axioms

Article

Cognitively Economical Heuristic for Multiple Sequence Alignment under Uncertainties

Milan Gnjatović [1,*], Nemanja Maček [2], Muzafer Saračević [3], Saša Adamović [4], Dušan Joksimović [1] and Darjan Karabašević [5]

[1] Department of Information Technology, University of Criminal Investigation and Police Studies, Cara Dušana 196, 11080 Beograd, Serbia
[2] School of Electrical and Computer Engineering, Academy of Technical and Art Applied Studies, Vojvode Stepe 283, 11000 Beograd, Serbia
[3] Department of Computer Sciences, University of Novi Pazar, Dimitrija Tucovića bb., 36300 Novi Pazar, Serbia
[4] Faculty of Informatics and Computing, Singidunum University, Danijelova 32, 11000 Beograd, Serbia
[5] Faculty of Applied Management, Economics and Finance, University Business Academy in Novi Sad, Jevrejska 24, 11000 Belgrade, Serbia
* Correspondence: milan.gnjatovic@kpu.edu.rs

Abstract: This paper introduces a heuristic for multiple sequence alignment aimed at improving real-time object recognition in short video streams with uncertainties. It builds upon the idea of the progressive alignment but is cognitively economical to the extent that the underlying edit distance approach is adapted to account for human working memory limitations. Thus, the proposed heuristic procedure has a reduced computational complexity compared to optimal multiple sequence alignment. On the other hand, its relevance was experimentally confirmed. An extrinsic evaluation conducted in real-life settings demonstrated a significant improvement in number recognition accuracy in short video streams under uncertainties caused by noise and incompleteness. The second line of evaluation demonstrated that the proposed heuristic outperforms humans in the post-processing of recognition hypotheses. This indicates that it may be combined with state-of-the-art machine learning approaches, which are typically not tailored to the task of object sequence recognition from a limited number of frames of incomplete data recorded in a dynamic scene situation.

Keywords: multiple sequence alignment; object recognition; uncertainty in vision task; cognitive economy

check for updates

Citation: Gnjatović, M.; Maček, N.; Saračević, M.; Adamović, S.; Joksimović, D.; Karabašević, D. Cognitively Economical Heuristic for Multiple Sequence Alignment under Uncertainties. *Axioms* **2023**, *12*, 3. https://doi.org/10.3390/axioms12010003

Academic Editor: Abbe Mowshowitz

Received: 24 October 2022
Revised: 7 December 2022
Accepted: 16 December 2022
Published: 21 December 2022

1. Introduction

The sources of noise and incompleteness in video streams are manifold and diverse. Captured objects may be non-uniformly illuminated, physically damaged, obscured by dirt or dust, etc. [1]. The human operator capturing a video stream may be negligent, physically affected (e.g., suffering from a tremor), or working under time constraints, which in turn reduces the number of quality image frames, etc. Thus, handling uncertainty caused by noise and incompleteness in video streams represents an important research task. This paper addresses a particular aspect of this research question—it introduces a cognitively economical heuristic for multiple sequence alignment aimed at improving real-time object recognition in short video streams with uncertainties.

Machine learning approaches have already been recognized to be able to outperform human observers in visual recognition tasks with static frame input involving low signal-to-noise ratio (e.g., noise robust convolutional neural networks for image classification [2–4], etc.). However, these approaches are not necessarily tailored to the task of object sequence recognition from a limited number of frames of incomplete data recorded in a dynamic scene situation. One way to overcome this problem is to introduce a pre-processing step devoted to image reconstruction from incomplete frames (cf. [5]). In contrast to this, the heuristic proposed in this paper is based on post-processing of the recognition hypotheses and allows for avoiding time-consuming image reconstruction.

One of the assumptions underlying this heuristic procedure is that there is an object recognition system that is independent and agnostic of the proposed approach. For the purpose of easier representation, let S be a set of object classes that can be detected and recognized by the given system. Without a loss of generality, the result of processing a single image frame is a recognition hypothesis that can be represented as follows:

$$\begin{aligned} h_i &\equiv (s_i, c_i) \\ &\equiv (s_i[0], c_i[0]), (s_i[1], c_i[1]), \ldots, (s_i[m-1], c_i[m-1]) \;, \end{aligned} \tag{1}$$

where

- m is a nonnegative integer ($m \in \mathbb{N}_0$),
- sequence s_i represents recognized objects, i.e.,

  $$(\forall\, 0 \leq k < m)(s_i[k] \in S) \;, \tag{2}$$

  and the order of elements in s_i is determined by the spatial order of recognized objects in the image reference system,
- sequence c_i contains the corresponding real-valued recognition confidences for objects in s_i.

For example, recognition hypothesis

$$h_i \equiv (s_i, c_i) \equiv (4, 0.931), (7, 0.834), (7, 0.877) \tag{3}$$

can be interpreted as follows: the sequence of recognized objects contains three digits, 4, 7, and 7 (i.e., sequence s_i), and their recognition confidences are 0.931, 0.834, and 0.877, respectively, (i.e., sequence c_i).

The second assumption is that multiple image frames are captured for each given spatial scene; i.e., multiple recognition attempts are performed, each of which generates a recognition hypothesis as described in Equation (1). For example, Figure 1 shows a set of image frame segments derived from a video stream captured by a mobile phone application. Each frame is processed separately by the external number recognition system introduced in [1,6]. In Table 1, for each frame, a recognition hypothesis is provided.

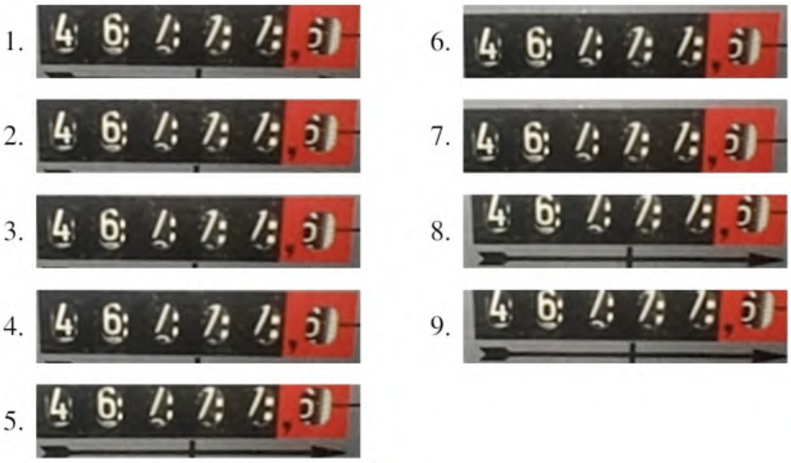


Figure 1. A set of image frame segments derived from a video stream captured by a mobile phone application. The size of fragments is 230 × 52 pixels (with 300 pixels/inch). For the purpose of presentation, images are scaled up.

