

Image detection using combinatorial auction

Simon Anuk ¹, Tamir Bendory ²,
and Amichai Painsky ³

¹School of Electrical Engineering, Tel Aviv University, Tel Aviv 6997801, Israel

²School of Electrical Engineering, Tel Aviv University, Tel Aviv 6997801, Israel

³The Industrial Engineering Department, Tel Aviv University, Tel Aviv 6997801, Israel

Corresponding author: Tamir Bendory (email: bendory@tauex.ac.il).

Tamir Bendory is supported in part by the BSF grant no. 2020159, the NSF-BSF grant no. 2019752, and the ISF grant no. 1924/21. Amichai Painsky is supported in part by ISF grant no. 963/21

ABSTRACT This paper studies the optimal solution of the classical problem of detecting the location of multiple image occurrences in a two-dimensional, noisy measurement. Assuming the image occurrences do not overlap, we formulate this task as a constrained maximum likelihood optimization problem. We show that the maximum likelihood estimator is equivalent to an instance of the winner determination problem from the field of combinatorial auction and that the solution can be obtained by searching over a binary tree. We then design a pruning mechanism that significantly accelerates the runtime of the search. We demonstrate on simulations and electron microscopy data sets that the proposed algorithm provides accurate detection in challenging regimes of high noise levels and densely packed image occurrences.

INDEX TERMS Combinatorial auction, gap statistics, image detection, winner determination problem

I. Introduction

This paper studies the problem of accurately detecting multiple occurrences of a template image $s \in \mathbb{R}^{W \times W}$ located in a noisy measurement y (the observed, acquired, data). Specifically, let $y \in \mathbb{R}^{N \times M}$ be a measurement of the form:

$$y[n, m] = \sum_{k=1}^K s \left[n - c_n^{(k)}, m - c_m^{(k)} \right] + w[n, m], \quad (1)$$

where $\mathbf{c}^{(k)} = (c_n^{(k)}, c_m^{(k)})$ is the upper-left corner coordinate of the k -th image occurrence. We model the noise term $w[n, m]$ as an i.i.d. Gaussian noise with mean zero and variance σ^2 . Our goal is to estimate the unknown image locations $\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(K)}$ from the observation y . While K —the number of image occurrences—is an unknown parameter, it is instructive to first assume that K is known; we later omit this assumption.

The problem of detecting multiple image occurrences in noisy data appears in various image processing applications, including fluorescence microscopy [1], [2], astronomy [3], anomaly detection [4], neuroimaging [5]–[7], and electron microscopy [8]–[10]. A detailed discussion on these applications is provided in [11]. However, although many algorithms were developed over the years (see, e.g. [4]) in most cases

there is no known statistically optimal method for detecting the images.

A particularly important motivation for this paper is single-particle cryo-electron microscopy (cryo-EM)—a leading technology for determining the three-dimensional structure of biological molecules [12]–[14]. In a cryo-EM experiment, multiple samples of biological molecules are frozen in a thin layer of ice. The samples might be densely packed, but they do not overlap. One of the first stages in the computational pipeline of cryo-EM involves locating images within a large and highly noisy measurement [8]–[10]. Motivated by cryo-EM, we assume that the image occurrences do not overlap, but otherwise may be arbitrarily spread in the measurement. Namely, each image occurrence is separated by at least W pixels from any other image occurrence, in at least one dimension:

$$\max \left(\left| c_n^{(k)} - c_n^{(\ell)} \right|, \left| c_m^{(k)} - c_m^{(\ell)} \right| \right) \geq W, \quad \forall k \neq \ell. \quad (2)$$

We refer to (2) as the *separation condition*. Figure 1 illustrates an example of a measurement that contains six image occurrences. The three image occurrences on the right side are densely packed (but satisfy the separation condition (2)), while the image occurrences on the left are separated by more than $2W$ pixels (twice the separation condition (2));

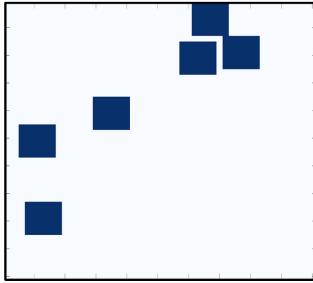


FIGURE 1. The left panel displays a noiseless measurement with six image occurrences: the images on the right side are densely packed, while the images on the left side are separated by more than $2W$ pixels. The right panel shows a noisy version of the same measurement, with $\text{SNR} = -30[\text{dB}]$. Our goal is to accurately estimate the location of the six images from the noisy measurement.

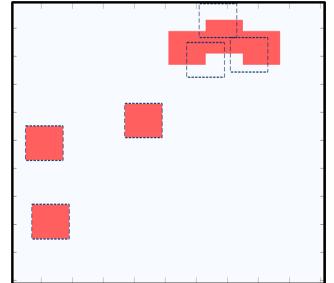
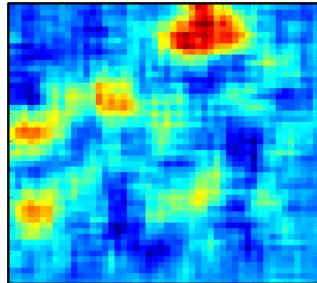
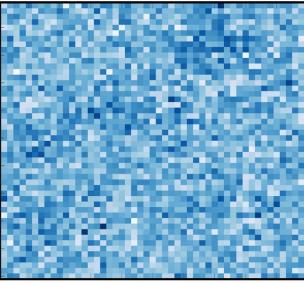


FIGURE 2. On the left, the correlation between the noisy measurement y , as depicted in Figure 1, and the template image s . On the right, the dashed lines display the noiseless measurement, as depicted in Figure 1. The red images are the output of the greedy algorithm, which fails to detect the location of the images on the right when the image occurrences are densely packed.

the distinction between these two scenarios plays a key role in this paper.

