

Распознавание образов

Накопительные контексты в задаче распознавания

В. Л. Арлазаров, А. Е. Марченко, Д. Л. Шоломов

Аннотация. Настоящая статья посвящена проблемам контекстного распознавания документа в видеопотоке. Особенностью данного подхода является наличие некоторой априорной информации, при помощи которой можно существенно повысить качество распознавания документа, записанного на бытовую видеокамеру или телефон. Использование контекста помогает получить улучшенное изображение по набору кадров, не распознаваемых в отдельности, например, в случае наличия блика, занимающего лишь небольшую часть кадра. Основными задачами, рассматриваемыми в статье, является контекстная обработка распознанного текста при наличии истории, интеграция данных видеопотока и выбор кадра с высоким качеством.

Ключевые слова: *контекстное распознавание, накопительный контекст, распознавание видеопотока, распознавание документа, интеграция видеоданных, выбор качественного кадра.*

Введение

Задача автоматического ввода текстов представляет собой задачу машинного распознавания документов, изначально предназначенных для восприятия человеком. Эти документы, как и вообще всякая информация, являются контекстозависимыми, то есть приобретают смысл только в некотором априорном контексте. Иначе говоря, воспринять информацию, содержащуюся в документе, возможно исключительно обладая некоторой предварительной информацией (например, о языке, на котором документ написан). Человеческий мозг в большинстве случаев способен применять контекстную информацию прямо на этапе распознавания, т. е. как бы модифицируя сами распознающие алгоритмы в соответствии с предполагаемым контекстом. К сожалению, современные компьютерные OCR-алгоритмы способны использовать контекстную информацию лишь ограниченного сорта — сведения о символьном наборе, шрифте, кегле и т. п. Практика показала, что для надежного ввода

документов этого недостаточно, поэтому в действующих системах применяется отдельный этап так называемой контекстной обработки результатов распознавания. Входными данными для этого этапа является поток распознанных символов (возможно, с вариантами и некоторой информацией об уверенности распознавания), а выходной информацией является окончательный вариант текста, воспринятого машиной.

1. Существующие подходы к контекстной обработке. Накопительные контекстные обработки

По своей сути контекстные обработки (или просто «контексты») используют некую априорную информацию о тексте. В самом простом и наиболее распространенном случае эта информация задается вручную при настройке системы. Например, можно

указать, на каком языке написан текст, чтобы обработчик мог ограничить множество распознанных слов известным словарем. Более сложные системы используют частотные словари слов и применяют статистические критерии для повышения надежности распознавания [1]. Еще более изощренные алгоритмы основаны на использовании статистической информации о соседствовании слов друг с другом [2, 3]. Метод N-грамм, описанный в работе [3], применяют также для восстановления текста на уровне отдельного слова — в случае, когда использовать полный частотный словарь не представляется возможным.

Для разбора документов сложной структуры применяются различные способы описания внешнего вида и (или) синтаксиса документа — строго заданного [4], частично заданного [5] или заданного вероятностно [6].

Однако данные системы не ориентированы на автоматическую коррекцию контекстной информации в ходе самого распознавания.

В [7] описывается подход, использующий несколько априори заданных частотных словарей («топиков»), и предполагающий, что частотное распределение слов в распознаваемом тексте представляет собой линейную комбинацию распределений слов в топиках. Коэффициенты этой линейной комбинации высчитываются исходя из предварительного анализа распознанного (некорректированного) текста.

В [8] обсуждаются методы адаптивного вычисления пороговых величин в задачах распознавания.

Система, автоматически подстраивающая механизм распознавания символов на основании предварительного анализа документа, описана в [9].

Публикации [7], [8] и [9] представляют частные случаи использования OCR-системами для самонастройки информации из исходного изображения либо из распознанного текста. В данной работе нас интересуют общие методы построения подобных систем, а именно методы построения контекстов, извлекающих информацию из текста, распознанного ранее. Именно такие контексты называются накопительными контекстными обработками. Накопительные обработки полагаются на то, что некоторые параметры текста на протяжении определенного массива входных данных остаются неизменными. Среди этих параметров присутствуют те, которые можно оценить, располагая некоторым объемом входных данных. В абсолютном большинстве случаев, чем больше объем входных данных, тем более точную оценку того или иного параметра возможно получить. А предполагаемый объем входных данных зависит от того, какого типа информация поступает на вход системы распознавания.

