EFFICIENT BAYESIAN DETECTION OF MULTIPLE EVENTS WITH A MINIMUM-DISTANCE CONSTRAINT

Georg Kail, Franz Hlawatsch, and Clemens Novak

Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Vienna, Austria; e-mail: gkail@nt.tuwien.ac.at

ABSTRACT

We propose a Bayesian method for detecting multiple events in signals under the practically relevant assumption that successive events may not be arbitrarily close and distant events are effectively independent. Our detector has low complexity since it involves only the (Monte Carlo approximation to the) one-dimensional marginal posteriors. However, its performance is good since the metric it minimizes depends on the entire event sequence. We also describe an efficient sequential implementation of our detector that is based on a tree representation and a recursive metric computation.

Index Terms— Event detection, pulse detection, Bayesian analysis, Monte Carlo method.

1. INTRODUCTION

In many applications, the signal of interest s_k contains an unknown number of "events" that occur at or around unknown times k_i. For example, these events may be signal components (e.g., shifted pulses) [1–3] or sudden changes of certain signal parameters (e.g., signal amplitude) [4,5]. Often, the system generating the signal does not allow successive event positions k_i, k_{i+1} to be arbitrarily close to each other. This can be formulated by a *minimum-distance constraint*

$$k_{i+1} - k_i \geq d_{\min} , \qquad (1)$$

for a prescribed minimum time distance $d_{\min} \geq 2$ (note that the k_i are indexed such that $k_{i+1} > k_i$). Furthermore, in many applications, distant events are effectively independent.

In this paper, we address the minimum-distance constrained detection of the presence and temporal positions k_i of events based on a noisy version of the signal s_k. Our treatment is general in that no specific signal model or statistical structure (apart from (1) and weak dependence between distant events) is assumed. We demonstrate certain limitations of classical maximum a-posteriori (MAP) detectors, and we propose a Bayesian detector that avoids these limitations through the use of a suitable decision metric. Based on a recursive formulation of this metric and a tree representation, we develop an efficient sequential implementation of the proposed detector. Using a simple approximation, the complexity per unit time of this implementation is independent of the sequence length.

This paper is organized as follows. In Section 2, the detection problem is formulated and classical MAP detectors are reviewed. Section 3 develops the new metric and describes the proposed detector. The efficient sequential detection algorithm is presented in Section 4. Finally, numerical results are provided in Section 5.

2. DETECTION PROBLEM AND CLASSICAL SOLUTIONS

The detection problem. Let the observed discrete-time signal x_k, $k = 1, \ldots, K$ be a noisy version of some signal of interest that depends on certain event positions $k_i \in \{1, \ldots, K\}$, with $i = 1, \ldots, I$.

Both the positions k_i and their number I are unknown. We will describe the event positions k_i by the binary "event indicator" sequence b_k, $k = 1, \ldots, K$, which is 1 at all event positions (i.e., if $k = k_i$ for some i) and 0 for all other k. Let $\mathbf{x}$ and $\mathbf{b}$ denote the length-K vectors corresponding to the sequences x_k and b_k, respectively.

Given an observed realization of $\mathbf{x}$, we would like to know how many events there are and at what positions they occur, for a prescribed minimum distance $d_{\min}$ in (1). We formulate this problem as the minimum-distance constrained detection of the event indicator vector $\mathbf{b}$ from $\mathbf{x}$. Adopting a Bayesian setting, we model $\mathbf{b}$ as random with prior and posterior probability mass functions (pmf's) $p(\mathbf{b})$ and $p(\mathbf{b}|\mathbf{x})$, respectively. The detection of $\mathbf{b}$ from $\mathbf{x}$ will then be based in some way on the posterior $p(\mathbf{b}|\mathbf{x})$. We do not assume any special structure of $p(\mathbf{b}|\mathbf{x})$ beyond (1) and the condition that temporally distant b_k are only weakly dependent.

