

Алгоритм принятия решения об останове процесса распознавания строки в видеопотоке на основе моделирования следующего интегрированного результата¹

К.Б. Булатов^{I,IV}, Е.П. Дорохов^{II,IV} и В.В. Арлазаров^{III,IV}

^I Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия

^{II} Московский авиационный институт (национальный исследовательский университет), г. Москва, Россия

^{III} Московский физико-технический институт (национальный исследовательский университет), г. Москва, Россия

^{IV} ООО «Смарт Энджинс Сервис», г. Москва, Россия

Поступила в редколлегию 5.06.2019

Аннотация—В работе рассматривается задача распознавания текстовой строки на последовательности кадров видеопотока, в которой возникает проблема останова — определения оптимального момента времени, на котором следует прекратить процесс захвата новых кадров и выдать окончательный результат. Предложен алгоритм останова процесса распознавания текстовой строки, предполагающий монотонность процесса получения интегрированных результатов, и основанный на моделировании интегрированного результата распознавания на следующем шаге процесса. Представлены результаты экспериментального исследования предложенного алгоритма, свидетельствующие о его эффективности и применимости в системах распознавания в видеопотоке.

КЛЮЧЕВЫЕ СЛОВА: распознавание в видеопотоке, мобильное распознавание, OCR, принятие решений, задачи останова

Современные мобильные устройства обладают цифровыми камерами достаточного качества, а также достаточными вычислительными ресурсами, чтобы имело смысл их использование в задачах анализа и распознавания объектов в реальном времени. В связи с этим, системы мобильного анализа и распознавания документов вызывают активный интерес исследователей в последние несколько лет [1–3].

В работах [1, 4–6] авторы показывают, что использование видеопотока в системах оптического распознавания и комбинирование покадровых результатов распознавания одного и того же объекта позволяют значительно увеличить точность результата распознавания. При этом в таких системах возникает дополнительная задача — определение момента (или кадра), на котором процесс распознавания объекта в видеопотоке следует остановить. Проблема останова особенно актуальна в контексте систем компьютерного зрения, автономно работающих на мобильном устройстве в реальном времени [1, 7, 8], в которых время получения результата не менее важно, чем его точность.

Как видно из современных публикаций по этой теме, в связи с расширением областей применимости оптического распознавания объектов в мобильных системах, представляется актуальным исследование и разработка новых методов останова процесса распознавания в видеопотоке, пригодных к использованию в задачах распознавания объектов различной природы.

¹ Работа выполнена при частичной финансовой поддержке РФФИ (проекты 17-29-03170, 17-29-03370)

Данная статья посвящена задаче останова процесса распознавания текстовой строки, как одного из важнейших объектов в системах распознавания документов. В первом разделе описана система распознавания объекта в видеопотоке и приведена постановка задачи останова. Во втором разделе описан алгоритм останова процесса распознавания текстовой строки на основе моделирования следующего интегрированного результата распознавания. В третьем разделе приведены результаты экспериментального исследования разработанного алгоритма.

1. ЗАДАЧА ОСТАНОВА В СИСТЕМЕ РАСПОЗНАВАНИЯ ОБЪЕКТА В ВИДЕОПОТОКЕ

Рассмотрим процесс распознавания объекта x в видеопотоке — т.е. на последовательности изображений, которая генерируется во времени. Пусть в каждый момент дискретного времени t доступно изображение $I_t(x)$ объекта x . Результат распознавания, учитывающий информацию, которая содержится в изображении $I_t(x)$ может быть доступен в момент времени $T(t) \geq t$.

Пусть в момент времени t происходит захват изображения $I_t(x)$, которое подается на вход модулю распознавания одиночного изображения F . Результат распознавания $F(I_t(x))$ становится доступным в момент времени $t' \geq t$ и регистрируется в модуле памяти системы. После этого происходит комбинирование результатов распознавания изображений, накопленных на текущий момент, и в момент времени $T(t) \geq t'$ происходит вывод результата распознавания $R_{T(t)}$, учитывающий информацию, которая содержится в изображениях с индексами $0, T^1(0), T^2(0), \dots, t$ (под надстрочным знаком функции $T(t)$ подразумевается не возведение в степень, а множественная композиция функций). Качество результата характеризуется близостью результата $R_{T(t)}$ к истинному значению x^* объекта x , согласно некоторой метрике. После вывода результата происходит захват очередного изображения $I_{T(t)}(x)$ и процесс захвата продолжается.