Assuming the template image s and the number of image occurrences K are known, the maximum likelihood estimator (MLE), for the sought locations, is given by

$$\arg \min_{\hat{\mathbf{c}}^{(1)}, \dots, \hat{\mathbf{c}}^{(K)}} \left\| y - \sum_{k=1}^K s \left[n - \hat{c}_n^{(k)}, m - \hat{c}_m^{(k)} \right] \right\|_2^2, \quad (3)$$

where we denote by $\hat{\mathbf{c}}^{(1)}, \dots, \hat{\mathbf{c}}^{(K)}$ the estimates of $\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(K)}$, respectively. Taking the separation condition (2) into account, the MLE results in a constrained optimization problem:

$$\begin{aligned} & \arg \max_{\hat{\mathbf{c}}^{(1)}, \dots, \hat{\mathbf{c}}^{(K)}} \sum_{k=1}^K \sum_{n=1}^N \sum_{m=1}^M y[n, m] s \left[n - \hat{c}_n^{(k)}, m - \hat{c}_m^{(k)} \right] \\ & \text{s.t. } \max \left(\left| \hat{c}_n^{(k)} - \hat{c}_n^{(\ell)} \right|, \left| \hat{c}_m^{(k)} - \hat{c}_m^{(\ell)} \right| \right) \geq W, \quad \forall k \neq \ell. \end{aligned} \quad (4)$$

Efficiently computing the optimal solution to this optimization problem is the prime goal of this work.

A popular and natural heuristic for solving (4) is as follows [8], [9], [15]–[17]. First, we correlate the template image s with the measurement y . We set $\hat{\mathbf{c}}^{(1)}$ as the location which maximizes the correlation. The next estimator, $\hat{\mathbf{c}}^{(2)}$, is chosen as the next maximum of the correlation that satisfies the separation condition (2) with respect to $\hat{\mathbf{c}}^{(1)}$. The same approach is applied sequentially for estimating the rest of the locations $\hat{\mathbf{c}}^{(3)}, \dots, \hat{\mathbf{c}}^{(K)}$. We refer to this algorithm as the *greedy algorithm*, and it can be thought of as a variation of the well-known template matching algorithm that takes the separation condition (2) into account. This algorithm is highly efficient since the correlations can be calculated using the FFT algorithm [18].

Figure 2 illustrates the correlation between the noisy measurement y of Figure 1 and the template image s , and the output of the greedy algorithm. Evidently, the greedy algorithm successfully detects the locations of the well-separated image occurrences on the left end of the measurement (when the images are separated by at least $2W$

pixels, twice the separation condition of (2)) but completely fails when the image occurrences are densely packed (right end). The reason is that correlation doubles the support of the image. Thus, the greedy algorithm is likely to consider two close image occurrences as one, especially in the presence of high levels of noise. In this paper, we design an algorithm that detects the image occurrences accurately for any measurement that satisfies the separation condition (2)—even for measurements with highly dense image occurrences—by computing the constrained MLE (4).

While many papers designed object detection algorithms, as far as we know, none of them is guaranteed to achieve the MLE [4], as we propose in this paper. For example, a CNN-based technique was designed in [19]; this method works well but has no theoretical guarantees and requires a large data set for training. Other papers focused on accelerating the template matching algorithm [20]. The MLE for the one-dimensional version of (1) was considered in [17] based on dynamic programming. A multiple hypothesis approach was suggested in [11], [21], [22], and the statistical properties of the problem were analyzed in [23].

In Section II, we reformulate the constrained MLE problem (4) as a Winner Determination Problem (WDP)—an optimization problem from the field of combinatorial auction. Section III shows how the problem can be solved exactly by binary tree search. While scanning the entire tree is intractable, we design an efficient pruning method to significantly accelerate the search, while guaranteeing to find the optimal solution. This section outlines the technical details of the proposed algorithm. Section IV presents results on simulated data, and Section V shows results on experimental electron microscopy datasets. Section VI concludes the paper.

II. Image detection as a combinatorial auction problem

This section begins by introducing the classical combinatorial auction problem and the WDP. Then, we show how the constrained optimization problem (4) can be reformulated as an optimization problem, which is highly similar to

the classical WDP. This formulation leads to an efficient algorithm, whose details are introduced in Section III.

A. Classical combinatorial auction

Consider an auction of a variety of different goods. Unlike a classical auction, where goods are sold individually, a bidder can place a bid on a bundle of goods in a combinatorial auction. The auction manager receives multiple price offers for different bundles of goods. The goal of the auction manager is to maximize his revenue by allocating the goods to the highest submitted bids. Naturally, the same good may not be allocated to multiple bidders. This means that if two bidders place bids over bundles with overlapping goods, only one of the bids can win.

Let $\mathcal{G} = \{\gamma_1, \dots, \gamma_A\}$ be a set of given goods, and let $\mathcal{B} = \{\beta_1, \dots, \beta_B\}$ be a set of bids. Each bid β_i includes a bundle of goods $g(\beta_i) \subseteq \mathcal{G}$, which the bidder seeks to acquire. For this bundle, the bidder of β_i proposes a price, $p(\beta_i)$. Let x_i be an indicator variable that corresponds to the winning bids. That is, we set $x_i = 1$ if β_i is a winning bid and $x_i = 0$ otherwise. An allocation, π , is a set of winning bids that do not overlap. Specifically, $\pi \subseteq \mathcal{B}$ is a subset of the bids, where $\forall \beta_2 \neq \beta_1 \in \pi, g(\beta_1) \cap g(\beta_2) = \emptyset$. Therefore, an allocation cannot include two bids that include the same good.

Using this notation, the combinatorial auction problem is given by:

$$\begin{aligned} \max_{x_1, \dots, x_B} \quad & \sum_{i=1}^B x_i p(\beta_i) \\ \text{s.t.} \quad & \sum_{i|\gamma \in g(\beta_i)} x_i \leq 1, \quad \forall \gamma \in \mathcal{G}, \\ & x_i \in \{0, 1\}, \quad \forall i. \end{aligned} \quad (5)$$

The first constraint ensures that each good can be allocated once at most, while the second constraint guarantees that each bid can be either included or excluded from the chosen allocation. We define the *optimal allocation* as the allocation that maximizes the revenue of the auction manager. The problem of determining the optimal allocation is known as the WDP [24]. It is known that the WDP is an *NP*-hard problem and cannot be approximated, within any constant, by a polynomial-time algorithm [25], [26].