2. Типы входных данных OCR-системы

Чтобы понять, какие именно параметры возможны, а также имеет смысл извлекать и применять вышеописанным образом, введем классификацию входных данных системы автоматического ввода документов.

Самым простым и очевидным видом входных данных системы машинного чтения является единичный документ. Возможности применения накопительных контекстов в этом случае ограничены, так как характеристики информации на протяжении одного документа, как правило, значительно меняются (информацию разного рода обычно помещают в разных частях документа). Например, товарная накладная состоит из заголовка («шапки»), сведений о продавце и покупателе, таблицы товаров с количеством, ценами и стоимостью, и нижней части с подписями и печатями. Эти части накладной существенно разнородны, т. е. отличаются друг от друга по большинству характеристик, начиная от словаря и кончая стилем выравнивания.

Накопительные контексты можно применять и для единичных документов, например, собирая сведения о качестве печати и сканирования, а также о шрифте. Такого рода контексты встроены в OCR-алгоритмы и успешно применяются в современных системах.

Второй вид входных данных — массив (поток) различных документов. Сходство между отдельными документами во многих случаях значительно больше, нежели между частями одного документа. Действительно, в потоке, как правило, содержатся документы ограниченного числа типов (нередко — только одного типа), состоящие из частей с идентичными или близкими характеристиками. Накопительные контексты в данном случае можно успешно применять для сбора информации и применения ее для обработки последующих документов.

Третий вид входных данных — массив изображений одного и того же документа. Такие данные получаются, например, при захвате видеопотока видеокмеры. Ясно, что сходство между отдельными кадрами в данном случае максимально (меняться могут только характеристики оцифровки). Применение накопительных контекстов здесь более чем оправдано, так как такой поток обладает наибольшим числом неизменных свойств — идентичен весь распознанный текст, равно как и расположение его на странице, шрифты и качество печати. Накопительные контекстные обработки потока изображений одного документа ниже будут рассмотрены подробнее.

Четвертый тип входных данных объединяет второй и третий — это массив изображений, состоящий из подмассивов, каждый из которых содержит разные изображения одного документа, причем грани-

цы подмассивов OCR-системе заранее неизвестны. Такие потоки так же получаются при захвате данных с камеры, если перед ней последовательно помещают серию документов, каждый из которых необходимо идентифицировать и распознать. На данных такого вида полностью применимы контекстные обработки для данных второго типа (массивов разных документов). Обработчики же для данных третьего типа в этих задачах можно применять после внесения в них определенных изменений, о которых пойдет речь ниже.

Как отмечено выше, третий и четвертый типы входных данных характерны для информации, полученной с камер. Особенности такого рода изображений стоит рассмотреть подробнее.

3. Характеристики изображений, полученных с видеокамер

Задачи распознавания текстов на первом этапе развития рассматривали почти исключительно изображения, оцифрованные при помощи сканеров. Достоинством такого способа оцифровки является высокое качество получаемого изображения. Типичные искажения картинки, вносимые сканерами, хорошо изучены за время существования OCR-систем (см., например, [10]). Цифровые камеры как возможные устройства ввода первоначально не рассматривались по причине, во-первых, малой распространенности, а, во-вторых, неудовлетворительного для задач распознавания текстов соотношения цены устройства к качеству картинки.

Со временем ситуация изменилась, наличие камер стало типичным, а качество их поднялось достаточно, чтобы говорить о возможности работы с ними OCR-программ.

Однако массовое внедрение этих устройств подняло ряд технических проблем, не характерных для сканеров.

В первую очередь — более низкое качество изображения. Даже самые совершенные современные камеры не достигают качества сканера, что отчасти связано с некоторыми непреодолимыми обстоятельствами — например, нестабильность освещения при применении камеры, а также неизбежность проективных искажений при отсутствии специальной фиксации документа перед объективом. Кроме того, картинка, захваченная с видеокамеры, обычно значительно более зашумленная, чем отсканированная (это связано, главным образом, с невозможностью применить специальное освещение). В довершение всего, разрешение матрицы камер, даже стандарта HD, в несколько раз меньше разрешения типичного сканера, даже домашнего. О практических следствиях всех этих проблем можно прочитать в [11].