Необходимо отметить, что в такой модели усилено влияние производительности алгоритмов распознавания одиночного изображения на выход системы. Действительно, уменьшение времени $(t' - t)$, необходимого для распознавания изображения $I_t(x)$, и времени $(T(t) - t')$, необходимого для получения обновленного интегрированного результата, позволяют обработать более количество информации об объекте x за одно и то же абсолютное время (т.е. за одно и то же время с точки зрения пользователя/оператора). Помимо этого в рамках подобной системы возникает задача останова процесса распознавания — поскольку захват изображений может быть не ограничен естественным образом, в момент времени $T(t)$ возникает задача принятия решения о том, что процесс захвата следует прекратить и накопленный к текущему моменту результат $R_{T(t)}$ принять за окончательный. Схема описанной системы распознавания представлена на рисунке 1.

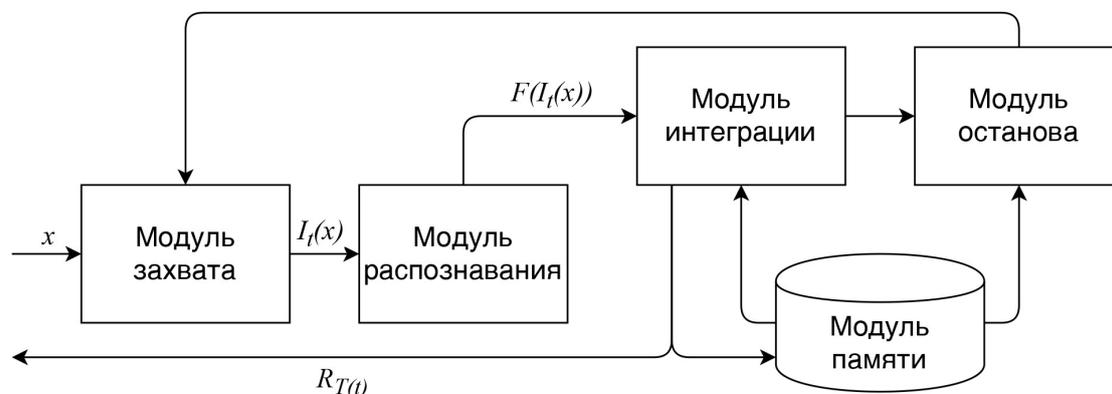


Рис. 1. Схема системы распознавания объекта в видеопотоке с остановом.

С точки зрения системной композиции процесс распознавания объектов в видеопотоке, при котором отдельные кадры распознаются независимо и комбинируются в единый результат, может рассматриваться как «anytime»-алгоритм (алгоритм с отсечением по времени: итерационный вычислительный алгоритм, который способен выдать наилучшее на данный момент решение в любое время, если процесс вычислений не доводится до естественного останова) [9]. Можно считать, что среди свойств «anytime»-алгоритмов процесс распознавания в видеопотоке обладает свойством *возможности прерывания* (interruptibility) (т.е. процесс может быть остановлен после обработки любого кадра и текущий интегрированный результат может быть принят за итоговый) и свойством *монотонности* (monotonicity) (т.е. качество интегрированных результатов в среднем не ухудшается). При этом процесс распознавания объекта в видеопотоке может не обладать свойством *определимого качества* (recognizable quality), т.е. качество текущего результата может быть неизвестно и невычислимо в момент исполнения алгоритма.

В качестве штрафного функционала системы в момент останова $t = t_{\text{stop}}$ предлагается рассматривать линейную комбинацию:

$$a \cdot \rho(R_{t_{\text{stop}}}, x^*) + b \cdot W(t_{\text{stop}}), \quad (1)$$

где a, b – константы, $\rho(R_t, x^*)$ – расстояние от интегрированного результата R_t до истинного значения объекта, характеризующее качество результата, а $W(t)$ – штрафная функция от времени. Константа a регулирует вклад стоимости ошибки распознавания, присутствующей в финальном результате, в общий функционал, в то время как константа b – вклад стоимости временных ресурсов, затраченных на распознавание объекта. Частным случаем штрафной функции $W(t)$ является количество обработанных одиночных изображений:

$$W(t) = \max\{i \mid T^i(0) \leq t\}. \quad (2)$$

Таким образом, задача останова процесса распознавания объекта в видеопотоке состоит в принятии решения в момент времени $T(t)$ о продолжении процесса или о его останове, по заданной последовательности вычисляемых интегрированных результатов распознавания объекта $R_{T^1 0}, R_{T^2 0}, \dots, R_{T(t)}$, оптимизируя функционал эффективности (1).

