https://doi.org/10.31891/2219-9365-2024-79-32
UDC 004.93.1

GUANXIANG Xia
Vinnytsia National Technical University
Hunan Mass Media Vocational and Technical College
xiaguanxiang@163.com
KOVTUN Viacheslav
Vinnytsia National Technical University
https://orcid.org/0000-0002-7624-7072
kovtun_v_v@vntu.edu.ua

THE METHOD FOR STOPPING THE STRING OBJECT RECOGNITION PROCESS IN A REAL-TIME VIDEO STREAM

This article explores the challenge of optimal stopping in real-time string object recognition within video streams, a critical issue for systems with constrained computational resources, such as mobile devices. The problem is framed as a decision-making task that balances the trade-off between computational costs and recognition accuracy. A mathematical model is proposed, introducing a stopping rule based on the expected distance between the current and subsequent integrated recognition results, ensuring efficiency and precision in dynamic environments. The developed algorithm leverages the normalized Levenshtein distance as a metric for assessing recognition accuracy, integrating results using the ROVER algorithm. Experimental validation was conducted on the MIDV-500 dataset, encompassing diverse text areas from identity documents, such as dates, machine-readable zones (MRZ), document numbers, and personal names. Using the Tesseract OCR engine (versions 4.1.1 and 5.5.0), the method demonstrated significant improvements in recognition accuracy and resource efficiency compared to traditional fixed-step and threshold-based approaches. The findings highlight the robustness and versatility of the proposed approach, as it adapts seamlessly to varying text recognition algorithms and computational environments. The article underscores the broader applicability of the method, suggesting extensions to other object types, including images and complex patterns. Furthermore, the integration with machine learning models and dynamic confidence scoring is proposed to enhance decision-making accuracy.
Keywords: real-time object recognition, machine learning integration, string objects, video streams, Optimal stopping, computational efficiency, dynamic confidence scoring.

ГУАНСЯНГ Ся
Вінницький національний технічний університет
Хунанський коледж масових медіа
КОВТУН В'ячеслав
Вінницький національний технічний університет

МЕТОД ЗУПИНКИ ПРОЦЕСУ РОЗПІЗНАВАННЯ РЯДКОВИХ ОБ'ЄКТІВ У ВІДЕО ПОТОЦІ РЕАЛЬНОГО ЧАСУ

У статті досліджується проблема оптимальної зупинки процесу розпізнавання рядкових об'єктів у відео потоці реального часу, що є критично важливою для систем з обмеженими обчислювальними ресурсами, такими як мобільні пристрої. Проблема формалізується як задача прийняття рішень, націлена на знаходження балансу між обчислювальними витратами та точністю розпізнавання. Представлена математична модель, яка вводить правило зупинки на основі очікуваної відстані між поточними та наступними інтегрованими результатами розпізнавання, забезпечуючи ефективність і точність у динамічних умовах. Розроблено алгоритм, який використовує нормалізовану відстань Левенштейна як метрику для оцінки точності розпізнавання, інтегруючи результати, отримані за допомогою алгоритму ROVER. Експериментальна перевірка була проведена на наборі даних MIDV-500, що містить різні рядкові об'єкти як елементи ідентифікаційних документів, зокрема, дати, зони для машинного зчитування, номери документів та власні імена. Використовуючи OCR-фреймворк Tesseract (версій 4.1.1 та 5.5.0), метод продемонстрував значні покращення точності розпізнавання та ефективності використання ресурсів порівняно з традиційними підходами, орієнтованими на використання фіксованих кроків й порогів. Результати дослідження підкреслюють надійність і універсальність запропонованого підходу, оскільки він безперешкодно адаптується до різних середовищ та алгоритмів розпізнавання тексту. Стаття підкреслює широку застосовність методу, пропонуючи розширення на інші типи об'єктів, включаючи фото та захисні зображення. Крім того, запропонована інтеграція з моделями машинного навчання та динамічним оцінюванням коректності розпізнавання для покращення точності прийняття рішень..
Ключові слова: розпізнавання об'єктів у реальному часі, інтеграція з машинним навчанням, рядкові об'єкти, відеопотоки, оптимальна зупинка, обчислювальна ефективність, динамічне оцінювання коректності розпізнавання.

