2018) on the MPII Multi-Person Pose data set (Andriluka et al. 2014), which consists of 418 images. Initially, the cost terms and problem structure are from Insafutdinov et al. (2016). They are modified to improve modeling power and optimization speed. Notably, each selected person must contain exactly one neck detection. Thus, during pricing, one needs to iterate only over selections of the neck detections containing one detection. The cardinality of $\mathcal{S}^r$ for each body part $r \in \mathcal{R}$ other than the neck is also restricted to be no greater than 50,000. The detailed modifications are as follows:

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 $\mathcal{D}^r$, $\forall r \in \mathcal{R}$, 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 $\mathcal{S}^r$, the procedure iterates over integer k from one to $|\mathcal{D}^r|$ and, at each iteration, adds to $\mathcal{S}^r$ the group of configurations containing exactly k detections in $\mathcal{D}^r$. If adding a group would make $|\mathcal{S}^r|$ exceed 50,000, then the group is not added, and the construction of $\mathcal{S}^r$ stops.

Table 2. Average Precision (in Percent) of the CG Algorithm vs. Levinkov et al. (2017)

Solution method	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

Notes. 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. CG, column generation.

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, Wang et al. (2018) exploit the fact that each person has a single neck detection as follows:

(a) Let $\hat{\theta}^0$ be a desired value for θ^0. Because Insafutdinov et al. (2016) do not model a prior on the number of people, $\hat{\theta}^0$ is set to a single value for the entire data set.

(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; that is,

$$\Omega = -1 + \min\left\{0, \hat{\theta}^0\right\} + \sum_{d \in \mathcal{D}} \min\left\{0, \theta^1_d\right\} + \sum_{\substack{d_1 \in \mathcal{D} \\ d_2 \in \mathcal{D}}} \min\left\{0, \theta^2_{d_1 d_2}\right\}. \tag{61}$$

(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 7.2.1.

The computational results 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%.

Table 2 compares 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 the 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, and thus poses associated with

Figure 6. (Color online) Output Examples

Source. Wang et al. (2018).
Notes. For each person, the locations of the detections of each body part are averaged to produce the corresponding colored dot, denoting the part position.

INFORMS Journal on Optimization, 2020, vol. 2, no. 3, pp. 167–191, © 2020 INFORMS

Table 3. Total Computing Times in Seconds and Time Reduction Using DOIs vs. No DOI (in Percent) on Different Systems

System	No DOI	Invariant DOIs	Varying DOIs	Reduction invariant	Reduction varying
One	2,092.6	1,450.5	1,290.5	30.7%	38.3%
Two	9,821.2	937.6	860.2	90.4%	91.2%

Notes. 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 reduction (%) achieved by using invariant and varying DOIs versus not using DOIs. MP, master problem; CG, column generation; DOI, dual-optimal inequalities.

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 having only one head/neck. Some sample outputs of the CG algorithm can be visualized in Figure 6. Note that the dynamic programming pricing problems are solved using the nested Benders decomposition technique of Wang et al. (2018), which provides a speedup factor of up to 500-fold for these instances.

7.2.1. 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 a 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.

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

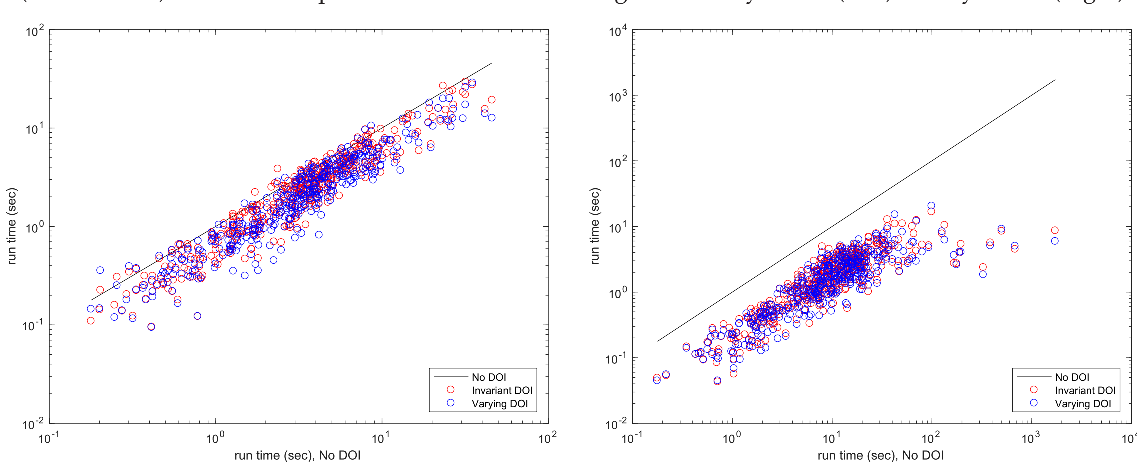
System 1: MATLAB 2016 linprog solver with default settings on a 2014 Macbook Pro.