2. АЛГОРИТМ ПРИНЯТИЯ РЕШЕНИЯ ОБ ОСТАНОВЕ ПРОЦЕССА РАСПОЗНАВАНИЯ СТРОКИ

Пусть \mathbb{X} обозначает множество всевозможных значений результата распознавания объекта x (т.е. множество всевозможных результатов распознавания текстовой строки), и пусть на множестве \mathbb{X} задана метрика ρ . Пусть истинным значением объекта x является $x^* \in \mathbb{X}$. Процесс распознавания в видеопотоке предполагает, что наблюдается последовательность случайных результатов распознавания одиночных изображений $\mathbf{X} = (X_1, X_2, \dots)$, один результат за один шаг процесса, и каждое наблюдение $x_i = F(I_i(x)) \in \mathbb{X}$ является реализацией X_i . Будем считать, что X_1, X_2, \dots имеют одинаковое совместное распределение с x^* .

Далее, пусть определена функции интеграции нескольких результатов распознавания объекта $R : \mathbb{X}^+ \rightarrow \mathbb{X}$ (под \mathbb{X}^+ подразумевается множество всех непустых последовательностей элементов из \mathbb{X}), при помощи которой в любой момент n может быть получен интегрированный результат $R_n = R(x_1, \dots, x_n)$. Процесс может быть остановлен в любое время $n > 0$ со следующей функцией штрафа:

$$L_n = \rho(R_n, x^*) + c \cdot n, \quad (3)$$

где c – стоимость наблюдения. Данная функция штрафа является уточнением штрафного функционала модели системы распознавания объекта в видеопотоке (1) со штрафной функцией времени (2), отражающей количество обработанных изображений.

Правило останова может быть представлено как случайная величина N (случайное время останова), распределение которой зависит от входных наблюдений. Задача состоит в выборе правила останова, доставляющего минимум функционалу ожидаемого убытка:

$$V(N) = E(L_N(X_1, X_2, \dots, X_N)). \quad (4)$$

Особым классом задач останова является класс *монотонных* задач, определяемый следующим образом. Пусть A_n обозначает событие $\{L_n \leq E_n(L_{n+1})\}$ (где под $E_n(X)$ подразумевается условное математическое ожидание $E(X | X_1 = x_2, \dots, X_n = x_n)$). Задача останова называется монотонной, если выполняется $A_0 \subset A_1 \subset A_2 \subset \dots$ (т.е. если событие A_n произошло на шаге n , то соответствующие события A_{n+1}, A_{n+2}, \dots также произойдут). Для монотонных задач останова с конечным горизонтом (т.е. если существует некоторый шаг n_{\max} , на котором процесс обязан остановиться) оптимальным является «близорукое правило останова», останавливающее процесс на шаге n , если текущее значение функции убытка не превосходит ожидаемого значения убытка при останове на шаге $n + 1$:

$$N^* = \min\{n \geq 0 : L_n \leq E_n(L_{n+1})\}. \quad (5)$$

Сформулируем следующее требование к функции интеграции R : ожидаемое расстояние между двумя соседними интегрированными результатами распознавания не возрастает со временем:

$$E(\rho(R_n, R_{n+1})) \geq E(\rho(R_{n+1}, R_{n+2})) \quad \forall n > 0. \quad (6)$$

Пользуясь таким предположением о функции интеграции R , можно показать, что задача останова (4) с функцией убытка (3) становится монотонной начиная с некоторого шага. Действительно обозначим через B_n событие $\{E_n(\rho(R_n, R_{n+1})) \leq c\}$ и рассмотрим задачу останова начиная с шага n , на котором событие B_n впервые произошло. События A_n , рассматриваемые в условии монотонности, принимают следующий вид:

$$\begin{aligned} A_n : \{ \rho(R_n, x^*) + cn \leq E_n(\rho(R_{n+1}, x^*)) + cn + c \} = \\ = \{ \rho(R_n, x^*) - E_n(\rho(R_{n+1}, x^*)) \leq c \}. \end{aligned} \quad (7)$$