THE PROBLEM STATEMENT IN GENERAL FORM AND ITS CONNECTION WITH IMPORTANT SCIENTIFIC OR PRACTICAL TASKS

The optimal stopping problem [1-3] has gained significant attention, particularly in the context of real-time computer vision systems, especially those implemented on mobile devices. This problem extends beyond simply achieving accurate object recognition [4], as it also addresses the critical aspect of time. In such systems, the time required to obtain a recognition outcome is often as crucial as the accuracy of the result, making the task of determining the best moment to stop processing a vital consideration [5]. The challenge lies in balancing the need for further information with the computational cost required to obtain it.

The optimal stopping problem refers to the decision-making process that dictates when to halt an ongoing task, optimizing the trade-off between continued information acquisition and system efficiency [6]. This decision becomes particularly complex in dynamic, real-time environments, where quick processing and precise outcomes are necessary. Delays in these systems can significantly impact performance, especially in mobile computing, where resources such as processing power and energy are limited.

In the context of real-time object recognition, the problem takes on added significance, as mobile devices need to make decisions about when to stop processing video streams in order to provide timely and accurate results [7]. This problem is closely related to various scientific and practical challenges, such as enhancing automated systems, improving energy efficiency, and optimizing real-time decision-making in environments with constrained resources. The solutions to this problem have a profound impact on fields like robotics, autonomous vehicles, and other mobile applications, where speed, precision, and resource management are crucial. This article explores the optimal stopping problem within the context of real-time object recognition, discussing its implications for mobile computing and proposing potential solutions to improve both efficiency and effectiveness in such systems.

THE PROBLEM STATEMENT AND REVIEW OF RECENT RESEARCH

Contemplate the object recognition problem in a real-time video stream. Let P denote the entire set of possible values of the recognizable object (e.g., the set of strings over a fixed alphabet in the case of recognizing text arias), with a determined metric $l: P \times P \to [0, \infty)$, on this set. The real-time video stream is used to recognize an object with a true value $\mathrm{P}^* \in P$. The recognition process assumes that a decision-maker observes a series of random recognition outcomes $\mathrm{P} = (\mathrm{P}_1, \mathrm{P}_2, \ldots)$, one result per step of the process, and every observation $\rho_i \in P$ is an execution of P_i.

We suppose that $\mathrm{P}_1, \mathrm{P}_2, \ldots$ shares the same joint distribution as P^*. Within this framework, we assume that confidence estimates for the recognition results of the object are unavailable. We also define a family of integration functions for multiple recognition results, which return a single integrated result $Y^{(m)}: P^m \to P$, as their output. At any moment m, the observation results $\mathrm{P}_1 = \rho_1, \ldots, \mathrm{P}_m = \rho_m$ are available, and an integrated result $Y_m = Y^{(m)}(\rho_1, \ldots, \rho_m)$ can be acquired. The process can be interrupted at any moment $m > 0$ with the following penalty function:

$$k_d l\left(Y_m \mathrm{P}^*\right) + k_f m, \tag{1}$$

where $k_d > 0$ represents the recognition error cost in relation to the range to the true value, and $k_f > 0$ denotes the every observation cost.

Since k_d and k_f are non-negative constants, a penalty function can be redefined devoid of altering the optimization problem structure as follows:

$$L_m \stackrel{def}{=} l\left(Y_m, \mathrm{P}^*\right) + km, \tag{2}$$

where $k = k_f / k_d$ represents the functional dependence of the penalty. The penalty value when stopping at step $m = 0$ (i.e., if no observations have been obtained) can be considered infinite.

The task is to select the step at which the observation process should be stopped to minimize the anticipated penalty. This formulation can be formalized using the notation from [106]. A stopping rule can be outlined as the series of functions:

$$\Psi \stackrel{def}{=} \left(\psi_0, \psi_1(\rho_1), \psi_2(\rho_1, \rho_2), \psi_3(\rho_1, \rho_2, \rho_3), \ldots\right), \tag{3}$$

where $\forall m: 0 \le \psi_m(\rho_1, \ldots, \rho_m) \le 1$ represents the stopping criterion. The function $\psi_m(\rho_1, \ldots, \rho_m)$ reflects the probability of stopping at moment m, considering that this step has been attained (i.e., based on the observations $\mathrm{P}_1 = \rho_1, \ldots, \mathrm{P}_m = \rho_m$ obtained up to that point).