System 2: MATLAB 2017 with the interior point solver on a powerful workstation with an Intel Core i7-6850K central processing unit (CPU) at 3.60 GHz.

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.

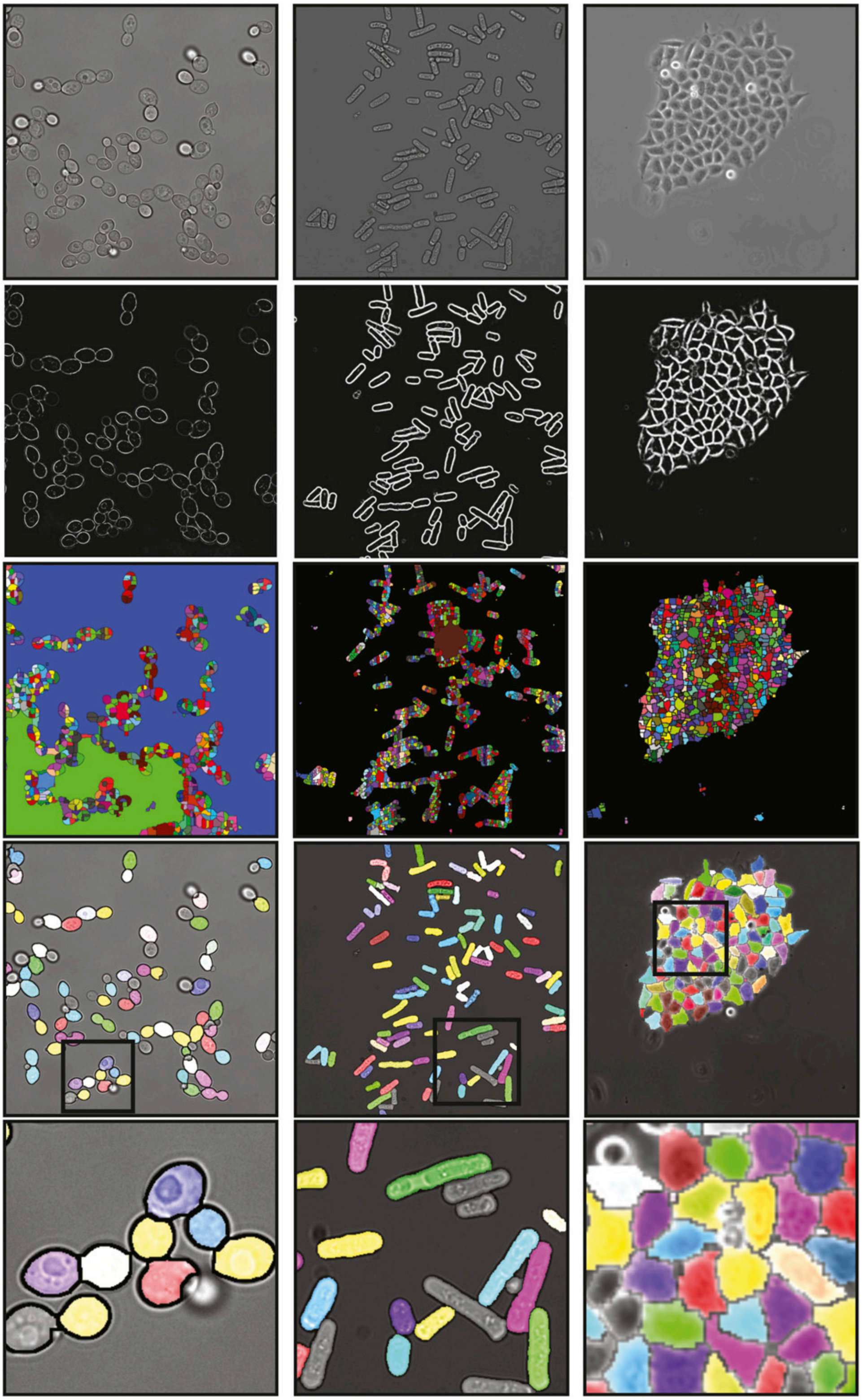
Table 3 compares the speedups obtained by the DOIs using the two systems, whereas Figure 7 provides a scatter plot of the time consumed using the DOIs for each system. Our experiments show that varying DOIs outperform invariant DOIs. The use of DOIs provides a large speedup for System 2 (over 90% CPU time reduction) but limited speedup for System 1 (~30%–38% CPU time reduction). Interestingly, compared with