From Bayes' theorem, $p(\mathbf{b}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{b}) \, p(\mathbf{b})$. The likelihood function $p(\mathbf{x}|\mathbf{b})$ describes how the event positions k_i (characterized by $\mathbf{b}$) affect the observed signal $\mathbf{x}$; it depends on the specific signal model considered and some statistical assumptions. The prior $p(\mathbf{b})$ provides a way to incorporate the minimum-distance constraint (1). Let $\mathcal{C} \subset \{0, 1\}^K$ be the set of all $\mathbf{b}$ satisfying (1). Then, we define $p(\mathbf{b}) \propto \tilde{p}(\mathbf{b})$ for $\mathbf{b} \in \mathcal{C}$ and $p(\mathbf{b}) = 0$ otherwise, where $\tilde{p}(\mathbf{b})$ is some auxiliary pmf (e.g., independent and identically distributed). Note that setting the prior $p(\mathbf{b})$ to zero for $\mathbf{b} \notin \mathcal{C}$ also forces the posterior $p(\mathbf{b}|\mathbf{x})$ to be zero for $\mathbf{b} \notin \mathcal{C}$. Note also that the b_k are strongly dependent for distances on the order of $d_{\min}$. In fact, the minimum-distance constraint (1) postulates that if $b_k = 1$ for a given k, then $b_{k+d} = 0$ for all d such that $|d| \leq d_{\min} - 1$.

Classical MAP detectors. An optimal method for detecting the event indicator sequence $\mathbf{b}$ is the MAP sequence detector [6]

$$\hat{\mathbf{b}}_{\text{MAP}}(\mathbf{x}) \triangleq \arg \max_{\mathbf{b} \in \mathcal{C}} p(\mathbf{b}|\mathbf{x}) , \qquad (2)$$

which minimizes the sequence error probability $\text{P}\{\hat{\mathbf{b}} \neq \mathbf{b}\}$. Unfortunately, this detector is of limited practical usefulness since finding the maximum of the (typically high-dimensional) posterior $p(\mathbf{b}|\mathbf{x})$ over the (typically large) set of hypotheses $\mathcal{C}$ is often infeasible.

An alternative optimal method is the MAP component detector, also known as "maximum posterior marginal/mode (MPM) detector" (e.g., [7]), which is defined as

$$\hat{b}_{k,\text{MAP}}(\mathbf{x}) \triangleq \arg \max_{b_k \in \{0,1\}} p(b_k|\mathbf{x}) , \quad k = 1, \ldots, K . \quad (3)$$

This detector minimizes the component error probability $\text{P}\{\hat{b}_k \neq b_k\}$. However, in regarding only one marginal at a time, the MAP component detector ignores a significant part of the information contained in the joint posterior $p(\mathbf{b}|\mathbf{x})$, which may result in counter-intuitive results. In particular, consider a fixed time interval $\mathcal{K}$ of length $|\mathcal{K}| \leq d_{\min}$, and suppose that for a given observed $\mathbf{x}$, the probability that there is exactly one event in $\mathcal{K}$ is 1. Equivalently, $\sum_{k \in \mathcal{K}} p(b_k = 1|\mathbf{x}) = 1$ since the events $b_k = 1$ are mutually exclusive for $k \in \mathcal{K}$ (there may not be more than one event in $\mathcal{K}$ because of the minimum-distance constraint). Suppose further that none of the possible event positions $k \in \mathcal{K}$ is much more likely than the

others, so $p(b_k = 1|\mathbf{x}) < 1/2$ for all $k \in \mathcal{K}$. It then follows that $\hat{b}_{k,\mathrm{MAP}}(\mathbf{x}) = 0$ for all $k \in \mathcal{K}$, i.e., the MAP component detector does not detect any event in $\mathcal{K}$. This is clearly counterintuitive as the probability that there is no event in $\mathcal{K}$ is zero.