Table 1. For each frame given in Figure 1, a recognition hypothesis is provided. The digit after the decimal point is intentionally discarded. The recognition confidences are normalized to the range $[0, 1]$.

Frame Recognition Hypotheses
$h_1 \equiv (s_1, c_1) \equiv (4, 0.931), (7, 0.834), (7, 0.877)$
$h_2 \equiv (s_2, c_2) \equiv (4, 0.933), (6, 0.883), (7, 0.828), (7, 0.827)$
$h_3 \equiv (s_3, c_3) \equiv (4, 0.907), (6, 0.880), (7, 0.829), (7, 0.840)$
$h_4 \equiv (s_4, c_4) \equiv (4, 0.928), (6, 0.875), (7, 0.843), (7, 0.886)$
$h_5 \equiv (s_5, c_5) \equiv (4, 0.921), (6, 0.883), (7, 0.851), (7, 0.791), (8, 0.640), (7, 0.781)$
$h_6 \equiv (s_6, c_6) \equiv (4, 0.909), (6, 0.869), (7, 0.836), (7, 0.830)$
$h_7 \equiv (s_7, c_7) \equiv (4, 0.907), (6, 0.861), (7, 0.833), (7, 0.882)$
$h_8 \equiv (s_8, c_8) \equiv (4, 0.881), (8, 0.838), (7, 0.809), (7, 0.846), (8, 0.651), (7, 0.819)$
$h_9 \equiv (s_9, c_9) \equiv (4, 0.891), (7, 0.813), (7, 0.860)$

It is the task of the heuristic proposed in this paper to post-process a set of recognition hypotheses and derive a single recognition result. To demonstrate the applicability of the proposed heuristic, it is extrinsically evaluated in a specific, real-life scenario of the automatic reading of electricity meters. The prototype system for multiple sequence alignment based on the proposed heuristic is engaged to post-process recognition hypotheses obtained from the external number-recognition subsystem (introduced in [1,6]). However, it should be noted that the prototype system for multiple sequence alignment is completely agnostic and independent of the underlying number recognition subsystem. This is in line with our commitment to introduce a heuristic that is not intended for a particular object recognition system or a specific recognition task, but rather for addressing the more general problem of the real-time recognition of spatially ordered objects in short video streams. An additional characteristic of our approach is that it is cognitively economical [7] and has a reduced computational complexity in comparison to optimal multiple sequence alignment.

The rest of this paper is organized as follows. Section 2 provides an overview of background and related work. The heuristic procedure is introduced in Section 3, and evaluated and discussed in Section 4. Section 5 concludes the paper.

2. Background and Related Work

Uncertainty management in cognitive agents is an important research question in the field of artificial intelligence [8,9]. This paper focuses on a particular aspect of this question, i.e., uncertainties in automatic object recognition in short video streams and addresses this research problem by means of multiple sequence alignment.

Under the alignment of two sequences s_i and s_j over alphabet S, it is usually meant that the sequences are modified by adding spaces, so that the resulting sequences $\hat{s_i}$ and $\hat{s_j}$ are of equal length L. The value of the alignment is defined as

$$\sum_{k=0}^{L-1} \text{score}(\hat{s_i}[k], \hat{s_j}[k]) \ , \tag{4}$$

where $\text{score}(\hat{s_i}[k], \hat{s_j}[k])$ is the score of two opposing symbols at position k in sequences $\hat{s_i}$ and $\hat{s_j}$, respectively. An optimal alignment minimizes the value given in Equation (4) [10].

One of the basic algorithms for finding the optimal alignment of two sequences is related to the widely acknowledged edit distance (i.e., the Levenshtein distance) [11–13]. The standard minimum edit distance between two sequences is defined as the minimum number of single-symbol edit operations (i.e., deletion of a symbol, insertion of a symbol, and replacement of a symbol by another symbol) required to transform one sequence into the other. In a more general case, the edit operations are weighted, and the calculation of the minimum edit distance can be described as follows. Let s_i and s_j be two sequences of lengths m and n, respectively, over alphabet S:

$$\begin{aligned} s_i &\equiv s_i[0], s_i[1], \ldots, s_i[m-1] \ , \\ s_j &\equiv s_j[0], s_j[1], \ldots, s_j[n-1] \ . \end{aligned} \tag{5}$$

To align these sequences, a distance matrix D of dimension $(m+1) \times (n+1)$ is generated. The symbol with index k in the first sequence, i.e., $s_i[k]$, is assigned the row of matrix D with index $(k+1)$, while the symbol with index l in the second sequence, i.e., $s_j[l]$, is assigned the column of matrix D with index $(l+1)$. The first row and the first column of matrix D are calculated as:

$$\begin{aligned} D[0,0] &= 0 \ , \\ D[k,0] &= D[k-1,0] + d(s_i[k-1]) \ , \\ D[0,l] &= D[0,l-1] + i(s_j[l-1]) \ , \end{aligned} \tag{6}$$

and the rest of matrix as:

$$D[k,l] = min \left\{ \begin{array}{l} D[k-1,l] + d(s_i[k-1]) \\ D[k,l-1] + i(s_j[l-1]) \\ D[k-1,l-1] + r(s_i[k-1], s_j[l-1]) \end{array} \right\} \tag{7}$$

where $1 \le k \le m$, $1 \le l \le n$ and

- $d(s_i[k-1])$ is the cost of deletion of symbol $s_i[k-1]$,
- $i(s_j[l-1])$ is the cost of insertion of symbol $s_j[l-1]$,
- $r(s_i[k-1], s_j[l-1])$ is the cost of replacement of symbol $s_i[k-1]$ by symbol $s_j[l-1]$.

The minimum edit distance between sequences s_i and s_j is equal to the bottom right cell of matrix D, i.e., $D[m,n]$. For example, if we set the costs of all single-symbol edit operations to one, i.e.,