Using a sequence of observations P and the stopping rule Ψ, it is possible to define the random stopping time M. Let $R\left(M = m \middle| \mathrm{P} = (\rho_1, \rho_2, \ldots)\right)$ denote the probability function for stopping at step m, given a specific sequence of observations P. This function is expressed in terms of the stopping rule (3) as follows:

$$R\left(M=0\middle|\mathrm{P}=\left(\rho_1,\rho_2,\ldots\right)\right)=\psi_0;$$

$$R\left(M=m\middle|\mathrm{P}=\left(\rho_1,\rho_2,\ldots\right)\right)=\psi_m\left(\rho_1,\ldots,\rho_m\right)\prod_{j=1}^{m-1}\left(1-\psi_j\left(\rho_1,\ldots,\rho_j\right)\right),\ \forall m\in\{1,2,\ldots\}; \tag{4}$$

$$R\left(M=\infty\middle|\mathrm{P}=\left(\rho_1,\rho_2,\ldots\right)\right)=1-\sum_{j=0}^{\infty}R\left(M=j\middle|\mathrm{P}=\left(\rho_1,\rho_2,\ldots\right)\right).$$

Alternatively, considering the random stopping time M, a stopping rule for moment $m=\{0,1,\ldots\}$ can likewise be represented as the conditional probability of stopping at the moment m, considering a specific series of observations P, and assuming a process did not terminate at earlier stages:

$$\psi_m\left(\mathrm{P}_1,\ldots,\mathrm{P}_m\right)=R\left(M=m\middle|M\geq m,\mathrm{P}=\left(\rho_1,\rho_2,\ldots\right)\right). \tag{5}$$

The task is to determine a stopping rule Ψ that minimizes the anticipated loss functional $W\left(\Psi\right)$. This can be formulated as follows:

$$W\left(\Psi\right)=\mathrm{E}\left(L_m\left(\mathrm{P}_1,\ldots,\mathrm{P}_M\right)\right). \tag{6}$$

The optimal stopping problem has received considerable attention across various disciplines, including decision theory, computer vision, and machine learning [8]. Advancements in real-time object recognition highlight the need to balance accuracy and computational efficiency, especially for mobile and resource-constrained systems. Fixed-step and threshold-based methods dominate current approaches but often fail to adapt dynamically to real-time challenges.

In optical character recognition (OCR), tools like Tesseract [9] have significantly improved text recognition accuracy [10]. However, they generally overlook the optimal stopping challenge for sequential tasks in video streams. Traditional methods, such as fixed-frame analysis or cluster-based thresholds, lack flexibility, leading to either premature termination or excessive computational costs.

Recent developments in anytime algorithms offer promising avenues to enhance adaptability. Dynamic decision-making mechanisms, such as confidence scores and integration functions, have demonstrated the potential to improve the trade-off between accuracy and processing speed. However, many of these approaches are resource-intensive or constrained by static parameters, limiting their scalability to various contexts [11].

This article advances the field by introducing a mathematical framework for optimal stopping in real-time string object recognition. By employing normalized Levenshtein distance [12] and ROVER-based integration [13], the proposed method dynamically determines stopping points without reliance on confidence metrics. This innovation addresses a critical gap by enabling efficient, adaptable mechanisms suitable for resource-limited systems, expanding the applicability of real-time recognition technologies across diverse domains.

FORMULATION OF THE ARTICLE'S OBJECTIVES

The aim of the study is to develop a method for determining the stopping point in the real-time recognition process of string objects that minimizes computational costs while improving the accuracy of results. To achieve this aim, the following research objectives must be accomplished:

1. Develop a mathematical model and algorithm for optimal stopping in the recognition process, considering accuracy and computational costs.
2. Conduct an experimental study of the proposed method on real data and compare its performance with existing approaches.
3. Assess the applicability of the method in systems with limited computational resources, including mobile devices.