Monte-Carlo approximations. A Monte-Carlo (MC) approximation [8] often renders difficult calculations feasible (e.g., [1–5, 7]). However, we will now argue that the above MAP detectors have practical limitations even in that case. The MC approximation is based on a sample $\mathcal{S} \triangleq \{\mathbf{b}^{(m)}\}_{m=1,\ldots,M}$ consisting of M realizations $\mathbf{b}^{(m)}$ of $\mathbf{b}$ that are drawn from the joint posterior $p(\mathbf{b}|\mathbf{x})$ by means of some sampling method such as the Gibbs sampler [8]. Let $q(\mathbf{b})$ denote the relative multiplicity of a given $\mathbf{b} \in \mathcal{C}$ in $\mathcal{S}$, i.e., the number of occurrences of $\mathbf{b}$ in $\mathcal{S}$ normalized by the sample size $|\mathcal{S}| = M$. Then $q(\mathbf{b})$ converges to the joint posterior $p(\mathbf{b}|\mathbf{x})$ as M increases [8]. Similarly, let $q(b_k)$ denote the relative multiplicity of b_k, i.e., the number of realizations $\mathbf{b}$ in $\mathcal{S}$ that have the given $b_k \in \{0, 1\}$ at position k, normalized by the sample size M. Then $q(b_k)$ converges to the marginal posterior $p(b_k|\mathbf{x})$ as M increases.

The MC approximation to the MAP sequence detector $\hat{\mathbf{b}}_{\mathrm{MAP}}(\mathbf{x})$, denoted $\hat{\mathbf{b}}_{\mathcal{S}}(\mathbf{x})$, is obtained by replacing $p(\mathbf{b}|\mathbf{x})$ with $q(\mathbf{b})$ in (2). In the practically relevant case of moderate sample size $M \ll |\mathcal{C}|$, the maximization in (2) may now be feasible, since the detector chooses only among the $\mathbf{b}$ contained in $\mathcal{S}$. However, replacing $p(\mathbf{b}|\mathbf{x})$ by $q(\mathbf{b})$ is problematic, since the number of *different* $\mathbf{b}$ contained in $\mathcal{S}$ (which is at most equal to $|\mathcal{S}| = M$) is much smaller than the number $|\mathcal{C}|$ of hypotheses. This means that most $\mathbf{b} \in \mathcal{C}$ must be expected to occur only once or not at all in $\mathcal{S}$, and thus $q(\mathbf{b})$ will be quite different from the true posterior $p(\mathbf{b}|\mathbf{x})$. In particular, a small sample size may introduce artificial dependencies of components that are actually independent. This is undesired, especially since we assumed weak dependencies between distant events.

The MC version of the MAP component detector $\hat{b}_{k,\mathrm{MAP}}(\mathbf{x})$, denoted $\hat{b}_{k,\mathcal{S}}(\mathbf{x})$, is obtained by replacing $p(b_k|\mathbf{x})$ with $q(b_k)$ in (3). Here, the limited sample size is no problem since there are only two possible hypotheses for b_k. Hence, $q(b_k)$ will be a good approximation to $p(b_k|\mathbf{x})$. However, $\hat{b}_{k,\mathcal{S}}(\mathbf{x})$ inherits the basic problem of $\hat{b}_{k,\mathrm{MAP}}(\mathbf{x})$, namely, that it ignores a significant part of the information contained in the joint posterior.

In [3], we considered a MAP block detector as a compromise between the MAP sequence and component detectors. This detector detects nonoverlapping blocks of $\mathbf{b}$ separately. This is justified if $\mathbf{b}$ consists of blocks of moderate length that are effectively independent of each other. However, if this is not the case, the MAP block detector will perform poorly since, similarly to the MAP component detector, it ignores statistical dependencies.