B. Image detection as a winner determination problem

We now cast the constrained optimization problem (4) as a variation of the WDP. We associate every pixel in the measurement with a single good and let \mathcal{B} be a set of bids. Accepting $\beta_{ij} \in \mathcal{B}$ implies that our estimator includes an image occurrence whose upper left corner is located at the (i, j) -th pixel. The goods requested by the bid, $g(\beta_{ij})$, are a $W \times W$ set of pixels, corresponding to a possible image occurrence in the measurement. The offered price of the bid, $p(\beta_{ij})$, is the correlation between the measurement and the

template image, namely,

$$p(\beta_{ij}) = \sum_{n=1}^N \sum_{m=1}^M y[n, m] s[n - i, m - j]. \quad (6)$$

The revenue of the auctioneer is the sum of offered prices by all accepted bids. Since we are estimating the upper left corner coordinates of images in the measurement, we require that the estimated coordinates will satisfy $\hat{c}_n^{(k)} \leq N - W + 1$ and $\hat{c}_m^{(k)} \leq M - W + 1$, so we do not exceed the measurement's dimension. We denote $\tilde{N} = N - W + 1$ and $\tilde{M} = M - W + 1$.

Let us define $\mathbf{X} \in \{0, 1\}^{\tilde{N} \times \tilde{M}}$ as a matrix of indicators, in which $x_{ij} \in \mathbf{X}$ is equal to 1 if the bid β_{ij} is included in the allocation, and 0 otherwise. To enforce the separation constraint (2), we require that

$$\mathbf{1}_w^T \tilde{\mathbf{X}}_{ij} \mathbf{1}_w \in \{0, 1\}, \quad \forall i, j, \quad (7)$$

where $\mathbf{1}_w$ is an all-ones $W \times 1$ column vector and $\tilde{\mathbf{X}}_{ij}$ is a $W \times W$ sub-matrix of \mathbf{X} , whose upper left entry is the (i, j) entry in the matrix \mathbf{X} . Using this notation, and assuming K is known, the constrained likelihood problem (4) can be written as:

$$\begin{aligned} \max_{x_1, \dots, x_{\tilde{N} \tilde{M}}} \quad & \sum_{i=1}^{\tilde{N}} \sum_{j=1}^{\tilde{M}} x_{ij} p(\beta_{ij}) \\ \text{s.t.} \quad & \mathbf{1}_{\tilde{N}}^T \mathbf{X} \mathbf{1}_{\tilde{M}} = K, \\ & \mathbf{1}_w^T \tilde{\mathbf{X}}_{ij} \mathbf{1}_w \leq 1, \quad \forall i, j, \\ & x_{ij} \in \{0, 1\}, \quad \forall i, j, \end{aligned} \quad (8)$$

where the last two conditions are, together, equivalent to (7). Therefore, solving (8) is equivalent to (4), where the non-zero entries of the optimal \mathbf{X} correspond to the MLE of the locations in (4).

We underscore that the optimization problem (8) is slightly different from the classical combinatorial auction (5). In the classical setup, the auctioneer aims to include as many bids as possible in the optimal allocation to maximize the revenue, whereas in our model the optimal allocation must include exactly K bids. Furthermore, in our model, $p(\beta_{ij})$ can be negative. Thus, we cannot use existing algorithms that solve the classical WDP (5), such as the one proposed in [24], and we design a new algorithm tailored for our constrained setup. Importantly, the designed algorithm must be computationally efficient: solving (8) in a brute-force manner requires searching over an exponential number of bids, rendering this approach intractable. For example, if $N, M = 30$ and $K = 5$, then the number of possible allocations is $\sim 10^{12}$. The next section introduces a pruning technique to solve (8), which dramatically reduces the number of explored allocations.

III. WDP for image detection

To delineate the proposed algorithm, we introduce further notation. Two bids, $\beta_i, \beta_j \in \mathcal{B}$, are considered to be conflicting if $g(\beta_i) \cap g(\beta_j) \neq \emptyset$. We extend the functions $g(\cdot)$ and $p(\cdot)$ to apply to allocations: $g(\pi) = \bigcup_{\beta \in \pi} g(\beta)$

and $p(\pi) = \sum_{\beta \in \pi} p(\beta)$. A partial allocation, π_k , is an allocation in which the number of bids is $k < K$. A bid β_i is considered to be conflicting with a partial allocation π_k if $g(\beta_i) \cap g(\pi_k) \neq \emptyset$. We denote the set of bids that conflict with a partial allocation π_k as C_{π_k} . Hence, given a partial allocation π_k , the set of allowed bids, from which a bid can be added to π_k , is $\bar{C}_{\pi_k} = \mathcal{B} \setminus C_{\pi_k}$. In addition, we define $\tilde{\pi}_{\text{opt}}$ as the current optimal allocation that was found by our algorithm so far.

A. Binary tree construction

The algorithm begins with forming a binary tree data structure. Binary trees, widely used in search algorithms in computer science [27], are suitable for examining every possible allocation through a systematic navigation to determine the optimal allocation.

Given a set of bids, each node in the tree corresponds to a simple binary question: whether to include or exclude the bid in the allocation. This results in a *full binary tree* where, except for the leaf nodes, each node has two children corresponding to the same bid. We introduce a simple example to illustrate the construction of the binary tree. Let us consider $y \in \mathbb{R}^4$, a template image of a unit size, $s \in \mathbb{R}$, and $K = 2$. As explained in Section II, we cast the image detection problem as a variation of the WDP. In this example, every bid includes a bundle of a single good (since the template image is of a unit size). We first set β_1 as the tree's root, and β_4 forms the tree's leaves. To explore an allocation that includes the current node, we navigate to the node's left child and add the current node to our partial allocation, π_k . Conversely, navigating to the right child represents the alternative choice of excluding the current node from π_k . Figure 3 demonstrates the binary tree we obtain in this simple example.