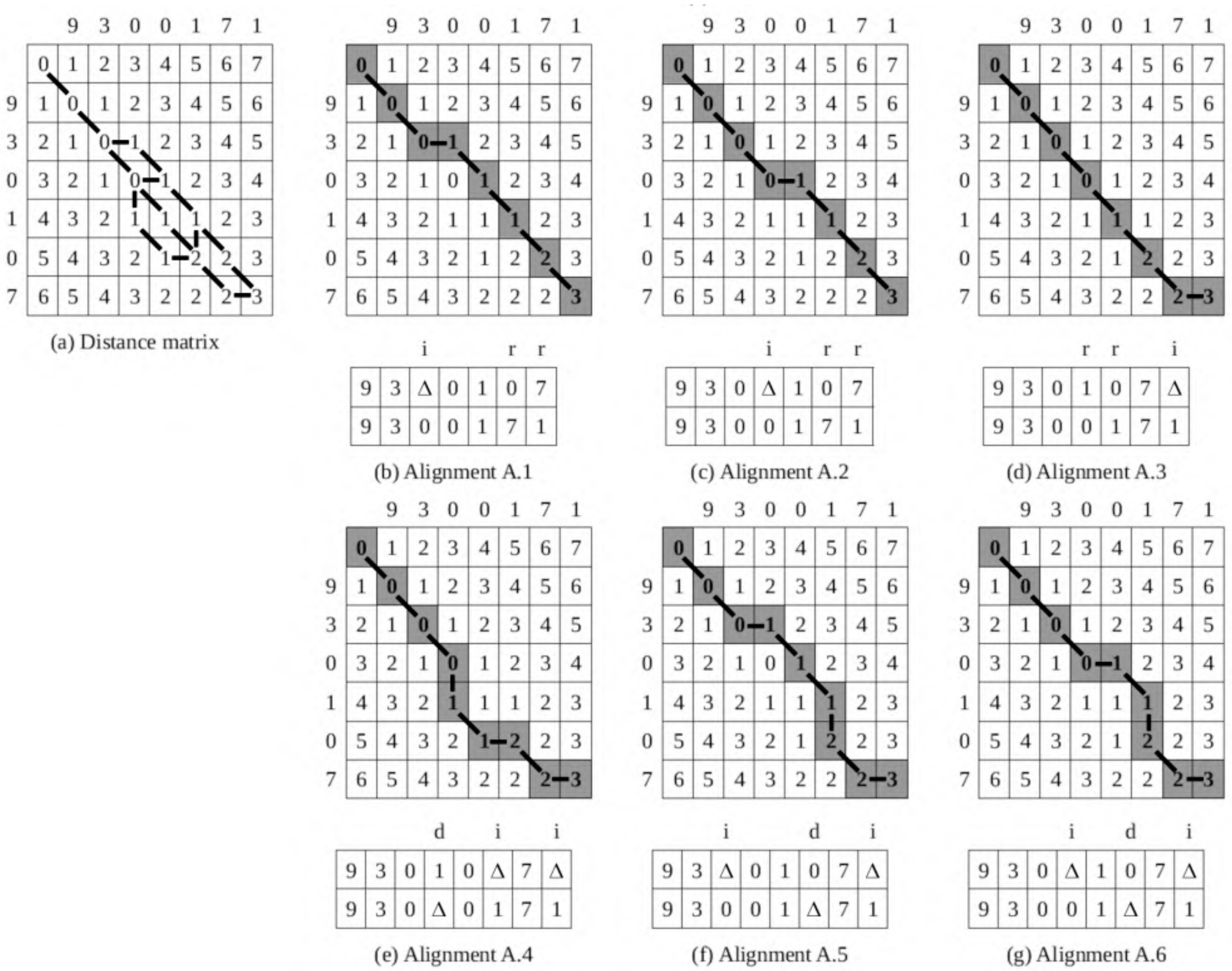
$$\begin{aligned} &(\forall \ \sigma \in S) \ (d(\sigma) = 1) \ , \\ &(\forall \ \sigma \in S) \ (i(\sigma) = 1) \ , \\ &(\forall \ \sigma_1, \sigma_2 \in S) \ (r(\sigma_1, \sigma_2) = \begin{cases} 0, & \text{if } \sigma_1 = \sigma_2 \\ 1, & \text{otherwise} \end{cases} \ . \end{aligned} \tag{8}$$

the minimum edit distance between sequences

$$s_1 = 930{,}107 \text{ and } s_2 = 9{,}300{,}171 \tag{9}$$

is equal to 3; i.e., three single-symbol edit operations are required to transform one sequence into the other. The underlying distance matrix is given in Figure 2a.

However, to find all optimal alignments between two sequences, it is necessary to backtrace from cell $D[m,n]$ to cell $D[0,0]$. For each cell in D (except cell $D[0,0]$), it is necessary to keep track of which matrix cell participated in the calculation of the value of the given cell (in accordance with Equations (5) and (6)). Each path starting at cell $D[0,0]$ and ending at cell $D[m,n]$ represents an optimal alignment. To continue the previous example, there are six possible minimum-cost paths in matrix D, i.e., six optimal alignments between sequences 930,107 and 9,300,171, as shown in Figure 2b–g. The minimum-cost paths are marked with a gray background. Horizontal edges in a path determine insertions, vertical edges determine deletions, and diagonal edges determine replacements. The alignments are given under each corresponding matrix. Single-symbol edit operations of insertion, deletion, and replacement are, respectively, denoted by letters i, d, and r. Spaces added to sequences are represented by special symbol $\triangle$ (i.e., a gap) not belonging to alphabet S.



		9	3	0	0	1	7	1
	0	1	2	3	4	5	6	7
9	1	0	1	2	3	4	5	6
3	2	1	0	1	2	3	4	5
0	3	2	1	0	1	2	3	4
1	4	3	2	1	1	1	2	3
0	5	4	3	2	1	2	2	3
7	6	5	4	3	2	2	2	3

(a) Distance matrix

		i			r	r
9	3	Δ	0	1	0	7
9	3	0	0	1	7	1

(b) Alignment A.1

			i		r	r
9	3	0	Δ	1	0	7
9	3	0	0	1	7	1

(c) Alignment A.2

			r	r		i
9	3	0	1	0	7	Δ
9	3	0	0	1	7	1

(d) Alignment A.3

			d		i		i
9	3	0	1	0	Δ	7	Δ
9	3	0	Δ	0	1	7	1

(e) Alignment A.4

		i			d		i
9	3	Δ	0	1	0	7	Δ
9	3	0	0	1	Δ	7	1

(f) Alignment A.5

			i		d		i
9	3	0	Δ	1	0	7	Δ
9	3	0	0	1	Δ	7	1

(g) Alignment A.6

Figure 2. Illustration of the minimum edit distance algorithm. The costs of all single-symbol edit operations are equal to one, cf. Equation (8). Abbreviations: i—insertion, d—deletion, r—replacement.

This edit distance algorithm can be generalized to multiple sequence alignment [14]. The basic score scheme applied is the SP-score, which considers the scores of all unordered pairs of opposing symbols at position k in all aligned sequences [10]. The sequence alignment is widely applied in computational biology (e.g., aligning protein sequences, cf. [15,16]), computational linguistics (e.g., spell correction, speech recognition, machine translation, information extraction, cf. [11]), and information security (e.g., impersonation attacks detection in cloud computing environments, cf. [17]), but also in image processing (e.g., scene detection in videos, online signature verification, cf. [18–21]). However, there is a practical limitation related to the fact that the problem of finding an optimal multiple sequence alignment is NP-hard [22]. The complexity of an optimal alignment of p sequences of average length n, based on the dynamic programming approach described above, is $O(n^p)$.