Figure 7. (Color online) Relative Computational Time When Using DOIs on System 1 (Left) and System 2 (Right)



Note. Each data point specifies the computing time to solve the MP with DOIs (vertical axis) relative to the computing time when not using DOIs (horizontal axis).

Figure 8. (Color online) Cell Segmentation Examples for Data Sets 1–3 (Left to Right)



Source. Zhang et al. (2017).
Note. Rows are (top to bottom): original image, cell boundary classifier prediction image, superpixels, color map of segmentation, and enlarged views of the inset (black square).

System 2, System 1 is an older computer running an older version of MATLAB, but the running times with System 1 are better than those with System 2 when no DOIs are used. This is a consequence of System 1 producing well-centered dual solutions and System 2 not doing so. The use of DOIs makes System 2 perform better than System 1, demonstrating the value of DOIs when the system is poorly selected.

7.3. Multicell Segmentation

In this section, we report the results obtained by Zhang et al. (2017) by applying CG for multicell segmentation on three different data sets 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, the open-source toolbox of Sommer et al. (2011) is used 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, less than 1% of the pixels per data set are used with generic features (e.g., Gaussian, Laplacian, and structured tensor). The output of this random forest classifier is also used to generate superpixels.

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

7.3.1. Segmentation Quality. In Figure 8, the output obtained by the CG algorithm is displayed for three images, one in each data set. For data set 2 (the middle column), observe that the CG approach successfully segments the cells in a problem instance where there are large variations of cell shape/size.

Next, the performance of the CG algorithm is compared with that of 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. 2014a) 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 computational results over all instances of each data set 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 with other methods, including those of Arteta et al. (2012, 2016) and Funke et al. (2015).

7.3.2. 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, average results over all these instances are reported. For each instance, the relative optimality gap between the upper and lower bounds obtained at the termination of CG is computed. For the three data sets, the proportions of instances that achieve an optimality gap of 10% or less are 99.3%, 80%, and 100% on data sets 1, 2, and 3, respectively, showing the high quality of the computed solutions.

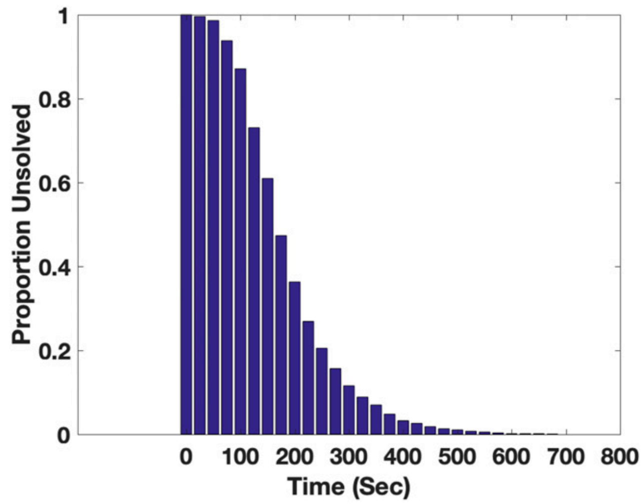
To illustrate the time required by the CG algorithm to solve these instances, Figure 9 shows the proportion of instances from data set 1 for which the computation time exceeds a given amount of time. Observe, for example, that around 50% of the instances are solved in less than 200 seconds. Given that the computation time is dominated by pricing and there are many pricing problems, parallelization can drastically accelerate CG.

Table 4. Comparative Average Results for Data Sets 1–3 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

	Data set 1					Data set 2					Data set 3				
	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. (2014a)	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. (2014a)	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

Notes. PCC and NPCC denote the planar correlation clustering and nonplanar correlation clustering algorithms of Zhang et al. (2014a). A "—" indicates that the result is unreported in the cited paper. CG, column generation.