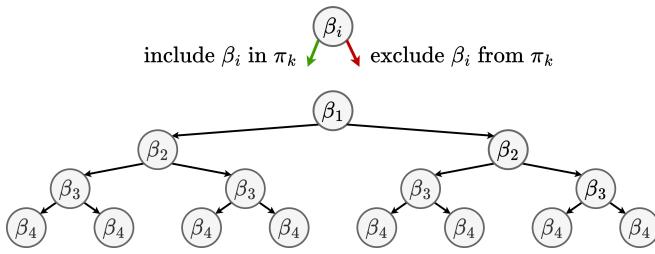


FIGURE 3. An example of a binary tree, corresponding to a measurement with four elements.

As previously mentioned, the binary tree enables us to conduct an exhaustive search across all possible allocations. Therefore, by using this search mechanism, the optimal allocation will inevitably be encountered during our search.

B. Tree pruning

Exhaustively searching through the tree leads necessarily to the optimal allocation, but is computationally infeasible even for problems of small dimensions. To overcome this issue,

we use the algorithmic technique of *branch-and-bound* and present a *pruning* method to avoid exploring a large portion of the allocations in advance by eliminating sub-trees that cannot yield the optimal allocation.

We illustrate this method using the example of Figure 3. We refer to a smaller binary tree, spanned by a node other than the root, as a *sub-tree*. Let us assume that our optimal allocation is $\pi_{\text{opt}} = \{\beta_1, \beta_3\}$ and that we have already encountered our optimal allocation π_{opt} in our search. At this point, we are unaware that this is the optimal allocation, so we continue exploring the remaining possible allocations. Assume we have chosen to exclude β_1 from the empty partial allocation, π_0 . Following that, two bids are left to be found in the sub-tree spanned by β_2 . Figure 4 demonstrates this sub-tree. The maximal revenue achievable by exploring this sub-tree is bounded by the sum of the prices of the two bids with the highest revenue within it. Since this revenue is less than what is achieved by the optimal allocation (by the assumption that we already encountered the optimal allocation), we can avoid exploring the allocations in this sub-tree. Thus, in this example, we reduce the computational burden by half, as we prune half of the binary tree.

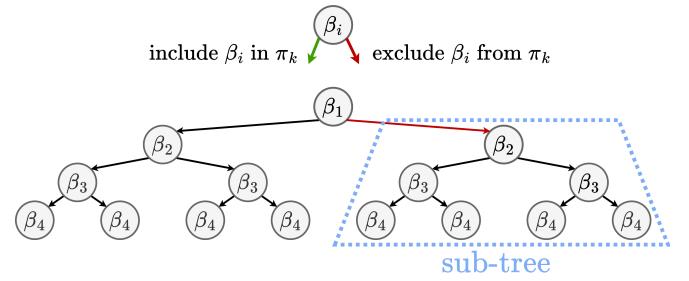


FIGURE 4. A sub-tree that corresponds to excluding β_1 from the partial allocation π_k . If the maximal revenue that can be achieved by exploring this sub-tree is less than the revenue of our current optimal allocation, π_{opt} , we prune it.

We now formulate this example as a general pruning mechanism. Say we have reached a partial allocation π_k . We sort the bids in \bar{C}_{π_k} (the set of allowed bids) in descending order according to their revenues. We denote the $K - k$ maximal revenue bids in \bar{C}_{π_k} as $\bar{C}_{\pi_k}^{(K-k)}$. Summing the revenues of the bids in $\bar{C}_{\pi_k}^{(K-k)}$ bounds on the remaining revenue to be allocated in this sub-tree

$$h(\pi_k) := \sum_{\beta \in \bar{C}_{\pi_k}^{(K-k)}} p(\beta). \quad (9)$$

This function will enable us to determine if the optimal allocation cannot be achieved within the current sub-tree, and thus we can prune it. Specifically, we prune the sub-tree if the sum of $p(\pi_k)$ and $h(\pi_k)$ is less than the revenue of our current optimal allocation, $p(\tilde{\pi}_{\text{opt}})$, i.e.,

$$p(\pi_k) + h(\pi_k) < p(\tilde{\pi}_{\text{opt}}). \quad (10)$$

Since we only prune sub-trees that do not lead to the optimal allocation, our proposed algorithm remains optimal. We note,

however, that $h(\pi_k)$ is not necessarily a tight upper bound as it ignores the constraint that the remaining $K - k$ bids must not conflict with each other.

C. Sorting the bids

Our algorithm examines each potential optimal allocation, and thus the optimality of the algorithm is unaffected by rearranging the order of the bids. We aim to reorder the bids in such a way that maximizes the efficiency of the pruning mechanism. Note that the earlier we secure a high revenue allocation $p(\tilde{\pi}_{\text{opt}})$, the more the stopping criterion (10) is met, resulting in fewer allocations to be examined. Hence, we begin the algorithm by sorting the bids by their revenue in descending order. The bid with the highest revenue is fixed as the tree's root, while the bid with the lowest revenue forms the tree's leaves. This strategy enhances the frequency in which the pruning condition is met, thereby boosting its efficiency and accelerating the algorithm's runtime. For example, in the experiments in Section IV, we observed that sorting the bids can lead to acceleration by a factor of around 10.

D. Algorithm description

We are now ready to describe our algorithm for detecting image occurrences in a noisy measurement. We will use the example above to explain the guiding principles. We start by sorting the bids in descending order by their revenues, as described in Section C. An example of a sorted tree is presented in Figure 5.

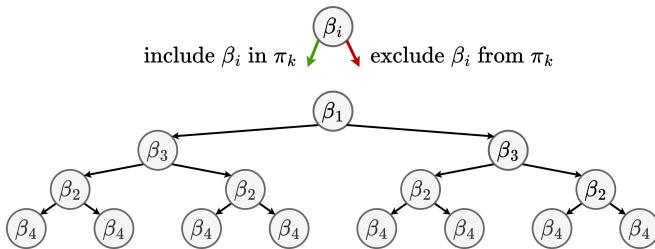


FIGURE 5. A sorted binary tree, corresponding to the tree that was presented in Figure 3.