To address the question of efficiency, a variety of heuristics were introduced [16,23] (cf. also [24]). A popular heuristic in the field of computational biology is so-called progressive alignment [16,25]. It works by first performing optimal pairwise sequence alignments and then clustering the sequences, e.g., by applying the mBed or k-means algorithms. In this paper, we introduce a novel cognitively economical heuristic procedure for multiple sequence alignment that builds upon the idea of the progressive alignment. At the methodological level, the novel aspects are as follows:

(i) The number of clusters is determined prior to the pairwise sequence alignment. Each distinct maximum-length sequence in a set of recognition hypotheses is declared as a cluster representative. All other non-maximum-length sequences were then assigned to the closest cluster representative, where the distance between two sequences was calculated by means of the adapted edit distance algorithm.

(ii) The proposed adaptation of the edit distance approach is inspired by human working memory limitations (cf. [26]). To reduce the “cognitive load” of our approach, we consider only “economical” sequence alignments that are optimal in terms of the standard minimum edit distance approach and in which no space is inserted into longer sequence. These two requirements for cognitive economy allow for the substantial reduction of the number of sequences derived in the alignment process by means of padding (as detailed in Section 3).

3. Heuristic for Multiple Sequence Alignment

In this section, we introduce a heuristic approach to multiple sequence alignment intended for improving real-time object recognition in short video streams under uncertainties. It includes two algorithms:

- The gap-minimum alignment algorithm, introduced and illustrated in Section 3.1, is intended for alignment of two recognition hypotheses.
- The cluster-based voting algorithm, introduced and illustrated in Section 3.2, builds upon the first algorithm and is intended for multiple recognition hypothesis alignment, based on which a single recognition result is derived.

3.1. Gap-Minimum Two Sequence Alignment

Let h_i and h_j be two recognition hypotheses, as described in Equation (1), of lengths m and n, respectively:

$$\begin{aligned} h_i &\equiv (s_i, c_i) \\ &\equiv (s_i[0], c_i[0]), (s_i[1], c_i[1]), \ldots, (s_i[m-1], c_i[m-1]) \; , \\ h_j &\equiv (s_j, c_j) \\ &\equiv (s_j[0], c_j[0]), (s_j[1], c_j[1]), \ldots, (s_j[n-1], c_j[n-1]) \; , \end{aligned} \tag{10}$$

where s_i and s_j are sequences over alphabet S, and c_i and c_j are sequences of the corresponding recognition confidence values. The cognitively economical idea underlying the proposed two-sequence-alignment approach is that we consider only sequence alignments that are optimal in terms of the standard minimum edit distance approach in which no space is inserted into the longer sequence. Thus, when two sequences of unequal lengths are aligned, the longer sequence always remains unchanged, while $|m - n|$ spaces are “economically” inserted into the shorter sequence. The algorithm can be described as follows.

Step 1.1: If sequences s_i and s_j are of equal length, i.e., $m = n$, then no particular alignment is performed, i.e.:

$$(\forall\, 0 \leq k < m)(s_i[k] \text{ is opposed to } s_j[k]) \; , \tag{11}$$

and the alignment process is terminated.

Step 1.2: Otherwise, if sequences s_i and s_j are not of equal length, let us assume, without loss of generality, that the length of s_i is less than the length of s_j, i.e., $m < n$. A distance matrix is generated, with the costs of all edit operations set to one, as described in Section 2. Let P be a set of all optimal alignments of sequences s_i and s_j derived from the distance matrix. We recall that all alignments in set P are determined by means of the minimal edit distance algorithm (i.e., the Levenshtein algorithm), and thus they contain a minimal number of single-symbol edit operations (i.e., deletion, insertion, and replacement) required to transform sequence s_i into sequence s_j. In general, it is easy to show that set P is never empty (i.e., it is always possible to find at least one alignment).

Step 1.3: From set P, containing all optimal alignments of sequences s_i and s_j, we select only those alignments in which no space is inserted into longer a sequence. Let $P_r \subseteq P$ be a set of selected alignments. If $P_r \neq \emptyset$, it is declared that sequences s_i and s_j cannot be economically aligned, and the alignment process is terminated. Otherwise, the algorithm proceeds to the next step.

Step 1.4: The value of each alignment $p \in P_r$ is calculated as the sum of confidence values of all symbols in a longer sequence s_j that are opposed to a space, i.e.:

$$v(p) = v((\hat{s}_i, \hat{c}_i), (s_j, c_j)) = \sum_{l=0}^{n-1} \text{score}(s_j[l]) \ , \tag{12}$$

where:

$$\text{score}(s_j[l]) = \begin{cases} c_j[l], & \text{if } s_j[l] \text{ is opposed to } \triangle \ , \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

The alignment in P_r with a minimum value is selected as the most cognitively economical alignment:

$$\hat{p} = \underset{p \in P_r}{\text{argmin}} \ v(p) \ . \tag{14}$$

The proposed gap-minimum two-sequence alignment algorithm is illustrated by the following examples.

Example 1. *Let us consider the alignment of the following recognition hypotheses:*

$$\begin{aligned} h_i \equiv &(s_i, c_i) \\ \equiv &(9, 0.822), (3, 0.765), (0, 0.746), (1, 0.815), \\ &(0, 0.831), (7, 0.672) \ , \\ h_j \equiv &(s_j, c_j) \\ \equiv &(9, 0.866), (3, 0.815), (0, 0.854), (0, 0.814), \\ &(1, 0.753), (7, 0.829), (1, 0.786) \ , \end{aligned} \tag{15}$$

Set P, generated in Step 1.2, contains six alignments, i.e., A.1–A.6, shown in Figure 2b–g. Set P_r, generated in Step 1.3, contains only three cognitively economical alignments (A.1–A.3) that satisfy the condition that no space is inserted into longer sequence s_j. The values of these alignments, calculated in Step 1.4 based on confidence values provided in Equation (15), are:

$$\begin{aligned} v(A.1) &= c_j[2] = 0.854 \ , \\ v(A.2) &= c_j[3] = 0.814 \ , \\ v(A.3) &= c_j[6] = 0.786 \ . \end{aligned} \tag{16}$$

The alignment A.3 has the minimum value and thus represents the most cognitively economical alignment:

$$\begin{aligned} \hat{h}_i \equiv &(\hat{s}_i, \hat{c}_i) \\ \equiv &(9, 0.822), (3, 0.765), (0, 0.746), (1, 0.815), \\ &(0, 0.831), (7, 0.672), (\triangle, 0.5) \ , \\ h_j \equiv &(s_j, c_j) \\ \equiv &(9, 0.866), (3, 0.815), (0, 0.854), (0, 0.814), \\ &(1, 0.753), (7, 0.829), (1, 0.786) \ , \end{aligned} \tag{17}$$

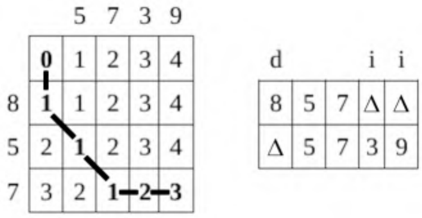
In the selected alignment, an initially longer recognition hypothesis h_j remains unchanged, while initially, the shorter recognition hypothesis h_i was transformed to $\hat{h}_i$ by inserting a space at the end of sequence s_i.