Однако данные с камер обладают одним важным преимуществом, по сравнению со сканерами — камеры дают видеопоток, т. е. не одну картинку, а серию изображений одного и того же документа. Следовательно, используя дополнительный объем информации, можно добиться более надежного распознавания.

Некоторые методы обработки различных изображений одного объекта освещены в [12] и [13]. Однако эти алгоритмы предназначены для работы с «сырыми», нераспознанными изображениями, т. е. относятся к области предобработок.

Другой способ использовать информацию из разных кадров — применить накопительные контексты. А именно, с каждым кадром накапливать данные о документе, постепенно повышая качество распознавания. Работа с массивом изображений вместо одного позволяет, в частности, избежать случайных ошибок OCR-системы — т. е. ошибок в единичных символах при в целом стабильном качестве. Однако подобного рода ошибки не так уж часты в современных распознающих модулях. Гораздо типичнее нестабильность, связанная с нестабильностью качества входных кадров.

Самая распространенная проблема отдельного кадра — нерезкое изображение, т. е. неверная фокусировка. В устройствах с автоматической фокусировкой эта проблема может проявляться для отдельных кадров или серий кадров. Аналогичный эффект возникает при резких перемещениях документа (изображение смазывается). Задача накопительных контекстов — используя информацию обо всех кадрах, выделить среди них нерезкие и исключить из рассмотрения (или, как минимум, понизить их вес при учете). Примеры см. на рис. 1 и 2.

Также при работе с камерами постоянно встает проблема бликов. Многие документы являются ламинированными, а также обладают определенными средствами защиты, которые представляют собой объекты, видимые только под определенным углом и (или) в свете определенного спектрального состава. Блики часто попадают на распознаваемую область, а защитные изображения нередко совмещены с текстом, однако проявляются все эти помехи не под каждым углом, т. е. только в определенных кадрах. Пример можно посмотреть на рис. 3.

На рис. 3 (а) четко виден номер 2515600058, но код RUS и последующая дата 910529 трудночитаемы из-за защитной надписи «РФ». На рис. 3 (б) наоборот, надпись «РФ» ушла из угла видимости, зато номер документа попал под блик и стал нечетким.

Накопительным контекстам в данном случае придется принимать решения не обо всем документе, а о каждом отдельном поле или даже символе, полагаясь на то, что для каждой распознаваемой области в видеофрагменте найдутся кадры, в которых она достаточно хорошо читаема.



а)



б)

Рис. 1. Изображения из одного видеопотока, не в фокусе (а) и в фокусе (б)

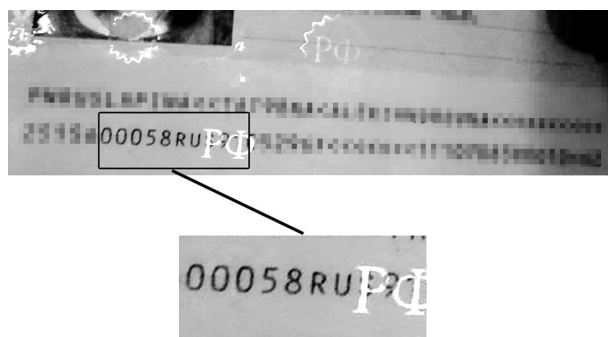


а)

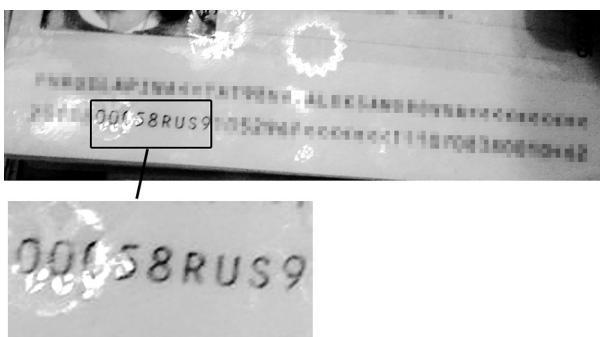


б)

Рис. 2. Изображения из одного видеопотока, смазанное (а) и не смазанное (б)



а)



б)

Рис. 3. Изображения из одного видеопотока — на разных кадрах блики затрудняют чтение разных символов

4. Типы контекстных параметров

По определению, накопительные контекстные обработки извлекают из распознанного текста некую информацию, которую впоследствии используют для обработки последующего текста. Семантическую единицу этой информации назовем контекстным параметром. Для дальнейшего рассмотрения выделим несколько типов контекстных параметров.