Our algorithm is designed such that when we encounter a node, it is added to our partial allocation only if it does not conflict with the partial allocation and if the partial allocation itself does not satisfy the pruning condition (10). We begin our search at the tree's root, β_1 . As this is the first node we encounter, we add the node to our partial allocation and move to the left child, β_3 . At this point, β_3 is added to the partial allocation, since it does not conflict with our partial allocation nor does the partial allocation itself satisfy the pruning condition (10). Since this is the first allocation we attain, this is the current optimal allocation. We note that the first allocation we explore is the output of the greedy algorithm. This means that the greedy algorithm is not necessarily optimal, since the optimal allocation may be

achieved in a later stage. In our specific simple example, indeed the greedy algorithm attains the optimal solution (since the first two bids do not conflict), but in order to explain the full mechanism of the algorithm, we describe the remaining steps.

We then remove the last bid we added, β_3 , and continue examining allocations excluding that bid. To do so, we move to the right child of β_3 . Upon reaching β_2 , we check if our partial allocation π_1 satisfies the pruning condition (10). Since the optimal allocation is $\pi_{\text{opt}} = \{\beta_1, \beta_3\}$, the pruning condition is met, and hence we prune the sub-tree spanned by β_2 and return to the parent, β_3 . Since we examined the sub-tree spanned by β_3 , we next return to the parent, β_1 , and remove β_1 from our partial allocation. This way, we examined all allocations that include β_1 , and we move to the node's right child. This process continues until the algorithm explores the entire tree. Upon completion, it returns the optimal allocation. The general algorithmic pipeline is described in Algorithm 1.

Algorithm 1 A WDP algorithm to solve (8)

Input: Set of bids \mathcal{B} and number of image occurrences K

Output: Optimal allocation π_{opt}

Initialize: $\text{curr} \leftarrow \text{root}$, $\pi_k = \{\}$, $\tilde{\pi}_{\text{opt}} = \{\}$

Sort \mathcal{B} by revenue and construct the binary tree

while $\text{curr} \neq \text{NULL}$ **do**

- if** $\text{size}(\pi_k) = K$ **then**

 - if** $p(\pi_k) > p(\tilde{\pi}_{\text{opt}})$ **then**

 - $\tilde{\pi}_{\text{opt}} = \pi_k$

end if

$\pi_k \leftarrow \pi_k \setminus \text{curr.parent}$

$\text{curr} \leftarrow \text{curr.parent}$

else

- if** have not explored left sub-tree **then**

 - if** $p(\pi_k) + h(\pi_k) < p(\tilde{\pi}_{\text{opt}})$ **then**

 - prune left and right sub-trees.

else if $\text{curr} \in \bar{C}_{\pi_k}$ **then**

- $\pi_k \leftarrow \pi_k \cup \text{curr}$
- $\text{curr} \leftarrow \text{curr.left}$

end if

else if right sub-tree exists **then**

- $\text{curr} \leftarrow \text{curr.right}$

else

- if** $\text{curr} \in \pi_k$ **then**

 - $\pi_k \leftarrow \pi_k \setminus \text{curr}$

end if

$\text{curr} \leftarrow \text{curr.parent}$

end if

end if

end while

E. Estimating the number of image occurrences using the gap statistics principle

In many real-world scenarios, such as electron microscopy [12], [22], the exact number of image occurrences K is unknown. This imposes a major challenge given that both the greedy algorithm and Algorithm 1 require a fixed K . A common practice is solving (4) for various K values, choosing the one that induces the steepest change in the objective value. This heuristic is called the “knee” method and has been widely adopted across diverse domains, such as clustering [28] and regularization [29].

Here, we propose employing the gap statistics principle, which can be thought of as a statistical technique to identify the “knee” [28]. This principle is widely applied to a variety of domains, as discussed for example in [30], [31]. The underlying idea of gap statistics is to adjust the objective value curve (as a function of the possible values of K) by comparing it with its expectation under a null reference. Let us define the gap statistic as:

$$\text{gap}(K) = p(\boldsymbol{\pi}_{\text{opt}}(K)) - \mathbb{E}^* p(\boldsymbol{\pi}_{\text{opt}}(K)), \quad (11)$$

where $p(\boldsymbol{\pi}_{\text{opt}}(K))$ represents the revenue of the optimal allocation for K image occurrences (computed by Algorithm 1), and $\mathbb{E}^* p(\boldsymbol{\pi}_{\text{opt}}(K))$ is the expectation for a measurement of dimensions $N \times M$ derived from a “null” reference distribution. To form the “null” distribution, we randomly rearrange the pixels of the measurement, resulting in an unstructured measurement, with no template image occurrences. Specifically, we estimate the expectation under the null by $\mathbb{E}^* p(\boldsymbol{\pi}_{\text{opt}}(K)) \approx \frac{1}{R} \sum_{r=1}^R p(\boldsymbol{\pi}_{\text{opt}}^r(K))$, where $p(\boldsymbol{\pi}_{\text{opt}}^r(K))$ is the revenue achieved by the optimal allocation, at the r -th rearrangement; in the experiments below, we set $R = 50$. The estimate of K —the number of image occurrences—is the K that maximizes $\text{gap}(K)$ and is denoted by \hat{K} .

IV. Numerical experiments

In this section, we conduct numerical experiments to compare Algorithm 1 with the alternative greedy algorithm. We use the F_1 -score to evaluate the performance of both methods [32]–[35]. The F_1 score is defined as:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{TPR}}{\text{Precision} + \text{TPR}}, \quad (12)$$

where Precision is the ratio of correct detections divided by the total number of detections, and TPR, which stands for True Positive Rate, is the ratio of correct detections divided by the total number of image occurrences. We note that in the case where the number of image occurrences K is known, Precision equals TPR.

We define a detection as correct if:

$$\max \left(\left| \hat{c}_n^{(k)} - c_n^{(k)} \right|, \left| \hat{c}_m^{(k)} - c_m^{(k)} \right| \right) \leq \frac{W}{2}. \quad (13)$$

That is, a detection is classified as correct if the estimated image location is within $\frac{W}{2}$ pixels from the true location.