При фиксированном x^* , на шаге n , пользуясь неравенством треугольника, можно получить соотношение между расстоянием от текущего результата распознавания до истинного значения, ожидаемым расстоянием до результата на следующем шаге и ожидаемым расстоянием от следующего результата до истинного значения:

$$\begin{aligned} \rho(R_n, x^*) \leq E_n(\rho(R_n, R_{n+1})) + E_n(\rho(R_{n+1}, x^*)) \Rightarrow \\ \Rightarrow \rho(R_n, x^*) - E_n(\rho(R_{n+1}, x^*)) \leq E_n(\rho(R_n, R_{n+1})). \end{aligned} \quad (8)$$

Если правая часть неравенства, полученного в (8), не превышает константы c , то и левая часть также не превышает c , и, следовательно, если происходит событие B_n , то и событие A_n (7) также должно произойти. Согласно предположению (6), если событие B_n произойдет, то и событие B_{n+1} также произойдет. Таким образом, $\forall n > 0 : B_n \subset A_n \wedge B_n \subset B_{n+1}$.

Из этого следует, что начиная с шага n , на котором событие B_n произошло впервые, события $A_n, A_{n+1}, A_{n+2}, \dots$ также произойдут, а значит задача останова может рассматриваться как монотонная задача начиная с этого шага, из чего в свою очередь следует оптимальность «близорукого правила» (5) среди всех правил останова, достигающих шага n (в случае, если задача имеет конечный горизонт).

Рассмотрим теперь правило останова, предписывающее останавливать процесс распознавания объекта в случае, если произошло событие B_n :

$$N_B = \min\{n > 0 : E_n(\rho(R_n, R_{n+1})) \leq c\}. \quad (9)$$

Если правило N_B требует останова на шаге n , то и «близорукое правило» (5) потребует останова на этом шаге, а значит это решение является оптимальным. Более того, если $\rho(R_n, x^*) - E_n(\rho(R_{n+1}, x^*)) > c$, то правило N_B не останавливает процесс, также как и оптимальное правило (5). Следовательно, в случае, если предположение (6) верно, правило N_B никогда не остановится раньше времени, и, если правило требует останова, то решение об останове оптимально.

Тем самым, для построения алгоритма останова процесса распознавания текстовой строки с функцией убытка (3), будет использоваться следующий подход:

1. Оценивается ожидаемое расстояние $\Delta_n = E_n(\rho(R_n, R_{n+1}))$ от текущего интегрированного результата распознавания объекта R_n (известного на шаге n) до неизвестного следующего результата R_{n+1} ;
2. Решение об останове процесса на шаге n принимается путем порогового отсечения расстояния, оцененного в пункте 1, таким образом аппроксимируя поведение правила N_B (9).

Схема подхода представлена на рисунке 2.

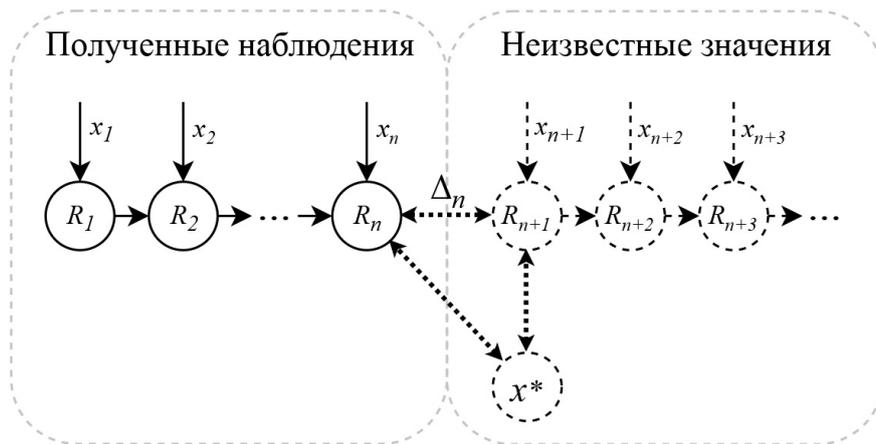


Рис. 2. Схема метода останова процесса распознавания: оценка ожидаемого расстояния от текущего интегрированного результата до следующего.