It should be noted that the confidence value of a space is set to 0.5 (cf. Equation (17), for the following reason. In the external recognition system [1] applied in this example, digit recognition confidence values are normalized in the $[0, 1]$ range, and an image segment is considered as potentially containing a digit only if its recognition confidence is beyond the threshold value of 0.5. In the general case, the confidence value of a space is set to the recognition confidence threshold value. By doing so, a space is considered less significant in the post-clustering voting process described in Section 4.2 than a potentially recognized digit.

Example 2. *It should be noted that in the proposed approach, it is possible that two sequences cannot be economically aligned. Let us consider the alignment of the following recognition hypotheses:*

$$\begin{aligned} h_i &\equiv (s_i, c_i) \equiv (8, c_i[0]), (5, c_i[1]), (7, c_i[2]) \ , \\ h_j &\equiv (s_j, c_j) \equiv (5, c_j[0]), (7, c_j[1]), (3, c_j[2]), (9, c_j[3]) \ , \end{aligned} \tag{18}$$

where the confidence values of particular digits are not specified, since they are irrelevant to this example. The distance matrix generated in Step 1.2 is given in Figure 3. It can be observed that there is only one optimal alignment in set P and that it does not satisfy the condition that no space is inserted into longer sequence s_j. Thus, set P_r is empty, i.e., sequences s_i and s_j cannot be economically aligned.



		5	7	3	9
	0	1	2	3	4
8	1	1	2	3	4
5	2	1	2	3	4
7	3	2	1	2	3

d			i	i
8	5	7	Δ	Δ
Δ	5	7	3	9

Figure 3. Distance matrix and alignment in Example 2.

3.2. Cluster-Based Multiple Sequence Alignment

Let H be a multiset of nonempty recognition hypotheses produced by an external recognition system, i.e.,

$$\begin{aligned} H = \{h_1, h_2, \ldots, h_q\} &= \{(s_1, c_1), (s_2, c_2), \ldots, (s_q, c_q)\} \\ &= \{((s_1[0], c_1[0]), \ldots, (s_1[m_1 - 1], c_1[m_1 - 1])) \\ &\quad ((s_2[0], c_2[0]), \ldots, (s_2[m_2 - 1], c_2[m_2 - 1])), \\ &\quad \ldots, \\ &\quad ((s_q[0], c_q[0]), \ldots, (s_q[m_q - 1], c_q[m_q - 1]))\} \ , \end{aligned} \tag{19}$$

where $g \geq 1$. It is important to note that H is defined as a multiset, i.e., a bag of recognition hypotheses, and not just as a set, in order to emphasize that it may include multiple instances for each of the recognition hypotheses it comprised. The proposed cluster-based multiple sequence alignment can be described as follows.

Step 2.1: Let H_t be a multiset containing recognition hypotheses from H with the maximum length, i.e.,

$$H_t = \{h_i \equiv (s_i, c_i) \mid (h_i \in H) \ \wedge \ (|s_i| = \max_{(s,c) \in H} |s|)\} \ . \tag{20}$$

Each hypothesis in H_t represents one cluster. If $H_t = H$ (i.e., if all recognition hypotheses in H are of equal length), each of the $|H|$ clusters contains exactly one recognition hypothesis from H, and the algorithm jumps to Step 2.3. Otherwise, the algorithm proceeds to Step 2.2.

Step 2.2: If $|H_t| < |H|$, each recognition hypothesis from set $H \setminus H_t$ is either assigned to exactly one cluster or discarded. More particularly, each hypothesis $h \in H \setminus H_t$ is independently aligned—by means of cognitively economical gap-minimum sequence alignment introduced in Section 3.1—to all hypotheses from set H_t, producing a set of alignments:

$$P(h) = \{(\hat{h}, h_t) \mid h_t \in H_t\} \,. \tag{21}$$

If $P(h) = \emptyset$, recognition hypothesis h cannot be economically aligned to any of hypotheses from H_t, and it is discarded. Otherwise, if $P(h) \neq \emptyset$, the alignment from $P(h)$ with minimum value is selected:

$$(\hat{h}, h_t) = \underset{p \in P(h)}{\operatorname{argmin}}\, v(p) \tag{22}$$

(cf. also Equations (12) and (13)), and hypothesis $\hat{h}$ (obtained by transforming observed recognition hypothesis h in the scope of the selected alignment) is assigned to the cluster represented by recognition hypothesis h_t. In a special case when there are multiple optimal instances for hypothesis h_t in multiset H_t, only one of them is randomly selected in Equation (22).

Step 2.3: It is easy to show that all recognition hypotheses (some of them being transformed by adding spaces) assigned to the clusters are of equal length $L_{max} = \max\limits_{(s,c) \in H} |s|$. In this step, they are all arrayed in rows, each of which contains L_{max} columns, and the order or rows is irrelevant. A new sequence s_f containing L_{max} symbols—one for each column—is generated by means of voting. For each column, a symbol from set $S \cup \{\triangle\}$ with the maximum sum of confidence values in the given column is selected. The final recognition result is obtained by removing all spaces from s_f.

Example 3. *To illustrate the proposed algorithm, we consider the set of recognition hypotheses given in Table 1. The algorithm execution is summarized in Table 2.*

Table 2. Illustration of the cluster-based multiple-sequence alignment algorithm. The recognition confidence values of the digits are given in Table 1. The recognition confidence value of a space is set to 0.5.