This convention is common in the image detection literature [11], [21]. We define the SNR of the measurement as

$$\text{SNR} = 10 \log \frac{KW^2}{\sigma^2 NM}, \quad (14)$$

where σ^2 is the variance of the noise. To compare our algorithm with the greedy algorithm, which works well when the supports of the correlated image occurrences do not overlap, we define another separation condition.

$$\max \left(\left| c_n^{(k)} - c_n^{(\ell)} \right|, \left| c_m^{(k)} - c_m^{(\ell)} \right| \right) \geq 2W, \quad \forall k \neq \ell. \quad (15)$$

We refer to this condition as the well-separated condition.

For a given set of parameters M, N, W, K , we generate a measurement y as follows. We begin by placing the first image occurrence at a random location (drawn from a uniform distribution) within the measurement. Then, we draw another location and place the image occurrence if it meets either (2) or (15). This process is repeated until all K images have been placed in the measurement. Finally, we add i.i.d. white Gaussian noise with zero mean and σ^2 variance to the measurement. The code to reproduce the numerical experiments is publicly available at <https://github.com/saimonanuk/Optimal-detection-of-non-overlapping-images-via-combinatorial-auction>.

A. A known number of image occurrences

In the first experiment, we assume that the number of image occurrences K is known, and both algorithms receive the true number of image occurrences K as input. We set $N = M = 40$ and $K = 4$, and choose a template image with all-ones entries of size $W = 3$. We generate the measurement such that the separation condition (2) is met, and at least one pair of image occurrences is separated precisely by W pixels. For each SNR level, we conduct 1000 trials, each with a fresh measurement. Figure 6 shows the average F_1 score of Algorithm 1 and the greedy algorithm as a function of the SNR. Evidently, Algorithm 1 outperforms the greedy algorithm in all SNR regimes. As expected, in high SNR regimes, Algorithm 1 achieves $F_1 = 1$. On the contrary, the greedy algorithm fails even in the high SNR regime, achieving $F_1 = 0.88$, due to the proximity of the image occurrences in the measurement. The average runtime per trial of Algorithm 1 was approximately 80 seconds, whereas the average runtime of the greedy algorithm was 0.1 seconds.

We repeated the same experiment, but now the image occurrences are well-separated according to (15). The results of this experiment are illustrated in Figure 7. As expected, in this case, both algorithms achieve the same results, and the running time remains the same.

B. An unknown number of image occurrences

We repeated the first experiment from the previous section (when the image occurrences satisfy (2) but not (15)), but now the number of image occurrences K is unknown. To estimate K , we used the gap statistics principle. This estimate is then used as an input for both algorithms to

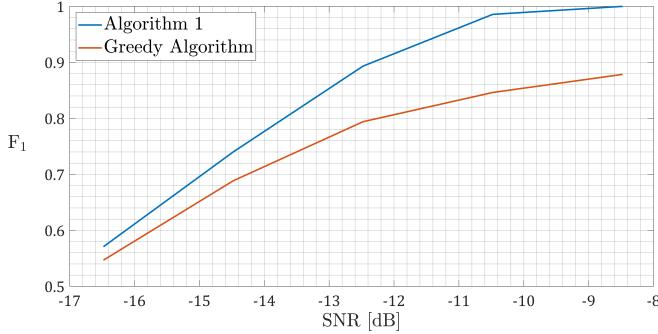


FIGURE 6. The average F_1 score, as a function of the SNR, of Algorithm 1 and the greedy algorithm, assuming the number of image occurrences is known.

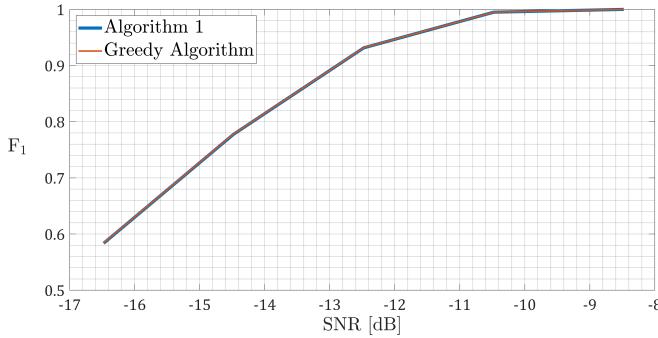


FIGURE 7. The average F_1 score, as a function of the SNR, of Algorithm 1 and the greedy algorithm, where the image occurrences satisfy the well-separated condition (15), assuming the number of image occurrences is known.

determine the image locations. The results are presented in Figure 8.

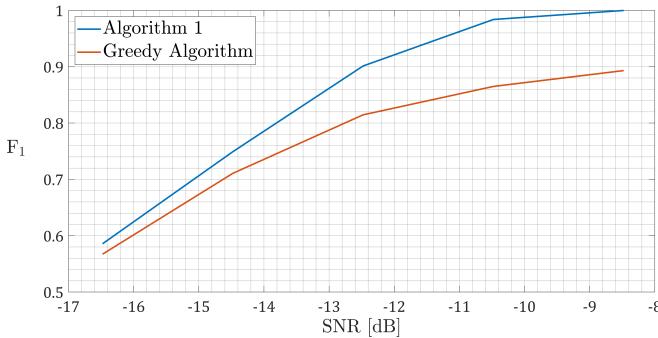


FIGURE 8. The average F_1 score, as a function of the SNR, of Algorithm 1 and the greedy algorithm assuming the number of image occurrences is unknown.

Similar to the previous case, in the unknown K scenario, Algorithm 1 outperforms the greedy algorithm in all SNR regimes. In addition, Algorithm 1 provides better estimates of the number of signal occurrences K . For example, for $\text{SNR} = -12.5$, Algorithm 1 gets 87.2% accuracy in estimating K precisely, while the greedy algorithm gets only 80%.

This gap drops with the SNR until they eventually converge to the same estimations when the SNR is low enough.