PRESENTATION OF THE MAIN MATERIAL

We formulate the following requirement for the integration functions $Y^{(m)}$: the Anticipated range among two sequential integrated results of recognition remains stable over time:

$$\mathrm{E}\left(l\left(Y_m,Y_{m+1}\right)\right)\geq\mathrm{E}\left(l\left(Y_{m+1},Y_{m+2}\right)\right),\ \forall m>0. \tag{7}$$

Within the scope of "anytime" algorithms, the requirement (7) implies that the problem shows the characteristic of diminishing returns. Based on this supposition concerning the integration functions $Y^{(m)}$, it can be shown that a stopping problem (6) with a loss function (2) turns into monotonic from a specific step onward.

Actually, suppose U_m represent occurrence $\left\{\mathrm{E}_m\left(l\left(Y_m, Y_{m+1},\right)\right) \leq k\right\}$, and examine the stopping problem (6) beginning at the step m, where occurrence U_m first took place. The occurrence B_m regarded in the monotonicity state be presented in the following way:

$$B_m:\left\{l\left(Y_m, \mathrm{P}^*\right)+km \leq \mathrm{E}_m\left(l\left(Y_{m+1}, \mathrm{P}^*\right)\right)+km+c\right\}=\left\{l\left(Y_m, \mathrm{P}^*\right)-\mathrm{E}_m\left(l\left(Y_{m+1}, \mathrm{P}^*\right)\right) \leq k\right\}. \quad (8)$$

For a fixed P^*, at step m, using the triangle inequality, we can acquire the following relation among the range from the ongoing result of recognition to the true value, the anticipated range to the result at the following step, and the anticipated range from the following result to the true value:

$$\begin{gathered} l\left(Y_m, \mathrm{P}^*\right) \leq \mathrm{E}_m\left(l\left(Y_m, Y_{m+1}\right)\right)+\mathrm{E}_m\left(l\left(Y_{m+1}, \mathrm{P}^*\right)\right) \Rightarrow \\ \Rightarrow l\left(Y_m, \mathrm{P}^*\right)-\mathrm{E}_m\left(l\left(Y_{m+1}, \mathrm{P}^*\right)\right) \leq \mathrm{E}_m\left(l\left(Y_m, Y_{m+1}\right)\right). \end{gathered} \quad (9)$$

This inequality provides a bound on the ongoing range, showing that the anticipated future improvement in recognition accuracy is at least partially determined by the change in range at the following step.

If the right term of the inequality holds in (9) does not exceed the non-negative constant k, then the left term also does not exceed k, and consequently, if occurrence U_m occurs, occurrence (8) must also occur. Moreover, using the assumption (7), we can infer that if occurrence U_m occurs, then occurrence U_{m+1} must also occur. Thus, we obtain:

$$\forall m>0: U_m \subset B_m,\ U_m \subset U_{m+1}. \quad (10)$$

From this, it follows that starting from step m, where occurrence U_m occurred for the first time, occurrence s $B_m, B_{m+1}, B_{m+2}, \ldots$ will also occur. Therefore, the stopping problem can be considered monotonic starting from this step, which implies the optimality of rule (6) within all stopping rules that attain step m in the case where the problem has a finite horizon.

Now, let us examine a stopping rule that instructs the determiner to stop the recognition process if occurrence U_m occurs:

$$M_U=\min\left\{m>0: \mathrm{E}_m\left(l\left(Y_m, Y_{m+1}\right)\right)\right\}. \quad (11)$$

This rule suggests that the recognition process is terminated when a specific occurrence , denoted as U_m, happens during the observation sequence.

If rule (11) requires stopping at step m, then rule M_B will also require stopping at this step. Since the problem becomes monotonic starting from step m, the decision of rule M_B is optimal, and therefore, the optimal rule M^* will also require stopping at this step. Moreover, if $l\left(Y_m, \mathrm{P}^*\right)-\mathrm{E}_m\left(l\left(Y_{m+1}, \mathrm{P}^*\right)\right)>k$ occurs, rule M_B does not stop the process, just like rule M^*, which follows the principle of optimality. Therefore, if assumption (7) holds true, rule M_B will never stop prematurely, and if the rule dictates stopping, the decision to stop is optimal.