В задаче распознавания строки в видеопотоке функция интеграции R может быть реализована при помощи алгоритма ROVER [10], который используется как для комбинирования результатов распознавания одного и того же изображения при помощи нескольких систем распознавания [11], так и в задаче комбинирования результатов распознавания объекта на различных изображениях [5, 12]. В качестве метрики ρ для текстовых строк предлагается использовать нормализованное расстояние Левенштейна [13]. Для того, чтобы аппроксимировать поведение правила останова N_B (9), на n -м шаге процесса необходимо вычислять оценку ожидаемого расстояния между соседними интегрированными результатами распознавания Δ_n , имея доступ к наблюдениям $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$, а также к функции интеграции R . Для вычисления оценки предлагается провести моделирование следующего интегрированного результата, исходя из предположения, что новое наблюдение $X_{n+1} = x_{n+1}$ будет близко к уже

полученным на предыдущих шагах наблюдениям:

$$\hat{\Delta}_n \stackrel{\text{def}}{=} \frac{1}{n+1} \left(\delta + \sum_{i=1}^n \rho(R_n, R(x_1, x_2, \dots, x_n, x_i)) \right), \quad (10)$$

где δ – настраиваемый параметр.

Верхняя оценка для длины интегрированных результатов R_n и R_{n+1} составляет $O(M \cdot n)$, где $M = \max_{i=1}^n |x_i|$ – максимальная длина входной строки. Трудоемкость прямого вычисления нормализованного расстояния Левенштейна между строками X и Y составляет $O(|X| \cdot |Y|)$, а трудоемкость алгоритма ROVER для интеграции n строк с максимальной длиной M составляет $O(M^2 n^2)$. Общая трудоемкость алгоритма принятия решения об останове процесса распознавания строки на шаге n составляет $O(M^2 n^3)$. В форме псевдокода алгоритм представлен ниже как Алгоритм 1.

Алгоритм 1 Алгоритм принятия решения об останове процесса распознавания строкового объекта на основе полученных наблюдений x_1, x_2, \dots, x_n , функции интеграции R , а также внешних параметров δ и c

Require: $n > 0$ and $\forall i \in \{1, \dots, n\} : |x_i| > 0$

```

1:  $R_N \leftarrow R(x_1, x_2, \dots, x_n)$ 
2:  $\hat{\Delta}_n \leftarrow \delta$ 
3: for  $i = 1$  to  $n$  do
4:    $R_{n+1} \leftarrow R(x_1, x_2, \dots, x_n, x_i)$ 
5:    $\hat{\Delta}_n \leftarrow \hat{\Delta}_n + \rho_L(R_{n+1}, R_n)$ 
6: end for
7:  $\hat{\Delta}_n \leftarrow \hat{\Delta}_n / (n + 1)$ 
8: if  $\hat{\Delta}_n \leq c$  then
9:   return ОСТАНОВ
10: else
11:   return ПРОДОЛЖИТЬ ПРОЦЕСС
12: end if

```

3. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ

Экспериментальное исследование проводилось на открытом пакете данных MIDV-500 [14], содержащем видеоролики 50 документов, удостоверяющих личность, различных типов (по 10 видеороликов для каждого документа, по 30 кадров в видеоролике) с размеченными идеальными позициями и значениями текстовых полей. Были проанализированы 4 группы полей: даты, записанные цифрами и знаками препинания, номер документа, строки машиночитаемой зоны и компоненты имени держателя документа, записанные латинским алфавитом. Рассматривались только кадры, на которых документ целиком присутствует в кадре (следовательно, видеопоследовательности в рассматриваемом подмножестве пакета данных имели разную длину, от 1 до 30 кадров). Для того, что минимизировать эффекты нормализации и обеспечить более ясное представление результатов, каждый видеоролик был дополнен до 30 кадров путем повторения видеоролика с начала (таким образом, все анализируемые ролики имели одну и ту же длину).

Каждое поле вырезалось из исходного изображения при помощи проективного преобразования, согласно совместной разметке идеальных границ документа и координат текстового поля, с добавленными отступами, равными 10% от наименьшей стороны текстового поля. Размер вырезаемых изображений текстовых полей соответствовал разрешению 300 точек на дюйм. Каждое вырезанное текстовое поле распознавалось при помощи библиотеки распознавания с

открытым исходным кодом Tesseract (версия 4.0.0) [15], используя параметры по умолчанию для английского языка. Все сравнения значений символов проводились вне зависимости от регистра, а также латинская буква «0» считалась идентичной цифре «0».