V. Cryo-electron microscopy examples

We applied Algorithm 1 and the greedy algorithm to two cryo-EM datasets, available at the EMPIAR repository [36]. We assume a disk-like shape for the images. Following standard procedures in this field, we performed a sequence of preprocessing steps before applying both algorithms. First, we whitened the measurement by manually selecting “noise-only” areas (i.e., without images), which were used to estimate the empirical covariance matrix of the noise. The data is then normalized by the inverse of this covariance matrix. Then, the data is downsampled by a factor of 12, concurrently reducing computational load and amplifying the SNR. Ultimately, we took the whitened, downsampled measurement and determined patches in which the image occurrences are not well-separated, namely, do not satisfy (15). We focused on small dense patches since this is the scenario in which Algorithm 1 shines. In well-separated (15) areas, it is recommended to use the computationally efficient greedy algorithm. Figure 9 displays two different measurements, after the preprocessing steps.

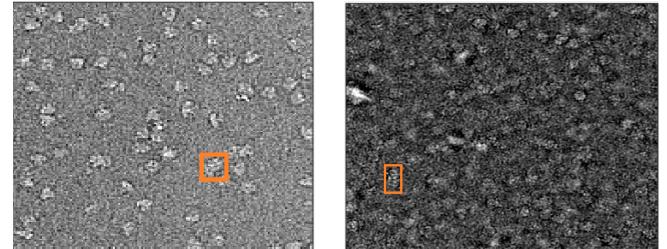


FIGURE 9. Two cryo-EM measurements after preprocessing steps. The left panel presents a measurement from the EMPIAR-10081 data set and the right panel shows a measurement from the EMPIAR-10217 data set. The orange squares are the dense areas used as the input for both algorithms.

A. EMPIAR 10081

Our first experiment is conducted on the EMPIAR-10081 data set of the Human HCN1 Hyperpolarization-Activated Channel [37]. The measurement is presented in Figure 9. We set the radius of the disc used as a template image to 3 pixels. The number of image occurrences is estimated based on gap statistics. Figure 10 shows the output of both algorithms. Algorithm 1 identifies two separated images. In contrast, the greedy algorithm identifies a single occurrence. For comparison, we also show the output of a popular detection algorithm for cryo-EM data sets (particle picker), called TOPAZ [10]. This algorithm is based on a deep learning technique (see also [38]).

B. EMPIAR 10217

Our second experiment is conducted on the EMPIAR-10217 data set of the bovine liver glutamate dehydrogenase [36].

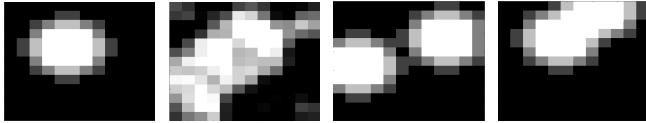


FIGURE 10. From left to right: the output of the greedy algorithm, the original patch taken from the data set EMPIAR 10081 as illustrated in Figure 9, the output of Algorithm 1, and the output of the cryo-EM particle picker TOPAZ [10].

We set the radius of the disc used as a template image to 6 pixels. The measurement is presented in Figure 9 and the results of both algorithms are displayed in Figure 11. Similar to the previous example, Algorithm 1 identifies two different particle images, while the greedy algorithm finds only one particle image in the measurement.

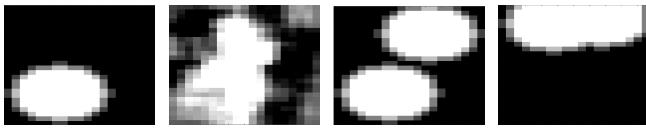


FIGURE 11. From left to right: the output of the greedy algorithm, the original patch taken from the data set EMPIAR 10217 as illustrated in Figure 9, the output of Algorithm 1, and the output of the cryo-EM particle picker TOPAZ [10].

VI. CONCLUSION

This paper studies the problem of detecting multiple image occurrences in a two-dimensional, noisy measurement. We approach this task by formulating it as a constrained maximum likelihood optimization problem. By showing the equivalence between this problem and a version of the winner determination problem, we design an efficient search algorithm over a binary tree. Our algorithm can find the optimal solution, even in scenarios characterized by high noise levels and densely packed image occurrences.

Our methodology can be extended in a couple of directions that we leave for future work. First, we assume white Gaussian noise, whereas in many applications, such as cryo-EM, that noise is not white [12]. Using the formulation of the detection problem as a combinatorial auction (8), changing the noise statistics will only change the revenue function p , while the rest of the optimization problem remains the same. Similarly, one may want to introduce a prior on the sought locations (in the Bayesian sense). In this case, the goal will be to maximize the posterior distribution under the separation condition. This again will not change the structure of the algorithm, just the definition of the revenue. Second, our pruning mechanism is quite conservative since it ignores the separation condition. An important future work includes designing a more effective pruning mechanism that will accelerate the search over the tree (and thus the algorithm's runtime), perhaps at the cost of a small bounded deviation from the optimal solution.

Acknowledgments

T.B. is supported in part by the BSF grant no. 2020159, the NSF-BSF grant no. 2019752, and the ISF grant no. 1924/21. A.P. is supported by the in part by ISF grant no. 963/21. The authors are grateful to Amnon Balanov for his assistance with TOPAZ.