3. THE PROPOSED DETECTOR

We now present a detector that does not suffer from the problems described above. The new detector is based on the marginal pmf $p(b_k|\mathbf{x})$. However, if calculation of $p(b_k|\mathbf{x})$ by marginalization of $p(\mathbf{b}|\mathbf{x})$ is infeasible, the MC approximation $q(b_k)$ may be used instead of $p(b_k|\mathbf{x})$ (as noted above, this approximation is accurate even for a small sample size M). If $q(b_k)$ is being used, the detector's complexity is quite low. At the same time, the detector's decision metric is based on the entire sequence, which avoids the information loss associated with separate optimization of the components b_k.

Hereafter, for notational simplicity, we write $p_k \triangleq p(b_k = 1|\mathbf{x})$ for the posterior probability that $b_k = 1$. We note that due to $b_k \in \{0, 1\}$, the posterior expectation of b_k is $\mathrm{E}\{b_k|\mathbf{x}\} = p_k$, and the number of events in an interval $[k_1, k_2]$ is $\sum_{k=k_1}^{k_2} b_k$. We will also consider the *expected number of events* in $[k_1, k_2]$ given $\mathbf{x}$:

$$Q_{k_1,k_2} \triangleq \mathrm{E}\left\{\sum_{k=k_1}^{k_2} b_k \,\middle|\, \mathbf{x}\right\} = \sum_{k=k_1}^{k_2} \mathrm{E}\{b_k|\mathbf{x}\} = \sum_{k=k_1}^{k_2} p_k.$$

Note that p_k and Q_{k_1,k_2} depend on $\mathbf{x}$. To avoid double counting at interval borders, we will use the modified quantity

$$Q'_{k_1,k_2} \triangleq Q_{k_1,k_2} - \frac{p_{k_1}}{2} - \frac{p_{k_2}}{2}, \quad 1 \le k_1 \le k_2 \le K,$$

complemented by the left-end and right-end definitions

$$Q'_{0,k} \triangleq Q_{1,k} - \frac{p_k}{2}, \qquad Q'_{k,K+1} \triangleq Q_{k,K} - \frac{p_k}{2}.$$

The new detector is motivated by the following four desiderata.

D1 Within any interval $[k_1, k_2]$ for which Q'_{k_1,k_2} is close to 1, one event should be detected.

D2 The detected events should be at positions k where p_k is large.

D3 The detected sequence of events should comply with the minimum-distance constraint (1).

D4 The detected events should be at positions k where p_k is not smaller than at the adjacent positions $k-1$ and $k+1$.

These desiderata are reflected either in the decision metric or by a constraint, as explained next.

D1. Suppose that a given detector detects N events at positions l_n, $n = 1, \ldots, N$ (hereafter briefly called "detected positions"). Then, based on the cumulative structure of Q_{k_1,k_2}, the following fact can be verified: If there is roughly one detected event within any interval $[k_1, k_2]$ such that $Q'_{k_1,k_2} \approx 1$, then $Q'_{l_{n-1},l_n} \approx 1$ for all $n \in [2, N]$, and vice versa.[1] Thus, D1 can be rephrased as follows: *Between any two successive detected positions l_{n-1}, l_n, the expected number of events Q'_{l_{n-1},l_n} should be close to* 1. The deviation of Q'_{l_{n-1},l_n} from 1 will be measured by $\sum_n |Q'_{l_{n-1},l_n} - 1|$. Thus, D1 corresponds to choosing the detected $\mathbf{b}$ (or, equivalently, N and the l_n) such that the following metric component is minimized:

$$w_1(\mathbf{b}) \triangleq \sum_{n=2}^{N} |Q'_{l_{n-1},l_n} - 1| + w_1^{(\mathrm{L})}(l_1) + w_1^{(\mathrm{R})}(l_N).$$