In the proposed method, we will suppose that a metric defined on the all possible object recognition results set is upper-bounded (i.e., $\exists G: \forall \rho_i, \rho_j \in P: 0 \leq l\left(\rho_i, \rho_j \leq G\right)$) and that the integration functions $Y^{(m)}$ produce results that, on mean, do not deteriorate over time:

$$\mathrm{E}\left(l\left(Y_m, \mathrm{P}^*\right)\right) \geq \mathrm{E}\left(l\left(Y_{m+1}, \mathrm{P}^*\right)\right),\ \forall m>0. \quad (12)$$

Thus, to solve the stopping problem (6) with the loss functional (2), the following method is proposed:

1. Estimate the anticipated range (in terms of the metric l) from the ongoing integrated object recognition result Y_m (known at step m) to the unknown result Y_{m+1} at the following step;

2. Make a decision to stop the process at step m by thresholding the range estimated in step 1, thereby approximating the behavior of the rule M_U.

Generally, the choice of a method for forecasting the following integrated object recognition result (or estimating the anticipated range among it and the ongoing integrated result) could rely on the character of the integration functions $Y^{(m)}$ and other specific characteristics of the problem.

Based on the proposed method, let us construct a stopping algorithm for the recognition process of a string object. Suppose the functions for integrating the recognition results of the string object $Y^{(m)}$ are given (which can be implemented using the ROVER method [14]. As the metric l for string objects, it is proposed to use the normalized generalized Levenshtein range. In order to approximate the behavior of the stopping rule M_U , at the m -th step of the process, it is necessary to compute the estimate of the anticipated range among adjacent integrated recognition results $\Delta_m \overset{def}{=} \mathrm{E}_m\left(l\left(Y_m, Y_{m+1}\right)\right)$, having access to the observations $\mathrm{P}_1 = \rho_1, \ldots, \mathrm{P}_m = \rho_m$. To calculate the estimate, it is proposed to simulate the following integrated result, considering that a new observation will be near to those obtained in the previous steps:

$$\Delta_m \overset{def}{=} \frac{1}{m+1}\left(\xi + \sum_{i=1}^{m} l\left(Y_m, Y\left(\rho_1, \rho_2, \ldots, \rho_m, \rho_i\right)\right)\right), \tag{13}$$

where ξ is an external adjustable parameter.

To apply the method described above, it is necessary to define the metric l and the integration functions $Y^{(m)}$ for the set of strings P . As a metric on the set of strings, it is proposed to use the normalized Levenshtein range:

$$l_L\left(\rho_i, \rho_j\right) \overset{def}{=} \frac{2 levenshtein\left(\rho_i, \rho_j\right)}{\left|\rho_i\right| + \left|\rho_j\right| + levenshtein\left(\rho_i, \rho_j\right)}, \tag{14}$$

where $\left|\rho_i\right|$ is the string ρ_i length, and $levenshtein\left(\rho_i, \rho_j\right)$ is a Levenshtein range among strings ρ_i and ρ_j . This metric values lie within the range [0, 1], and its normalization is performed while preserving the triangle inequality.

As integration functions $Y^{(m)}$, the ROVER algorithm [13] was used. To implement the method, it is necessary to introduce a threshold λ for evaluating the empty class, which is considered in the voting module. In the experiments conducted the threshold value $\lambda = 0.6$ was used.

The experimental study was conducted on the open MIDV-500 dataset [14], which holds 50 different types of identity certificates videos (with 10 video clips for each certificate; 30 frames per video) with explained ideal locations and content of text areas. Four groups of arias were analyzed: dates recorded with numbers and punctuation marks, Machine-Readable Zone (MRZ) strings, certificate number, and elements of the certificate holder's name written in the Latin alphabet.

Only frames where the certificate was completely were regarded (consequently, the video sequences in the analyzed dataset subset had varying lengths, ranging between 1 frame up to 30 frames). To provide a clearer representation of the results and minimize normalization effects, each video clip was lengthened up to 30 frames by repeating it from the beginning. Thus, all analyzed clips had a uniform length of 30 frames.