Figure 9. (Color online) Proportion of Instances from Data Set 1 That Take a Longer Computational Time Than a Given Amount of Time



8. Conclusion

In this review 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 review 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. As shown in the recent works of Wang et al. (2017b, 2018) and Zhang et al. (2017), the well-studied computer vision applications of multiperson tracking, multiperson pose estimation, and multicell segmentation can be modeled as a MWSP problem. Because the set of hypotheses in the MWSP model grows exponentially with the size of the set of observations, a CG algorithm is employed to solve it. In these applications, the CG pricing problems correspond to tractable optimization problems, which are either dynamic programs or small-scale binary integer programs. To tighten the MP, SRIs can be used for these 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 the applications considered.

In future work, we recommend that the cost of hypotheses be dependent on a higher-level variable. For example, in the context of multiperson 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 multicell segmentation, where the higher-level variable could correspond to the cell species or cell orientation (with regard to rotation). The higher-level variable can be a latent variable, and the corresponding learning problem can be attacked using a latent structured SVM (Yu and Joachims 2009).

References

Andres B, Kappes JH, Beier T, Kothe U, Hamprecht FA (2011) Probabilistic image segmentation with closedness constraints. *Proc. 13th Internat. Conf. Comput. Vision* (IEEE, Piscataway, NJ), 2611–2618.

Andriluka M, Pishchulin L, Gehler P, Schiele B (2014) 2D human pose estimation: New benchmark and state of the art analysis. *Proc. 27th Conf. Comput. Vision Pattern Recognition* (IEEE, Piscataway, NJ), 3686–3693.

Arteta C, Lempitsky V, Noble J, Zisserman A (2012) Learning to detect cells using non-overlapping extremal regions. *Proc. 15th Internat. Conf. Medical Image Comput. Comput.-Assisted Intervention* (Springer, Berlin, Heidelberg), 348–356.

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

Bansal N, Blum A, Chawla S (2004) Correlation clustering. *J. 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. *Oper. Res.* 46:316–329.

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

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

Bernardin K, Stiefelhagen R (2008) Evaluating multiple object tracking performance: The CLEAR MOT metrics. *EURASIP J. Image Video Processing* 2008:246309.

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

Boykov Y, Kolmogorov V (2004) An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Machine Intelligence* 26(9):1124–1137.

Butt A, Collins R (2013) Multi-target tracking by Lagrangian relaxation to min-cost network flow. *Proc. 26th Conf. Comput. Vision Pattern Recognition* (IEEE, Piscataway, NJ), 1846–1853.

Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. *Proc. 18th Conf. Comput. Vision Pattern Recognition*, vol. 1 (IEEE, Piscataway, NJ), 886–893.

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: A large-scale hierarchical image database. *Proc. 22nd Conf. Comput. Vision Pattern Recognition* (IEEE, Piscataway, NJ), 248–255.

Desai C, Ramanan D, Fowlkes CC (2011) Discriminative models for multi-class object layout. *Internat. J. Comput. Vision* 95(1):1–12.

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

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

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 Conf. Comput. Vision Pattern Recognition* (IEEE, Piscataway, NJ), 1–8.

Funke J, Hamprecht F, Zhang C (2015) Learning to segment: Training hierarchical segmentation under a topological loss. Navab N, Hornegger J, Wells W, Frangi A, eds. *Proc. 18th Internat. Conf. Medical Image Comput. Comput.-Assisted Intervention*, Lecture Notes in Computer Science, vol. 9351 (Springer, Cham, Switzerland), 268–275.

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

Gschwind T, Irnich S (2016) Dual inequalities for stabilized column generation revisited. *INFORMS J. Comput.* 28(1):175–194.

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 Eur. Conf. Computer Vision* (Springer, Cham, Switzerland), 34–50.

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

Joncour C, Michel S, Sadykov R, Vanderbeck F (2010) Column generation based primal heuristics. *Electronic Notes Discrete Math.* 36:695–702.

Kappes JH, Speth M, Reinelt G, Schnörr C (2016) Higher-order segmentation via multicuts. *Comput. Visage Image Understanding* 143(C):104–119.

Karp RM (1972) Reducibility among combinatorial problems. *Proc. Sympos. Complexity Comput. Comput.* (Springer, Boston), 85–103.

Kolmogorov V (2006) Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Machine Intelligence* 28(10):1568–1583.

Komodakis N, Paragios N, Tziritas G (2007) MRF optimization via dual decomposition: Message-passing revisited. *Proc. 11th Internat. Conf. Comput. Vision* (IEEE, Piscataway, NJ), 1–8.