5. Непрерывно-метрические контекстные параметры

Пусть S — множество исходных данных контекстной обработки (множество результатов OCR-модуля). Обозначим как D выходное множество контекстной обработки (множество финальных распознанных текстов).

Обозначим саму контекстную обработку как функцию $c_{tx} : S \rightarrow D$.

Введем также функцию, ставящую в соответствие результат оптического распознавания и текст, который в действительности был напечатан («идеальный» результат): $c_{id} : S \rightarrow D$.

Введем функцию расстояния (метрики) на множестве D :

$$d_{ist} : D \times D \rightarrow R.$$

Семантика этой функции зависит от поставленной задачи. Например, для многих задач хорошо подходит расстояние Левенштейна.

Допустим, параметр p функции c_{tx} задан на метрическом пространстве P , и ρ_p — метрика этого пространства.

Непрерывно-метрическим назовем параметр с множеством значений P , удовлетворяющий условию:

$$\begin{aligned} \forall p \in P, s \in S, \varepsilon > 0 \exists \delta > 0 \forall p_1 : \rho_p(p, p_1) < \delta \rightarrow \\ \rightarrow |d_{ist}(c_{tx}(p, s), c_{id}(p, s)) - \\ - d_{ist}(c_{tx}(p_1, s), c_{id}(p_1, s))| < \varepsilon. \end{aligned} \quad (1)$$

То есть, формулируя неформально, малые отклонения непрерывно-метрического параметра дадут малые различия в качестве контекстной обработки (если под качеством понимать дистанцию до оригинального текста).

Пример непрерывно-метрического параметра: доля несловарных слов в тексте.

6. Сепарированные параметры

Используя обозначения из предыдущего пункта, назовем *сепарированным* контекстный параметр, удовлетворяющий условию:

$$\begin{aligned} \exists \delta > 0 \forall p_1, p_2 \in P : p_1 \neq p_2 \exists s \in S \rightarrow \\ \rightarrow |d_{ist}(c_{tx}(p_1, s), c_{id}(p_1, s)) - \\ - d_{ist}(c_{tx}(p_2, s), c_{id}(p_2, s))| \geq \delta. \end{aligned} \quad (2)$$

Иначе говоря, для любой пары различных значений параметра соответствующие результаты контекстной обработки всегда будут достаточно далеки друг от друга.

Пример сепарированного параметра: число символов в строке или в поле.

7. Методы оценки для параметров разного типа

Процесс получения оценки контекстного параметра назовем *интеграцией* (на том основании, что, собирая информацию на массиве входных данных, программный модуль в результате выдает единственное, интегрированное значение).

Алгоритмы интеграции принимают наиболее простой вид на входных данных третьего типа (массив разных изображений одного документа). Ограничимся в этом разделе рассмотрением методов для данных указанного типа.

Большинство нижеприведенных методов используют оценку достоверности того или иного параметра в отдельном изображении, которую мы назовем *частной оценкой*. Эта оценка строится на основе существующих алгоритмов определения достоверности результатов распознавания (см., например, [14]).

Наиболее очевидным способом интеграции непрерывно-метрического параметра является усреднение. А именно, извлекая оценочное значение параметра из каждого изображения массива, мы затем вычисляем среднее арифметическое этих значений, и объявляем его *интегральной* оценкой параметра:

$$p_{itg} = \sum_{i=1}^n p_i / n, \quad (3)$$

где p_{itg} — интегральная оценка, p_i — частная оценка для i -го изображения, n — число изображений.

Немного более сложным является случай, когда вместе с оценкой параметра из изображения извлекается достоверность этой оценки. В этой ситуации нужно прибегнуть к взвешенному усреднению:

$$p_{itg} = \frac{\sum_{i=1}^n p_i \cdot c_i}{\sum_{i=1}^n c_i}, \quad (4)$$

где c_i — достоверность частной оценки p_i .

Для параметров, совершающих редкие шумовые скачки большой амплитуды, хорошо подходит интеграция путем нахождения медианы. Пусть L и H — множества индексов частных оценок, такие, что

$$\begin{aligned} \forall i \in L \rightarrow p_i \leq p_{lig}, \\ \forall i \in H \rightarrow p_i > p_{lig}. \end{aligned} \quad (5)$$

Тогда p_{lig} будет называться медианной оценкой если и только если

$$\|L\| - \|H\| \leq 1 \quad (6)$$

($|X|$ обозначает мощность множества X).