Step 2.1			**Step 2.2**			**Step 2.3**					
Hypotheses		**Cluster**	**Hypotheses**		**Cluster**	**Arraying**					
h_1	477	?	$\hat{h}_1$	4 △ △7 △ 7	C_8	4	△	△	7	△	7
h_2	4677	?	$\hat{h}_2$	4677 △ △	C_5	4	6	7	7	△	△
h_3	4677	?	$\hat{h}_3$	4677 △ △	C_5	4	6	7	7	△	△
h_4	4677	?	$\hat{h}_4$	4677 △ △	C_5	4	6	7	7	△	△
h_5	467787	C_5	$\hat{h}_5$	467787	C_5	4	6	7	7	8	7
h_6	4677	?	$\hat{h}_6$	4677 △ △	C_5	4	6	7	7	△	△
h_7	4677	?	$\hat{h}_7$	4677 △ △	C_5	4	6	7	7	△	△
h_8	487787	C_8	$\hat{h}_8$	487787	C_8	4	8	7	7	8	7
h_9	477	?	$\hat{h}_9$	4 △ △7 △ 7	C_8	4	△	△	7	△	7
						↓	↓	↓	↓	↓	↓
					Voting:	4	6	7	7	△	7
					Final result:	46777					

In the given set, there are two recognition hypotheses of the maximum length, h_5 and h_8. Therefore, there are two clusters in Step 2.1, which we refer to as C_5 and C_8, respectively. In Step 2.2, recognition hypotheses h_1 and h_9 are economically aligned to h_8 and thus assigned to cluster C_8, while recognition hypotheses h_2, h_3, h_4, h_6, and h_7 are aligned to h_5 and assigned to cluster C_5. After the arraying, voting, and removing spaces from the voting result in Step 2.3, the final recognition result is obtained: 46,777.

It is important to note that although none of the initial recognition hypotheses represents the number given in Figure 1, the proposed algorithm generated the correct recognition result (we recall that the digit after the decimal point in Figure 1 is intentionally discarded).

4. Evaluation and Discussion

The evaluation of the introduced heuristic procedure is performed along two lines. The first evaluation line is aimed at demonstrating that the proposed approach improves real-time object recognition in video streams under uncertainties. Thus, we perform an extrinsic evaluation of the approach in real-life settings (cf. Section 4.1). The second evaluation line is aimed at comparing the proposed approach to the post-processing of recognition hypotheses with human performance (cf. Section 4.2).

4.1. Extrinsic Evaluation

To perform an extrinsic evaluation of the introduced heuristic, it was embedded in an Android-based number recognition system intended for the automatic reading of electricity meters. This recognition system integrates two subsystems that are independent of each other: (i) the number recognition subsystem introduced in [1] that processes each image frame separately and (ii) the post-processing subsystem based on the proposed approach to multiple sequence alignment. For each rate of an electricity meter, a set of image frames is extracted. The first subsystem generates one recognition hypothesis per image frame. The second subsystem post-processes the recognition hypotheses obtained from the first subsystem, by means of cognitively economical multiple sequence alignment described in this paper, and derives a single recognition result.

Two healthy subjects used this integrated number recognition system to automatically read electricity meters in real-life conditions. The experimental settings were designed to reduce the confounding variables:

- Hardware and software: the subjects used Android-based mobile phones of the same type and with the same software settings.
- Subjects: the subjects were of the same gender (male), and comparable in height and expertise in recording electric meters with an Android-based mobile phone. They did not have any insight into the post-processing results.
- Electricity meters: both subjects recorded the same set of 100 electricity meters, including 5 m with one rate, and 95 m with two rates.
- Ambient: to achieve the same ambient conditions, each electricity meter was recorded first by one subject and then immediately after by another.
- Recording span: when reading an electricity meter, the digit recognition system was set to record until ten image frames were recorded or the recording time reached three seconds.

The image frame corpora are described in Table 3. The subjects recorded the same set of electricity meters, but image frames of one particular electricity meter rate were discarded for both the subjects, while image frames of two electricity meters are missing in Subject 2. Thus, Subject 1 captured 2011 image frames that can be divided into 194 disjoint sets (each representing a particular electricity meter rate). The average number of image frames per set for this subject is 10.366 (with standard deviation of 2.109). Subject 2 captured 1873 image frames that can be divided into 190 disjoint sets. The average number of image frames per set for this subject is 9.858 (±1.538). In total, the corpus contains 3884 image frames, which can be divided into 384 sets. The average number of image frames per set is 10.116 (±1.866). In the text below, we refer to the corpora of image frames captured by Subject 1 and Subject 2 as Corpus 1 and Corpus 2, respectively.

The details on the recognition accuracy at the number level are also provided in Table 3. It is important to note that neither of the subsystems had any predefined expectation in terms of the number of digits in a correct recognition hypothesis. The external number recognition system correctly recognized 48.15% (i.e., 1870 of 3884) of image frames: 50.32% accuracy was obtained for Corpus 1 and 45.81% for Corpus 2. This low recognition accuracy is caused by significant noise and incompleteness contained in the image frames. As a small illustration of the corpus quality, we point out that the average root-mean-square contrast [27] of images in the corpora is relatively low at 0.218, with standard deviation of 0.041. However, it is important to note that such a low-quality image corpora was intentionally adopted to increase the recognition uncertainty and to confront the post-processing subsystem, which is the actual object of the evaluation, with a real-life challenge.

Table 3. Image frame corpus and recognition accuracy.

System	Image Frames	Corpus 1	Corpus 2	Total
External number recognition system [1]	**# image frames**	2011	1873	3884
	average root mean square contrast	0.227 (±0.040)	0.208 (±0.039)	0.218 (±0.041)
	# correctly recognized image frames	1012 (50.32%)	858 (45.81%)	1870 (48.15%)
Proposed post-processing subsystem	**# hypothesis sets**	194	190	384
	average # hypotheses per set	10.366 (±2.109)	9.858 (±1.538)	10.116 (±1.866)
	average # correct hypotheses per set	5.201 (±3.766)	4.437 (±3.647)	4.823 (±3.727)
	# correctly aligned recognition result	137 (70.62%)	125 (65.79%)	262 (68.23%)

It can be observed that the proposed heuristic procedure for alignment of recognition hypotheses significantly increased the recognition accuracy, from 48.15%, obtained prior to the embedding of the introduced heuristic procedure, to 68.23%, obtained after the embedding. In total, 68.23% (i.e., 262 of 384) of hypothesis sets were correctly aligned (70.62% accuracy obtained for Corpus 1, and 65.79% for Corpus 2). To additionally describe the performance, we make the following points (summarized in Table 4):

- A total of 86 of 384 recognition hypothesis sets do not contain correct recognition hypotheses (39 sets in Corpus 1 and 47 sets in Corpus 2). However, the hypotheses from seven of these sets were correctly aligned; i.e., the correct recognition results were derived (in 1 of 39 sets in Corpus 1 and 6 of 47 sets in Corpus 2). One of these sets and the derivation of the recognition result are presented above in Example 3.
- A total of 114 of 384 sets contain at least one correct recognition hypothesis, but the number of correct hypotheses in each of these sets is less than or equal to the half of the number of hypotheses in a given set (59 sets in Corpus 1 and 55 sets in Corpus 2). The correct recognition results were derived for 72 of these sets (40 of 59 sets in Corpus 1 and 32 of 55 sets in Corpus 2).
- A total 184 of 384 sets contain correct recognition hypotheses, and the number of correct hypotheses in each of these sets is greater than the half of the number of hypotheses in a given set (96 sets in Corpus 1 and 88 sets in Corpus 2). The correct recognition results were derived for all these sets except one (from Corpus 2).