Для оценки правила останова был построен профиль эффективности, графически показывающий зависимости среднего количества обработанных наблюдений и соответствующего среднего расстояния от полученного интегрированного результата в момент останова до истинного значения, при различных значениях стоимости наблюдения s . Подобный профиль эффективности отражает размен между временем, необходимым для обработки видеопоследовательности, и точностью полученного результата распознавания, а также позволяет визуально сравнить различные стратегии останова.

В качестве контрольного правила использовалось простое правило подсчета N_K , которое требует останавливать процесс распознавания на фиксированном шаге K . Дополнительно исследовались два варианта правила останова, описанного в [16]. Поскольку оригинальная работа опирается на использование показателей уверенности результата распознавания, которые недоступны в рамках исследуемой в данной статье модели, правило останова, описанное в [16], вырождается в пороговое отсечение размера наибольшего кластера идентичных результатов распознавания, накопленных к моменту n . Таким образом, построено два контрольных правила останова: N_{CX} , производящее пороговое отсечение размера наибольшего кластера идентичных результатов кадрового распознавания x_1, x_2, \dots, x_n , и правило N_{CR} , аналогично рассматривающий интегрированные результаты распознавания R_1, R_2, \dots, R_n . Правило останова N_B (9), реализованное при помощи Алгоритма 1, применялось начиная с шага $n = 2$. В проведенных экспериментах для вычисления оценки (10) использовалось значение параметра $\delta = 0.2$.

Сравнительные профили эффективности представлены на рисунке 3. Более низкое положение кривой отражает большую эффективность правила останова. Можно отметить, что в среднем разработанный алгоритм обладает большей эффективностью, чем другие исследованные методы.

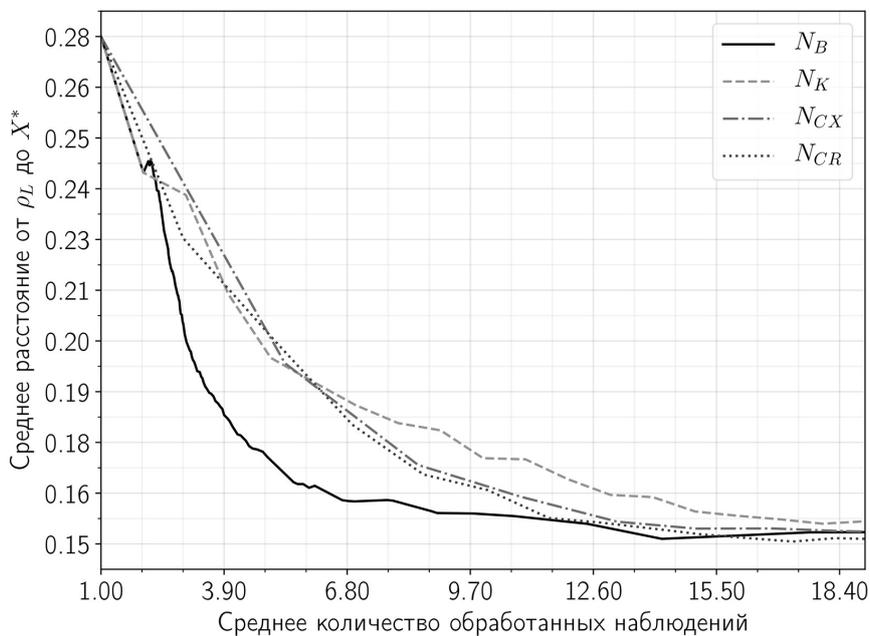


Рис. 3. Сравнительные профили эффективности правил останова.

4. ЗАКЛЮЧЕНИЕ

В работе была рассмотрена задача распознавания текстовой строки в видеопотоке и связанная с ней задача оптимального останова. Рассмотрена динамическая модель системы распознавания объекта в видеопотоке с модулем памяти и модулем принятия решения об останове процесса, и возникающая в ней постановка задачи оптимизации штрафной функции как линейной комбинации ошибки результата распознавания и времени, необходимого для его получения.

Предложен алгоритм останова процесса распознавания текстовой строки в видеопотоке, рассматривающий процесс распознавания как монотонную задачу последовательного принятия решений, моделирующий после обработки каждого кадрового результата возможное значение интегрированного результата на следующем шаге, и производящий пороговое отсечение оценки ожидаемого расстояния от текущего интегрированного результата до следующего. Предложенный алгоритм экспериментально протестирован на открытом пакете данных, содержащем видеопоследовательности текстовых полей документов, и показал свою эффективность.