Each aria was clipped from the initial image using a projective transformation, based on the combined annotation of the ideal certificate boundaries and the coordinates of the text aria, with added margins matching 15% of the shortest side of the text aria. The size of the cropped text aria images matched to the 300 dots per inch resolution. Each cropped text aria was recognized applying the Free Tesseract recognition software (versions v4.1.1 and v5.5.0 with default parameters for the English. Every character value comparisons were case- neutral, and the the digit "0" was considered identical to Latin letter "O."

Table 1 provides, for all group of text arias, a number of unique arias in the MIDV-500 dataset, the entire number of text aria images (over all frames where the certificate is fully visible), and the mean length of the frame sequence. The table also presents the mean range P_i among the recognition result for a single frame and the ground truth value P^*, as well as among the integrated result for the video clip and the ground truth value P^*, both before (Y_{fin}) and after (Y_{30}) padding in terms of the metric defined in (14)).

To evaluate the effectiveness of the stopping rule, an efficiency profile can be created, visually showing the relationship among the maen number of analyzed observations and the corresponding mean range from the obtained integrated result at the stopping point to the ground truth value, while varying the observation cost k . Such an efficiency profile reflects the trade-off among the required time for video sequence processing and the accuracy of the obtained recognition result, as well as allows for a visual comparison of different stopping strategies.

Table 1.

Mean Values of Metric l_L to Ground Truth for Recognition Results Using the Tesseract Library for Text Arias of the MIDV-500 Dataset:
P_i - recognition result for a single frame, Y_{fin} - integrated recognition result for the video clip obtained using a modified ROVER algorithm, Y_{30} - integrated recognition result for the padded video clip obtained using a modified ROVER algorithm.

	Document number	Date	Name	MRZ sring	All arears
Unique arears	48	91	79	30	248
Total number of clips	436	824	719	260	2239
Total number of images	9329	17735	15587	5096	47747
Mean clip length	21.397	21.523	21.679	19.600	21.325
Tesseract v4.1.1					
$\mathrm{E}\left(l_L\left(\mathrm{P}_i,\mathrm{P}^*\right)\right)$	0.422	0.360	0.443	0.258	0.388
$\mathrm{E}\left(l_L\left(Y_{fin},\mathrm{P}^*\right)\right)$	0.326	0.244	0.338	0.162	0.281
$\mathrm{E}\left(l_L\left(Y_{30},\mathrm{P}^*\right)\right)$	0.323	0.246	0.336	0.164	0.280
Tesseract v5.5.0					
$\mathrm{E}\left(l_L\left(\mathrm{P}_i,\mathrm{P}^*\right)\right)$	0.287	0.238	0.250	0.339	0.262
$\mathrm{E}\left(l_L\left(Y_{fin},\mathrm{P}^*\right)\right)$	0.160	0.123	0.125	0.277	0.149
$\mathrm{E}\left(l_L\left(Y_{30},\mathrm{P}^*\right)\right)$	0.163	0.125	0.127	0.279	0.151

The control rule used was a simple counting rule M_N, which requires stopping the recognition process at a set step N. Furthermore, two variants of a stopping rule, described in [1], were investigated. Considering the original work depends on the use of assurance metrics for the recognition results, which are not available within the framework of the model studied in this chapter, the stopping rule described in [1] degenerates into a threshold cut-off of the size of the largest cluster of matching results of recognition collected up to step m. Therefore, two control stopping rules were constructed:

- $M_{K\mathrm{P}}$, which performs a threshold cut-off of the size of the largest cluster of matching frame-by-frame results of recognition $\rho_1,\ldots,\rho_m$,
- M_{KY}, which similarly considers the integrated recognition results $Y_1,\ldots,Y_m$.

Finally, the stopping rule (11), developed in current chapter, assessments at each step the anticipated range Δ_m to the following integrated result and stops the process when this estimate becomes less than or matching a threshold. A stopping rule M_U applies just starting from step $m=2$ (i.e., from the step where the estimation in (13) becomes more justified).