Here, the left-end term $w_1^{(\mathrm{L})}(l_1)$ and the right-end term $w_1^{(\mathrm{R})}(l_N)$ express our desire that the expected numbers of events to the left of l_1 and to the right of l_N are both close to *or less than* 1:

$$w_1^{(\mathrm{L})}(l_1) \triangleq [Q'_{0,l_1} - 1]^+, \quad w_1^{(\mathrm{R})}(l_N) \triangleq [Q'_{l_N,K+1} - 1]^+, \quad (4)$$

where $[\cdot]^+$ means that a negative value is replaced by zero. We note that D1—i.e., minimization of $w_1(\mathbf{b})$ alone—would essentially define the *relative* positions of the detected events with respect to each other. Fixing one detected event position, e.g. l_1, determines all the others. However, the value of l_1 obtained by minimizing $w_1(\mathbf{b})$ is rather arbitrary within some set of almost equivalent solutions.

D2. The *absolute* position of the detected events is strongly influenced by D2. We will formulate D2 as the maximization of $\sum_{n=1}^{N} p_{l_n}$ or, equivalently, minimization of the metric component

$$w_2(\mathbf{b}) \triangleq \sum_{n=1}^{N} (1 - p_{l_n}).$$

D3. Rather than enforcing D3, i.e., $\mathbf{b} \in \mathcal{C}$ by explicitly restricting the set of hypotheses $\mathbf{b}$, we will use the metric component (cf. [9])

$$w_3(\mathbf{b}) \triangleq \begin{cases} 0, & \mathbf{b} \in \mathcal{C} \\ \infty, & \text{otherwise.} \end{cases}$$

---

[1] Our formulation of this equivalence is fuzzy only because of the discrete nature of k. Replacing k by a continuous variable and, consequently, Q' by an integral would make the equivalence exact.

Since ∞ overrides the other metric components, all hypotheses $\mathbf{b} \notin \mathcal{C}$ are excluded. We can reformulate $w_3(\mathbf{b})$ in terms of the l_n as

$$w_3(\mathbf{b}) = \sum_{n=2}^{N} \tilde{w}_3(l_{n-1}, l_n),$$

where $\tilde{w}_3(k_1, k_2) \triangleq 0$ for $|k_1 - k_2| \geq d_{\min}$ and $\tilde{w}_3(k_1, k_2) \triangleq \infty$ otherwise.

D4. To reduce the number of hypotheses to be compared, D4 is formulated as a hard constraint $\mathbf{b} \in \mathcal{W}$ rather than via the metric. Let $\tilde{\mathcal{K}} = \{\tilde{k}_j\}_{j=1,\ldots,J}$ denote the set of all k such that $p_k \geq p_{k-1}$ and $p_k \geq p_{k+1}$. Then, $\mathcal{W}$ is defined as the set of all hypotheses $\mathbf{b}$ such that all detected positions l_n are contained in $\tilde{\mathcal{K}}$. This reduces the number of hypotheses from 2^K to 2^J. However, the actual constraint set (also taking into account D3) is $\mathcal{W} \cap \mathcal{C}$, and thus still smaller. This set is nonempty, since it always contains the all-zero sequence $\mathbf{b} = \mathbf{0}$.

Definition of the detector. Putting together D1 through D4, the proposed detector is finally defined via the constrained minimization

$$\hat{\mathbf{b}}(\mathbf{x}) \triangleq \arg\min_{\mathbf{b} \in \mathcal{W}} \{w(\mathbf{b})\}, \qquad (5)$$

with

$$\begin{aligned} w(\mathbf{b}) &\triangleq w_1(\mathbf{b}) + \gamma w_2(\mathbf{b}) + w_3(\mathbf{b}) \\ &= \sum_{n=2}^{N} |Q'_{l_{n-1}, l_n} - 1| + w_1^{(\mathrm{L})}(l_1) + w_1^{(\mathrm{R})}(l_N) \\ &\quad + \gamma \sum_{n=1}^{N} (1 - p_{l_n}) + \sum_{n=2}^{N} \tilde{w}_3(l_{n-1}, l_n). \end{aligned} \qquad (6)$$

Here, γ is a positive weight emphasizing D1 or D2 (recall that D3 and D4 are strict conditions).