Leal-Taixe L, Pons-Moll G, Rosenhahn B (2012) Branch-and-price global optimization for multi-view multi-target tracking. *Proc. 25th Conf. Comput. Vision Pattern Recognition* (IEEE, Piscataway, NJ), 1987–1994.

Leal-Taixé L, Milan A, Reid I, Roth S, Schindler K (2015) MOT Challenge 2015: Toward a benchmark for multi-target tracking. Preprint, submitted April 8, https://arxiv.org/abs/1504.01942.

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 Conf. Comput. Vision Pattern Recognition* (IEEE, Piscataway, NJ), 6012–6020.

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 Conf. Comput. Vision Pattern Recognition* (IEEE, Piscataway, NJ), 4929–4937.

Ren X, Malik J (2003) Learning a classification model for segmentation. *Proc. 16th Internat. Conf. Comput. Vision Pattern Recognition* (IEEE, Piscataway, NJ), 10–17.

Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. *Proc. 18th Internat. Conf. Medical Image Comput. Comput.-Assisted Intervention* (Springer, Cham, Switzerland), 234–241.

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

Silberman N, Sontag D, Fergus R (2014) Instance segmentation of indoor scenes using a coverage loss. Fleet D, Pajdla T, Schiele B, Tuytelaars T, eds. *Proc. 14th Eur. Conf. Comput. Vision*, Lecture Notes in Computer Science, vol. 8689 (Springer, Cham, Switzerland), 616–631.

Sommer C, Straehle C, Koethe U, Hamprecht FA (2011) Ilastik: Interactive learning and segmentation toolkit. *Proc. 8th Internat. Sympos. Biomedical Imaging* (IEEE, Piscataway, NJ), 230–233.

Sontag D, Meltzer T, Globerson A, Jaakkola T, Weiss Y (2008) Tightening LP relaxations for MAP using message passing. *Proc. 24th Conf. Uncertainty Artificial Intelligence* (AUAI Press, Arlington, VA), 503–510.

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

Valério de Carvalho JM (2005) Using extra dual cuts to accelerate column generation. *INFORMS J. Comput.* 17(2):175–182.

Wang S, Fowlkes C (2015) Learning optimal parameters for multi-target tracking. *Proc. 26th British Machine Vision Conf.* (BMVA Press, UK), 484–501.

Wang S, Kording K, Yarkony J (2017a) Exploiting skeletal structure in computer vision annotation with Benders decomposition. Preprint, submitted September 13, https://arxiv.org/abs/1709.04411.

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 Eur. Conf. Comput. Vision* (Springer, Cham, Switzerland), 652–666.

Wang S, Wolf S, Fowlkes C, Yarkony J (2017b) Tracking objects with higher order interactions via delayed column generation. *Proc. 20th Internat. Conf. Artificial Intelligence Statist.* (PMLR), 1132–1140.

Wang S, Zhang C, Gonzalez-Ballester MA, Ihler A, Yarkony J (2017c) Multi-person pose estimation via column generation. Preprint, submitted September 18, https://arxiv.org/abs/1709.05982.

Yarkony J, Fowlkes C (2015) Planar ultrametrics for image segmentation. *Proc. 28th Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA), 64–72.

Yarkony J, Ihler A, Fowlkes C (2012) Fast planar correlation clustering for image segmentation. *Proc. 12th Eur. Conf. Comput. Vision* (Springer, Cham, Switzerland), 1169–1176.

Yu CN, Joachims T (2009) Learning structural SVMs with latent variables. *Proc. 26th Internat. Conf. Machine Learn.* (ACM, New York), 1169–1176.

Zhang C, Yarkony J, Hamprecht FA (2014a) Cell detection and segmentation using correlation clustering. *Proc. 17th Internat. Conf. Medical Image Comput. Comput.-Assisted Intervention* (Springer, Cham, Switzerland), 9–16.

Zhang C, Huber F, Knop M, Hamprecht FA (2014b) Yeast cell detection and segmentation in bright field microscopy. *Proc. 11th Internat. Sympos. on Biomedical Imaging* (IEEE, Piscataway, NJ), 1267–1270.

Zhang C, Wang S, Gonzalez-Ballester MA, Yarkony J (2017) Efficient column generation for cell detection and segmentation. Preprint, submitted September 21, https://arxiv.org/abs/1709.07337.

Zhang L, Li Y, Nevatia R (2008) Global data association for multi-object tracking using network flows. *Proc. 21st Conf. Comput. Vision Pattern Recognition* (IEEE, Piscataway, NJ), 1–8.