СПИСОК ЛИТЕРАТУРЫ

1. Bulatov K., V. Arlazarov, T. Chernov, O. Slavin, and D. Nikolaev. Smart IDReader: Document recognition in video stream. *Proc. International Conference on Document Analysis and Recognition*, 2017, vol. 6, pp. 39–44.
2. Puybureau E., and T. Geraud. Real-time document detection in smartphone videos. *25th International Conference on Image Processing (ICIP)*, 2018, pp. 498–502.
3. Esser D., K. Muthmann, and D. Schuster. Information extraction efficiency of business documents captured with smartphones and tablets. *Proc. ACM Symposium on Document Engineering*, 2013, pp. 111–114.
4. Bulatov K., A. Lynchenko, and V. Krivtsov. Optimal frame-by-frame result combination strategy for OCR in video stream. *ICMV 2017, International Society for Optics, Photonics (SPIE)*, 2018, article no. 106961Z.
5. Булатов К.Б., Кирсанов В.Ю., Арлазаров В.В., Николаев Д.П., Полевой Д.В. Методы интеграции результатов распознавания текстовых полей документов в видеопотоке мобильного устройства. *Вестник РФФИ*, 2016, №. 4, с. 109–115.
6. Elleuch H., A. Wali, A. Samet, and A.M. Alimi. Interacting with mobile devices by fusion eye and hand gestures recognition systems based on decision tree approach. *ICMV 2016, International Society for Optics, Photonics (SPIE)*, 2017, article no. 103410D.
7. Hartl A., C. Arth, and D. Schmalstieg. Real-time detection and recognition of machine-readable zones with mobile devices. *Proc. 10th International Conference of Computer Vision Theory and Applications*, 2015, vol. 3, pp. 79–87.
8. Sourvanos N., and G. Tsatiris. Challenges in input preprocessing for mobile OCR applications: A realistic testing scenario. *9th International Conference on Information, Intelligence, Systems and Applications*, 2018, pp. 1–5.
9. Zilberstein S. Using anytime algorithms in intelligent systems. *AI Magazine*, 1996, vol. 17, pp. 73–83.
10. Fiscus J.G. A post-processing system to yield reduced word error rates: Recognizer Output Voting Error Reduction (ROVER). *IEEE Workshop on Automatic Speech Recognition and Understanding*, 1997, pp. 347–354.
11. Stuner B., C. Chatelain, and T. Paquet. LV-ROVER: Lexicon Verified Recognizer Output Voting Error Reduction. *arXiv e-prints*, 2017, arXiv: 1707.07432.
12. Wemhoener D., I.Z. Yalniz, and R. Manmatha. Creating an Improved Version using Noisy OCR from Multiple Editions. *Proc. 12th International Conference on Document Analysis and Recognition*, 2013, pp. 160–164.

13. Yujian L., and L. Bo. A normalized Levenshtein distance metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, vol. 29, no. 6, pp. 1091–1095.
14. Arlazarov V.V., K. Bulatov, T. Chernov, and V.L. Arlazarov. A Dataset for Identity Documents Analysis and Recognition on Mobile Devices in Video Stream. *arXiv e-prints*, 2018, arXiv: 1807.05786.
15. Smith R. An overview of the Tesseract OCR engine. *Proc. 9th International Conference on Document Analysis and Recognition*, 2007, vol. 2, pp. 629–633.
16. Arlazarov V., K. Bulatov, T. Manzhikov, O. Slavin, and I. Janiszewski. Method of determining the necessary number of observations for video stream documents. *ICMV 2017, International Society for Optics, Photonics (SPIE)*, 2018, article no. 106961X.

A stopping decision algorithm for string recognition in a video stream based on modelling of the next integrated result

Bulatov K.B., Dorokhov E.P. and Arlazarov V.V.

Abstract—In this paper we consider a task of text string recognition in a sequence of video frames. In such task a particular problem arises – a problem of stopping decision, i.e. determining the optimal moment at which the capturing process should be terminated and the final result should be returned. An algorithm is proposed for the stopping of the text string recognition process, which assumes monotonicity of the integrated results acquisition process, and which is based on modelling the integrated recognition results on the next process stage. Experimental results are presented for the proposed algorithm, which show its efficiency and applicability for text string recognition systems, operating in a video stream.

KEYWORDS: recognition in a video stream, mobile recognition, OCR, decision theory, stopping problems