REFERENCES

- [1] C. Geisler, A. Schönle, C. Von Middendorff, H. Bock, C. Eggeling, A. Egner, and S. Hell, “Resolution of $\lambda/10$ in fluorescence microscopy using fast single molecule photo-switching,” *Applied Physics A*, vol. 88, pp. 223–226, 2007.
- [2] A. Egner, C. Geisler, C. Von Middendorff, H. Bock, D. Wenzel, R. Medda, M. Andresen, A. C. Stiel, S. Jakobs, C. Eggeling, *et al.*, “Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters,” *Biophysical journal*, vol. 93, no. 9, pp. 3285–3290, 2007.
- [3] P. Brutt, C. Genovese, C. J. Miller, R. C. Nichol, and L. Wasserman, “Spike hunting in galaxy spectra,” 2005.
- [4] T. Ehret, A. Davy, J.-M. Morel, and M. Delbracio, “Image anomalies: A review and synthesis of detection methods,” *Journal of Mathematical Imaging and Vision*, vol. 61, pp. 710–743, 2019.
- [5] K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans, “A unified statistical approach for determining significant signals in images of cerebral activation,” *Human brain mapping*, vol. 4, no. 1, pp. 58–73, 1996.
- [6] C. R. Genovese, N. A. Lazar, and T. Nichols, “Thresholding of statistical maps in functional neuroimaging using the false discovery rate,” *Neuroimage*, vol. 15, no. 4, pp. 870–878, 2002.
- [7] J. E. Taylor and K. J. Worsley, “Detecting sparse signals in random fields, with an application to brain mapping,” *Journal of the American Statistical Association*, vol. 102, no. 479, pp. 913–928, 2007.
- [8] A. Eldar, B. Landa, and Y. Shkolnisky, “KLT picker: Particle picking using data-driven optimal templates,” *Journal of structural biology*, vol. 210, no. 2, p. 107473, 2020.
- [9] A. Heimowitz, J. Andén, and A. Singer, “APPLE picker: Automatic particle picking, a low-effort cryo-EM framework,” *Journal of structural biology*, vol. 204, no. 2, pp. 215–227, 2018.
- [10] T. Bepler, A. Morin, M. Rapp, J. Brasch, L. Shapiro, A. J. Noble, and B. Berger, “Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs,” *Nature methods*, vol. 16, no. 11, pp. 1153–1160, 2019.
- [11] D. Cheng and A. Schwartzman, “Multiple testing of local maxima for detection of peaks in random fields,” *Annals of statistics*, vol. 45, no. 2, p. 529, 2017.
- [12] T. Bendory, A. Bartesaghi, and A. Singer, “Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities,” *IEEE signal processing magazine*, vol. 37, no. 2, pp. 58–76, 2020.
- [13] A. Singer and F. J. Sigworth, “Computational methods for single-particle electron cryomicroscopy,” *Annual review of biomedical data science*, vol. 3, pp. 163–190, 2020.
- [14] J. Frank, *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*. Oxford university press, 2006.
- [15] B. P. Prasad and V. Parthasarathy, “Detection and classification of cardiovascular abnormalities using FFT based multi-objective genetic algorithm,” *Biotechnology & Biotechnological Equipment*, vol. 32, no. 1, pp. 183–193, 2018.
- [16] M. Fukunishi, D. McDuff, and N. Tsumura, “Improvements in remote video based estimation of heart rate variability using the Welch FFT method,” *Artificial Life and Robotics*, vol. 23, pp. 15–22, 2018.
- [17] M. Roth, A. Painsky, and T. Bendory, “Detecting non-overlapping signals with dynamic programming,” *Entropy*, vol. 25, no. 2, p. 250, 2023.
- [18] S. Rapuano and F. J. Harris, “An introduction to FFT and time domain windows,” *IEEE instrumentation & measurement magazine*, vol. 10, no. 6, pp. 32–44, 2007.
- [19] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *Advances in neural information processing systems*, vol. 28, 2015.

[20] A. Mahmood and S. Khan, "Correlation-coefficient-based fast template matching through partial elimination," *IEEE Transactions on image processing*, vol. 21, no. 4, pp. 2099–2108, 2011.

[21] A. Schwartzman, Y. Gavrilov, and R. J. Adler, "Multiple testing of local maxima for detection of peaks in 1D," *Annals of statistics*, vol. 39, no. 6, p. 3290, 2011.

[22] A. Eldar, K. M. Waknin, S. Davenport, T. Bendory, A. Schwartzman, and Y. Shkolnisky, "Object detection under the linear subspace model with application to cryo-EM images," *arXiv preprint arXiv:2405.00364*, 2024.

[23] M. Dadon, W. Huleihel, and T. Bendory, "Detection and recovery of hidden submatrices," *IEEE Transactions on Signal and Information Processing over Networks*, 2024.

[24] K. E. Leyton-Brown, *Resource allocation in competitive multiagent systems*. stanford university, 2003.

[25] M. H. Rothkopf, A. Pekeć, and R. M. Harstad, "Computationally manageable combinatorial auctions," *Management science*, vol. 44, no. 8, pp. 1131–1147, 1998.

[26] T. Sandholm, "Algorithm for optimal winner determination in combinatorial auctions," *Artificial intelligence*, vol. 135, no. 1-2, pp. 1–54, 2002.

[27] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, no. 9, pp. 509–517, 1975.

[28] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.

[29] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM review*, vol. 34, no. 4, pp. 561–580, 1992.

[30] A. Painsky and S. Rosset, "Exclusive row biclustering for gene expression using a combinatorial auction approach," in *2012 IEEE 12th International Conference on Data Mining*, pp. 1056–1061, IEEE, 2012.

[31] A. Painsky and S. Rosset, "Optimal set cover formulation for exclusive row biclustering of gene expression," *Journal of Computer Science and Technology*, vol. 29, no. 3, pp. 423–435, 2014.

[32] A. Ghaddar and P. Langlais, "Robust lexical features for improved neural network named-entity recognition," *arXiv preprint arXiv:1806.03489*, 2018.

[33] H. Huang, H. Xu, X. Wang, and W. Silamu, "Maximum F1-score discriminative training criterion for automatic mispronunciation detection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 4, pp. 787–797, 2015.

[34] A. Fujino, H. Isozaki, and J. Suzuki, "Multi-label text categorization with model combination based on f1-score maximization," in *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II*, 2008.

[35] A. Painsky, "Quality assessment and evaluation criteria in supervised learning," *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, pp. 171–195, 2023.

[36] A. Iudin, P. K. Korir, J. Salavert-Torres, G. J. Kleywegt, and A. Patwardhan, "EMPIAR: a public archive for raw electron microscopy image data," *Nature methods*, vol. 13, no. 5, pp. 387–388, 2016.

[37] C.-H. Lee and R. MacKinnon, "Structures of the human hcnl hyperpolarization-activated channel," *Cell*, vol. 168, no. 1, pp. 111–120, 2017.

[38] A. Dhakal, R. Gyawali, L. Wang, and J. Cheng, "CryoTransformer: a transformer model for picking protein particles from Cryo-EM micrographs," *Bioinformatics*, vol. 40, no. 3, p. btae109, 2024.