Table 4. An overview of image frame sets.

# Correct Recognition	Corpus 1		Corpus 2		Total	
Hypotheses in a Set	# Sets	# Correctly Aligned	# Sets	# Correctly Aligned	# Sets	# Correctly Aligned
0	39	1	47	6	86	7
$\leq$half of the set, $\neq$0	59	40	55	32	114	72
$>$half of the set	96	96	88	87	184	183
Total	194	137	190	125	384	262

To evaluate the recognition performance at the digit level, the confusion matrices for Corpora 1 and 2 are given in Tables 5 and 6, respectively. The results can be summarized as follows. In Corpus 1, 93.50% digits are correctly recognized, 2.99% incorrectly recognized, and 3.51% not detected. In Corpus 2, 92.73% digits are correctly recognized, 4.21% incorrectly recognized, and 3.06% not detected. It can be also derived that, in total, 93.10% digits are correctly recognized, 3.60% incorrectly recognized, and 3.28% not detected.

Table 5. Confusion matrix for Corpus 1 (INS—segment incorrectly recognized as a digit; ND—digit not detected).

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**ND**	**Total**
0	92	1	0	0	0	0	0	0	3	0	3	99
1	0	90	0	0	0	0	0	0	0	0	3	93
2	0	0	94	0	1	0	0	0	0	0	6	101
3	0	2	0	115	0	0	0	0	0	2	2	121
4	0	1	0	0	101	0	0	0	0	0	3	105
5	0	1	0	1	0	74	1	0	0	0	1	78
6	4	0	0	1	0	1	80	0	2	0	6	94
7	0	1	0	0	0	0	0	86	4	0	4	95
8	1	0	0	0	0	0	0	1	89	0	4	95
9	0	0	0	1	0	0	0	0	0	85	2	88
INS	0	5	0	1	1	0	1	1	5	0	–	14

Table 6. Confusion matrix for Corpus 2 (INS—segment incorrectly recognized as a digit; ND—digit not detected).

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**ND**	**Total**
0	88	2	0	2	0	0	0	0	2	0	4	98
1	0	92	0	0	0	0	0	0	0	0	1	93
2	0	0	89	0	4	0	0	4	0	0	2	99
3	0	0	0	103	0	0	0	0	0	1	6	110
4	0	0	0	0	99	0	0	1	0	0	5	105
5	1	0	0	4	0	70	0	0	0	0	1	76
6	6	0	0	1	0	3	78	0	1	0	4	93
7	0	0	1	0	1	0	0	88	1	0	1	92
8	1	0	0	0	0	0	0	0	93	0	3	97
9	2	0	0	0	0	0	0	0	2	80	2	86
INS	0	12	1	5	1	0	2	2	10	9	–	42

4.2. Comparison to Human Performance

In the second evaluation phase, the proposed heuristic procedure is compared to human performance in the post-processing of recognition hypotheses. For this purpose, an additional naïve healthy subject was given the same 384 sets of recognition hypotheses to which the post-processing subsystem was confronted in the first evaluation phase. The subject did not know how many digits were expected in a correct recognition hypothesis. For each of the given sets, the task of the subject was to try to derive the correct recognition result. The recognition hypotheses were presented on a screen. Paper and pencil were available to the subject. The time was not limited.

The results of the comparative analysis are summarized in Table 7. For Corpus 1, the human and the system had the same recognition accuracy. They both derived correct results for 137 of 194 sets, with the overlapping of 129 sets. In addition, of 194 sets in Corpus 1, the human and the system derived the same results for 157 sets. For Corpus 2, the system outperformed the human. Of 190 sets, the system derived the correct results for 125 sets, and the human for 121 sets, with the overlapping of 114 sets. In addition, of 190 sets in Corpus 2, the human and system derived the same results for 141 sets. In total, the human and the system derived the same results for 77.60% of all recognition hypothesis sets.

The difference in performance can be explained as follows. Table 3 shows that the average number of correct recognition hypotheses per set is lower in Corpus 2 (5.201 ± 3.766) than in Corpus 1 (4.437 ± 3.647), most probably due to the recording style of Subject 2.

In line with this, in the last rows of Tables 5 and 6 it can be observed that the number of image segments that were incorrectly recognized as digits by the underlying recognition system is greater for Corpus 2 than for Corpus 1. Thus, in Corpus 2 both the human and the system were confronted with more recognition hypotheses containing false-positive segments. We recall that neither of them knew how many digits are expected in a correct recognition hypothesis. The system performed better due to its cognitively economical design. The longer-than-necessary recognition hypotheses are not additionally extended by spaces (i.e., gap-minimum alignment), and the false-positive segments are more effectively eliminated in the post-clustering voting.

Table 7. Comparative analysis (C1—Corpus 1; C2—Corpus 2).

	# Sets	**# Correctly Derived Results**			**Total Overlap**
		Heuristic	**Human**	**Overlap**	
C1	194 (100%)	137 (70.62%)	137 (70.62%)	129 (66.49%)	157 (80.93%)
C2	190 (100%)	125 (65.79%)	121 (63.68%)	114 (60%)	141 (74.21%)
Total	384 (100%)	262 (68.23%)	258 (67.19%)	243 (63.28%)	298 (77.60%)

5. Conclusions

In this paper, we introduced a heuristic approach to multiple sequence alignment under uncertainties. The proposed approach was cognitively economical to the extent that it accounted for human working memory limitations and thus had a reduced computational complexity in comparison to the optimal multiple sequence alignment. On the other hand, its relevance was experimentally confirmed.

The evaluation was performed along two lines. First, an extrinsic evaluation conducted in real-life settings demonstrated that the proposed approach improves the accuracy of number recognition in short video streams under uncertainties caused by noise and incompleteness. At the number level (i.e., sequence of digits), the recognition accuracy of a given external recognition system was increased from 48.15%, obtained prior to the embedding of the introduced heuristic procedure, to 68.23%, obtained after the embedding. At the digit level, the improved performance is reflected through the recognition accuracy of 93.10%.

In the second evaluation phase, the proposed heuristic procedure was compared to human performance in the post-processing of recognition hypotheses. A naïve subject was given the same 384 sets of recognition hypotheses to which the post-processing subsystem was confronted in the first evaluation phase. For each of the given sets, the task of the subject was to try to derive the correct recognition result. It was demonstrated that the proposed approach outperformed the human. This indicates that the proposed heuristic for post-processing of the recognition hypotheses may be combined with machine learning approaches, which are typically not tailored for the task of object sequence recognition from a limited number of frames of incomplete data recorded in a dynamic scene situation.