Fig. 1 illustrates the effectiveness of the stopping rules for all groups of text arias recognized using the Tesseract library (versions v4.1.1 and v5.5.0). A lower position of the curve reflects greater effectiveness of the stopping rule. It is worth noting that, on average, the presented stopping rule (11) demonstrates higher effectiveness compared to other examined methods. It is worth mentioning that the stopping method (11) exhibits high effectiveness without any modifications for the two different Tesseract softwore versions, that use varied generations of text string recognition algorithms.

The experimental results presented in the table and graphs demonstrated the effectiveness of the proposed method for optimal stopping in the process of recognizing string objects. The method significantly reduces the average error of integrated recognition results compared to single-frame results and traditional approaches such as fixed-step or threshold-based methods. For all types of data, including dates, MRZ strings, names, and document numbers, a reduction in error is observed in terms of the normalized Levenshtein metric. The graphs illustrate that the proposed method achieves lower errors while processing fewer frames, ensuring an optimal balance between accuracy and computational costs. Additionally, the method's robustness is confirmed by its consistent performance across different versions of the Tesseract software, highlighting its versatility and adaptability to various text recognition algorithms.

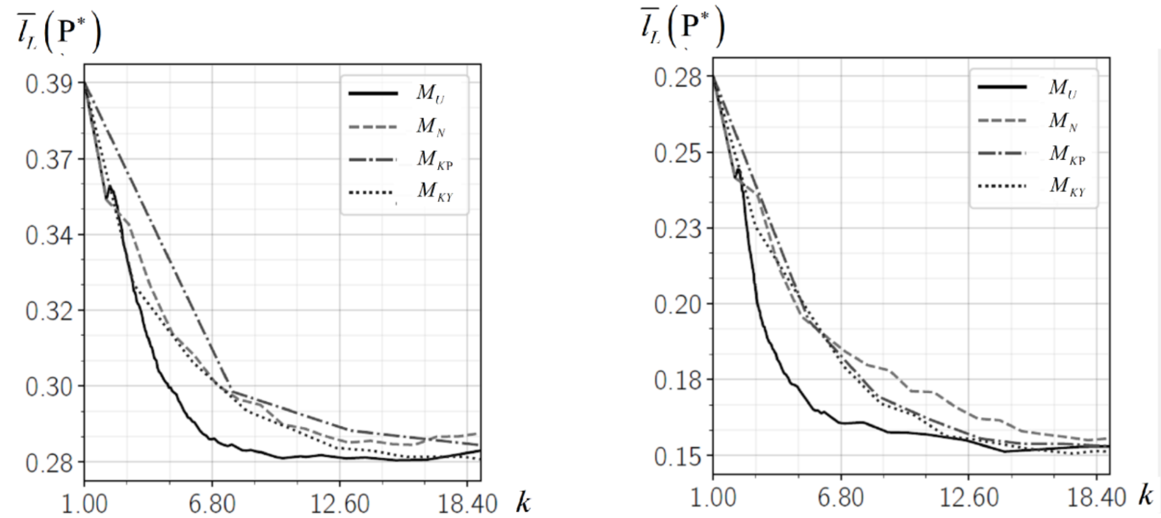

Fig. 1. Comparative study of the effectiveness of stopping rules: a graph showing the relationship among the mean range among the obtained result at the stopping point and the true value, and the mean number of analyzed frames before stopping, with varying observation cost k, at the configurable parameter value $\xi = 0.2$. The recognition of text arias was performed using the Tesseract library v4.1.1 (left) and v5.5.0 (right).

CONCLUSIONS FROM THE PRESENT STUDY AND PROSPECTS FOR FURTHER RESEARCH IN THIS AREA

This article addressed the task of stopping the process of object recognition in a real-time video stream, a novel and critical problem, especially related to developing optical recognition systems designed for mobile devices. A formal problem formulation was proposed, aligned with the classical stop problem framework, and a method was introduced that treats the object recognition process in a real-time video stream as a process, where the stopping point becomes monotonic after a certain step.

Based on the proposed method, an algorithm for stopping the string object recognition process in a real-time video stream was developed. The range among the ongoing and the following integrated results was estimated by modeling the following integrated result using the accumulated observations.