4. EFFICIENT SEQUENTIAL IMPLEMENTATION

In (5), $w(\mathbf{b})$ is minimized over $\mathcal{W}$ or, equivalently, with respect to N and the position set $\{l_n\}_{n=1,\ldots,N} \subseteq \tilde{\mathcal{K}} = \{\tilde{k}_j\}_{j=1,\ldots,J}$. At each of the J admissible positions $\tilde{k}_1, \ldots, \tilde{k}_J$, b_k may be 0 or 1. Thus, the constraint set $\mathcal{W}$ consists of 2^J sequences $\mathbf{b}$, which, typically, is still too much. Fortunately, the structure of the decision metric $w(\mathbf{b})$ in (6) allows for an efficient implementation that is sequential in $\tilde{k}_j$.

Tree representation. In Fig. 1, $\mathcal{W}$ is represented by a binary tree whose levels (except the root level) correspond to the admissible positions $\tilde{k}_j$. Thus, there are $J + 1$ levels including the root. We assume that the $\tilde{k}_j$ are indexed in increasing order, i.e., $\tilde{k}_{j+1} > \tilde{k}_j$, and displayed from left to right. The root and the leaves correspond to levels $j = 0$ and $j = J$, respectively. The two branches leaving a node at level j correspond to the binary decision $b_{\tilde{k}_{j+1}} = 0$ (upward branch) or $b_{\tilde{k}_{j+1}} = 1$ (downward branch). At level j, half of the 2^j nodes correspond to $b_{\tilde{k}_j} = 0$, the other half to $b_{\tilde{k}_j} = 1$; we will refer to them as 0-nodes and 1-nodes, respectively. Each of the 2^J leaves (equivalently, paths from root to leaf) corresponds to one $\mathbf{b} \in \mathcal{W}$.

Pruning the tree. The tree can be pruned, based on a left-right decomposition of the metric $w(\mathbf{b})$ at 1-nodes. Each 1-node corresponds to a detected position. Consider a fixed 1-node $\mathcal{N}$ at level j, corresponding to the detected position $l_n = \tilde{k}_j$. Let $\mathbf{b}$ with $b_{\tilde{k}_j} = 1$ correspond to a path that passes through $\mathcal{N}$. This path is uniquely defined to the left of $\mathcal{N}$ (from root to $\mathcal{N}$), but to the right of $\mathcal{N}$ it may be any of the path segments connecting $\mathcal{N}$ to one of the leaves. These path segments constitute a subtree that originates from $\mathcal{N}$. From (6), it follows that the metric of $\mathbf{b}$ can be decomposed as
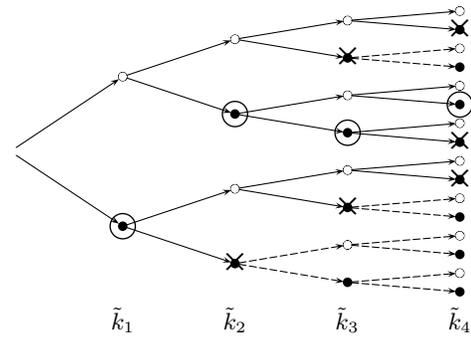

Fig. 1. Tree representation of the constraint set $\mathcal{W}$ associated with D4 (example). White and black bullets denote 0-nodes and 1-nodes, respectively. At each level, only one 1-node survives (large circle). Dashed branches belong to discarded subtrees.