Медианную оценку можно обобщить и на множество оценок с достоверностями. В этом случае условие (6) примет вид:

$$|w(L) - w(H)| \rightarrow \min_{p_{lig}}, \quad (7)$$

где

$$w(X) = \sum_{i \in X} p_i \times c_i. \quad (8)$$

Иногда необходимо получить оценку, близкую не к середине распределения, а к одному из краев, но, тем не менее, исключая редкие большие выбросы. В таком случае применимы квантили множества с уровнем, отличным от 0,5 (квантиль уровня 0,5 является медианой). Для таких оценок условие (6) принимает вид:

$$\left| \frac{\|L\|}{\|H\|} - \alpha \right| \rightarrow \min_{p_{lig}}, \quad (9)$$

где α — уровень квантиля.

Для оценок с достоверностями, соответственно, получаем

$$\left| \frac{w(L)}{w(H)} - \alpha \right| \rightarrow \min_{p_{lig}}. \quad (10)$$

Совсем иным образом обстоит дело для сепарированных параметров. Как ясно из определения, для сепарированных параметров не имеют смысла никакие операции сравнения и вычисления расстояния, кроме собственно равенства. Поэтому как усреднение, так и медианная оценка в данном случае неприменимы. Разумным методом интеграции сепарированного параметра является нахождение пика плотности распределения. Пусть $I_v = \{i | p_i = v\}$. Тогда p_{lig} определяется выражением:

$$|I_{p_{lig}}| \rightarrow \max_{p_{lig}}. \quad (11)$$

На оценки с достоверностями это условие обобщается следующим образом:

$$w(I_{p_{lig}}) \rightarrow \max_{p_{lig}}, \quad (12)$$

где

$$w(I) = \sum_{i \in I} c_i. \quad (13)$$

8. Обобщение методов интеграции на входные данные других типов

Принципиальной особенностью данных четвертого типа является наличие в потоке данных разных документов. Так как на разных документах интегрируемый параметр может принимать разные значения, напрямую использовать вышеприведенные методы на потоках 4-го типа нельзя.

Заметим, однако, что чем ближе друг к другу в потоке расположены изображения, тем выше вероятность, что они принадлежат одному документу. Соответственно, возникает идея применить к частным оценкам весовую функцию, имеющую максимум в исследуемом кадре и затухающую в прямом и обратном направлении от центра (данному условию отвечает, например, гауссова функция). Ширину ветвей функции необходимо подбирать исходя из типичной величины однодокументного подмассива и оценки погрешности частных оценок.

Поскольку все вышеприведенные методы обобщены нами на случай оценок с достоверностями, весовую функцию можно заложить в значения достоверностей, домножив каждую из достоверностей на значение весовой функции в данном кадре.

В реальных задачах распознавания часто возникает необходимость распознавать потоки «на лету», т. е. давать ответ, не имея возможности обратиться к будущим кадрам. Вышеприведенный метод без изменений работает и в этом случае: будущие кадры просто исключаются из суммы; иначе говоря, весовая функция теряет свой положительный «хвост», приравниваясь к нулю на всех значениях справа от текущего кадра.

Недостатком метода в приведенном виде является то, что для вычисления интегрируемого параметра в каждом кадре придется выполнять большой объем вычислений, так как для расчета текущей интегральной оценки используется весь массив предшествующих частных оценок. Хотелось бы найти более эффективный способ с меньшим числом операций на каждом шаге. Такие способы есть для каждого из методов оценки, приведенных в предыдущем параграфе.

Заметим, что для метода усреднения, в случае, когда весовая функция не применяется (т. е. во всех точках равна единице), нет необходимости выполнять весь объем вычислений для каждого кадра. Достаточно сохранять на каждом шаге числитель

и знаменатель выражения (4) (обозначим их, соответственно, g_k и h_k), а для следующего шага добавить к числителю $p_{k+1} \cdot c_{k+1}$, а к знаменателю — c_{k+1} , где k — индекс текущего кадра. К сожалению, для неконстантной весовой функции этот способ не подходит. Однако можно попробовать найти затухающую весовую функцию, которая упростит процесс вычислений. В этом качестве можно предложить функцию экспоненты. Эта функция не обладает пиком в середине и двумя затухающими «хвостами», однако если система не может обращаться к будущим кадрам, «положительный хвост» функции веса не нужен. То есть весовые коэффициенты будут иметь вид:

$$v_i = e^{\alpha \cdot (k-i)} \quad (14)$$

или

$$v_i = a^{(k-i)}, \quad (15)$$

где $i \leq k$.