The method was experimentally tested in the task of recognizing text arias in identity certificates using the open MIDV-500 dataset and the widely accessible Tesseract text recognition library. It was demonstrated that the precented stopping rule is more effective than thresholding based on the number of analized frames or thresholding based on the size of identical results largest cluster, despite the point that confidence recognition results scores were excluded from the model.

Future research in this area offers several promising directions. The proposed method can be extended to other object types, such as images or complex patterns, and integrated with machine learning models to enhance decision-making accuracy using dynamic confidence scores. Further optimization is needed for deployment on edge devices and embedded systems with stringent constraints on computational power and energy consumption. Practical case studies in fields such as autonomous vehicles, robotics, and mobile augmented reality could validate the method's effectiveness in real-world applications. Refining the stopping algorithm through adaptive thresholding mechanisms and exploring advanced integration functions could improve flexibility and robustness. Additionally, interdisciplinary applications in areas like healthcare for real-time medical imaging and security for surveillance systems highlight the broad potential of this approach. These directions provide opportunities to advance the adaptability and practical impact of the proposed method.

References

1. Gupta A., Anpalagan A., Guan L., Khwaja A. S. Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues // Array. – 2021. – Vol. 10. – P. 100057. – doi: 10.1016/j.array.2021.100057.
2. Leszczuk M., Janowski L., Nawała J., Boev A. Objective Video Quality Assessment Method for Object Recognition Tasks // Electronics. – 2024. – Vol. 13, No. 9. – P. 1750. – doi: 10.3390/electronics13091750.
3. Luna E., San Miguel J. C., Ortego D., Martínez J. M. Abandoned Object Detection in Video-Surveillance: Survey and Comparison // Sensors. – 2018. – Vol. 18, No. 12. – P. 4290. – doi: 10.3390/s18124290.
4. Kushnir D. Methods and means for real-time object recognition accuracy increase in video images on ios mobile platform // Computer Systems and Network. – 2021. – Vol. 3, No. 1. – P. 80–88. – doi: 10.23939/csn2021.01.080.
5. Taye M. M. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions // Computers. – 2023. – Vol. 12, No. 5. – P. 91. – doi: 10.3390/computers12050091.
6. Wu Y. Optimal Stopping and Loading Rules Considering Multiple Attempts and Task Success Criteria // Mathematics. – 2023. – Vol. 11, No. 4. – P. 1065. – doi: 10.3390/math11041065.
7. Malligere Shivanna V., Guo J.-I. Object Detection, Recognition, and Tracking Algorithms for ADASs—A Study on Recent Trends // Sensors. – 2023. – Vol. 24, No. 1. – P. 249. – doi: 10.3390/s24010249.

8. Elyan E., et al. Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward // Artificial Intelligence Surgery. – 2022. – OAE Publishing Inc. – doi: 10.20517/ais.2021.15.
9. Robby G. A., Tandra A., Susanto I., Harefa J., Chowanda A. Implementation of Optical Character Recognition using Tesseract with the Javanese Script Target in Android Application // Procedia Computer Science. – 2019. – Vol. 157. – P. 499–505. – doi: 10.1016/j.procs.2019.09.006.
10. Prakisya N. P. T., Kusmanto B. T., Hatta P. Comparative Analysis of Google Vision OCR with Tesseract on Newspaper Text Recognition // Media of Computer Science. – 2024. – Vol. 1, No. 1. – P. 31–46. – doi: 10.69616/mcs.v1i1.178.
11. Paul J., et al. Digital transformation: A multidisciplinary perspective and future research agenda // International Journal of Consumer Studies. – 2024. – Vol. 48, No. 2. – doi: 10.1111/ijcs.13015.
12. Yujian L., Bo L. A Normalized Levenshtein Distance Metric // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2007. – Vol. 29, No. 6. – P. 1091–1095. – doi: 10.1109/tpami.2007.1078.
13. Chiodini S., et al. Rover Relative Localization Testing in Martian Relevant Environment // 2019 IEEE 5th International Workshop on Metrology for AeroSpace (MetroAeroSpace). – 2019. – P. 473–478. – doi: 10.1109/metroaerospace.2019.8869561.
14. Iglovikov V. ternaus/midv-500-models v0.0.2 // Zenodo. – 2020. – doi: 10.5281/ZENODO.4263532.