Author Contributions: Conceptualization, M.G.; methodology, M.G.; software, M.G. and N.M.; validation, M.S., S.A. and D.J.; formal analysis, M.S., S.A. and D.K.; investigation, M.G., N.M. and D.K.; writing—original draft preparation, M.G.; writing—review and editing, M.G. and N.M. All authors have read and agreed to the published version of the manuscript.

Funding: The work of M.G. was partially funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia, under the Research Grants III44008 and TR32035. The work of S.A. was partially funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia, under the Research Grant III44006. The work of D.J. was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia Grant, under the Research Grant No. 337-00-426/2021-09, and by the National Key R&D Program of China under the Research Grant No. 2021YFE0110500.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Gnjatović, M.; Maček, N.; Adamović, S. Putting Humans Back in the Loop: A Study in Human-Machine Cooperative Learning. *Acta Polytech. Hung.* **2020**, *17*, 191–210. [CrossRef]
2. Singh, P.; Diwakar, M.; Gupta, R.; Kumar, S.; Chakraborty, A.; Bajal, E.; Jindal, M.; Shetty, D.K.; Sharma, J.; Dayal, H.; et al. A Method Noise-Based Convolutional Neural Network Technique for CT Image Denoising. *Electronics* **2022**, *11*, 3535 . [CrossRef]
3. Momeny, M.; Latif, A.M.; Agha Sarram, M.; Sheikhpour, R.; Zhang, Y.D. A noise robust convolutional neural network for image classification. *Results Eng.* **2021**, *10*, 100225. [CrossRef]
4. De Man, R.; Gang, G.J.; Li, X.; Wang, G. Comparison of deep learning and human observer performance for detection and characterization of simulated lesions. *J. Med Imaging* **2019**, *6*, 025503. [CrossRef] [PubMed]
5. Montalt-Tordera, J.; Muthurangu, V.; Hauptmann, A.; Steeden, J.A. Machine learning in Magnetic Resonance Imaging: Image reconstruction. *Phys. Med. Eur. J. Med. Phys.* **2021**, *83*, 79–87. [CrossRef] [PubMed]
6. Gnjatović, M.; Maček, N.; Adamović, S. A Non-Connectionist Two-Stage Approach to Digit Recognition in the Presence of Noise. In Proceedings of the 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Naples, Italy, 23–25 October 2019; pp. 15–20. [CrossRef]
7. Finton, D.J. Cognitive-Economy Assumptions for Learning. In *Encyclopedia of the Sciences of Learning*; Seel, N., Ed.; Springer: Boston, MA, USA, 2012; pp. 626–628. ._565. [CrossRef]
8. Hui, W.; Yu, L. The uncertainty and explainability in object recognition. *J. Exp. Theor. Artif. Intell.* **2021**, *33*, 807–826. [CrossRef]
9. Heydari, M.; Mylonas, A.; Tafreshi, V.H.F.; Benkhelifa, E.; Singh, S. Known unknowns: Indeterminacy in authentication in IoT. *Future Gener. Comput. Syst.* **2020**, *111*, 278–287. [CrossRef]
10. Wang, L.; Jiang, T. On the complexity of multiple sequence alignment. *J. Comput. Biol.* **1944**, *1*, 337–348. . cmb.1994.1.337. [CrossRef]
11. Jurafsky, D.; Martin, J.H. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*, 2nd ed.; Prentice-Hall: Hoboken, NJ, USA, 2009.
12. Levenshtein, V.I. Binary codes capable of correcting deletions, insertions, and reversals, Cybernetics and Control Theory. *Cybern. Control. Theory* **1966**, *10*, 707–710.
13. Wagner, R.A.; Fischer, M.J. The String-to-String Correction Problem. *J. Assoc. Comput. Mach.* **1974**, *21*, 168–173. [CrossRef]
14. Chao, J.; Tang, F.; Xu, L. Developments in Algorithms for Sequence Alignment: A Review. *Biomolecules* **2022**, *12*, 546. [CrossRef]
15. Alkuhlani, A.; Gad, W.; Roushdy, M.; Voskoglou, M.G.; Salem, A.b.M. PTG-PLM: Predicting Post-Translational Glycosylation and Glycation Sites Using Protein Language Models and Deep Learning. *Axioms* **2022**, *11*, 469. [CrossRef]
16. Daugelaite, J.; O' Driscoll, A.; Sleator, R.D. An Overview of Multiple Sequence Alignments and Cloud Computing in Bioinformatics. *ISRN Biomath.* **2013**, *2013*, 615630. [CrossRef]
17. Kholidy, H.A. Detecting impersonation attacks in cloud computing environments using a centric user profiling approach. *Future Gener. Comput. Syst.* **2021**, *117*, 299–320. [CrossRef]
18. Campbell, J.; Lewis, J.P.; Seol, Y. Sequence alignment with the Hilbert-Schmidt independence criterion. In Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production, London, UK, 13–14 december 2018; pp. 1–8. [CrossRef]
19. Chasanis, V.T.; Likas, A.C.; Galatsanos, N.P. Scene Detection in Videos Using Shot Clustering and Sequence Alignment. *IEEE Trans. Multimed.* **2009**, *11*, 89–100. [CrossRef]
20. Dogan, P.; Li, B.; Sigal, L.; Gross, M. A Neural Multi-Sequence Alignment TeCHnique (NeuMATCH). In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 8749–8758. [CrossRef]
21. Schimke, S.; Vielhauer, C.; Dittmann, J. Using adapted Levenshtein distance for on-line signature authentication. In Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, 26 August 2004; Volume 2, pp. 931–934. [CrossRef]
22. Just, W. Computational complexity of multiple sequence alignment with SP-score. *J. Comput. Biol.* **2001**, *8*, 615–623. [CrossRef]
23. Herman, J.L.; Novák, A.; Lyngsø, R.; Szabó, A.; Miklós, I.; Hein, J. Efficient representation of uncertainty in multiple sequence alignments using directed acyclic graphs. *BMC Bioinform.* **2015**, *16*, 108. [CrossRef]
24. Ma, L.; Chamberlain, R.D.; Agrawal, K.; Tian, C.; Hu, Z. Analysis of classic algorithms on highly-threaded many-core architectures. *Future Gener. Comput. Syst.* **2018**, *82*, 528–543. [CrossRef]
25. Feng, D.F.; Doolittle, R.F. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **1987**, *25*, 351–360. [CrossRef]

26. Miller, G. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychol. Rev.* **1956**, *63*, 81–97. [CrossRef]
27. Peli, E. Contrast in complex images. *J. Opt. Soc. Am. A* **1990**, *7*, 2032–2040. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.