$$w(\mathbf{b}) = w_{\mathrm{L}}(l_1, \ldots, l_n) + w_{\mathrm{R}}(l_n, \ldots, l_N), \qquad (7)$$

with the "left partial metric" $w_{\mathrm{L}}(l_1, \ldots, l_n) \triangleq \sum_{\nu=2}^{n} |Q'_{l_{\nu-1}, l_\nu} - 1| + w_1^{(\mathrm{L})}(l_1) + \gamma \sum_{\nu=1}^{n}(1 - p_{l_\nu}) + \sum_{\nu=2}^{n} \tilde{w}_3(l_{\nu-1}, l_\nu)$ and the "right partial metric" $w_{\mathrm{R}}(l_n, \ldots, l_N) \triangleq \sum_{\nu=n+1}^{N} |Q'_{l_{\nu-1}, l_\nu} - 1| + w_1^{(\mathrm{R})}(l_N) + \gamma \sum_{\nu=n+1}^{N}(1 - p_{l_\nu}) + \sum_{\nu=n+1}^{N} \tilde{w}_3(l_{\nu-1}, l_\nu)$.

Different 1-nodes $\mathcal{N}_m$ at the same level j correspond to the same detected position $l_{n_m} = \tilde{k}_j$ (note that the index of l depends on the 1-node), but the paths passing through these 1-nodes are different. The left partial metric (LPM) $w_{\mathrm{L}}(l_1, \ldots, l_{n_m})$ is generally different for these 1-nodes. However, the set of possible values for the right partial metric (RPM) $w_{\mathrm{R}}(l_{n_m}, \ldots, l_N)$ is the same, since the subtrees originating from the 1-nodes are equivalent. Because of (7), this means that only the 1-node with the smallest LPM $w_{\mathrm{L}}(l_1, \ldots, l_{n_m})$ can be part of the minimum-metric path, and thus all the other 1-nodes at level j and the entire subtrees originating from them can be discarded. (A similar consideration does not apply to the 0-nodes because these do not correspond to a detected position l_n, and hence no LPMs and RPMs can be associated with them.) The resulting pruning procedure yields a dramatic reduction of the number of paths (see Fig. 1). Before pruning, level j contains twice as many nodes as level $j - 1$. Out of the 1-nodes, which are half of them, only one is retained after pruning. Thus, passing to the next level, the number of left path segments to be retained is increased by 1, rather than doubled. There are two nodes at level $j = 1$, and therefore, at level j, there are $j + 1$ surviving nodes (or left path segments), namely, j 0-nodes and one 1-node. At the final level $j = J$, the number of surviving leaves (or paths, or hypotheses $\mathbf{b}$) is only $J + 1$.

Sequential operation. At each level, we have to compute the LPM $w_{\mathrm{L}}(l_1, \ldots, l_n)$ for each 1-node. A sequential processing is possible because $w_{\mathrm{L}}(l_1, \ldots, l_n)$ can be computed recursively according to

$$w_{\mathrm{L}}(l_1, \ldots, l_n) = w_{\mathrm{L}}(l_1, \ldots, l_{n-1}) + \Delta w_n, \qquad (8)$$

with $\Delta w_n \triangleq |Q'_{l_{n-1}, l_n} - 1| + \gamma(1 - p_{l_n}) + \tilde{w}_3(l_{n-1}, l_n)$. This recursion is initialized by $w_{\mathrm{L}}(l_1) = w_1^{(\mathrm{L})}(l_1) + \gamma(1 - p_{l_1})$. At level j, we update the LPMs of the j 1-nodes according to (8) and identify the surviving 1-node with minimum LPM. Processing the entire tree requires $\sum_{j=1}^{J} j = J(J+1)/2$ such updates. At the final level J, we calculate the RPMs for the $J + 1$ leaves and add them to the LPMs. Since there are no detected positions to the right of $\tilde{k}_J$, these RPMs reduce to the respective values of $w_1^{(\mathrm{R})}(l_N)$.

We have thus calculated the final metrics $w(\mathbf{b})$ of all $J + 1$ paths (or sequences $\mathbf{b}$) that still qualify for the minimum-metric path; these paths constitute a significantly reduced subset of the constraint set $\mathcal{W}$. Finally, we choose the detection result $\hat{\mathbf{b}}(\mathbf{x})$ as the $\mathbf{b}$ with min-