Тогда интегральная оценка для k -го кадра вычисляется по рекуррентной формуле:

$$p_{\text{итг},k} = \frac{g_{k-1} \cdot a + p_k \cdot c_k}{h_{k-1} \cdot a + c_k}. \quad (16)$$

Эффективный алгоритм интеграции параметра с оценкой на основе квантилей оказывается немного сложнее. Необходимо на каждом шаге поддерживать условие, налагаемое на множества L и H (в самом общем случае — условие (10)). Для этого на каждом шаге может потребоваться перемещение нескольких элементов из L в H или наоборот. Заметим, что, поскольку любая оценка с индексом из L не больше любой оценки с индексом из H , то перемещать из L в H можно только элементы L , соответствующие наибольшим оценкам (а из H в L , соответственно, только элементы H , соответствующие наименьшим оценкам). Известный способ быстрого нахождения наибольшего или наименьшего элемента массива — организация элементов в виде пирамиды (кучи). Если постоянно поддерживать две пирамиды, одна из которых содержит оценки «нижнего» множества, а другая — «верхнего», то интегральная оценка будет всегда известна. Поддержание пирамидальных структур будет занимать порядка $\ln k$ операций для каждого кадра. Обозначим пирамиду оценок с индексами из L как P_L , а с индексами из H — как P_H . Кроме самих пирамид нужно также хранить и поддерживать значения их весов, $w_H = w(H)$ и $w_L = w(L)$.

Таким образом, вычисление оценки для кадра k будет происходить по следующему алгоритму:

1. Добавить частную оценку p_k с весом c_k в P_L или P_H , соблюдая условие (5). Добавить ее вес к w_H или w_L соответственно.

2. Вычислить $r = \left| \frac{w_L}{w_H} - \alpha \right|$.

3. Переместить из P_L в P_H наибольший элемент P_L . Вычесть его вес из w_L и добавить к w_H . Вычис-

- лит $r' = \left| \frac{w_L}{w_H} - \alpha \right|$.

4. Если $r' < r$, перейти к шагу 2. Иначе перейти к следующему шагу.

5. Вычислить $r = \left| \frac{w_L}{w_H} - \alpha \right|$.

6. Переместить из P_H в P_L наименьший элемент P_H . Вычесть его вес из w_H и добавить к w_L . Вычис-

- лит $r' = \left| \frac{w_L}{w_H} - \alpha \right|$.

7. Если $r' < r$, перейти к шагу 5. Иначе завершить алгоритм.

Чтобы учесть влияние экспоненциальной весовой функции, можно было бы на каждом шаге домножать веса всех элементов P_H и P_L на a , однако это ухудшит время работы алгоритма с $\ln k$ до k . Вместо этого можно вес каждой новой частной оценки перед добавлением умножать на a^k .

9. Выбор достоверного кадра

Существенной особенностью алгоритмов интеграции на входных потоках четвертого типа является наличие «краевых» эффектов — на кадрах, близких к границе между документами, оценки параметров становятся недостоверными. Это происходит вследствие попадания под интегрирующую функцию значений параметра для двух смежных документов. Распознавание в целом на таких кадрах тоже будет недостоверным.

Необходимо уметь отличать такие недостоверные кадры, чтобы исключить их из рассмотрения.

В предлагаемом ниже методе решения поставленной задачи используется оценка ответа типичной современной OCR-системы. Результат работы системы оптического распознавания текста обычно представлен в виде массива литер, а каждая литера представляет собой так называемый вектор альтернатив, т. е. массив пар $\langle l_i, c_i \rangle$, где l_i — код распознанной буквы, а c_i — вероятность нахождения данной буквы в данной позиции. Как правило, пары в векторе альтернатив отсортированы в порядке убывания c_i . Простейшую оценку достоверности распознавания литеры можно определить по формуле

$$c_{\text{итр}} = c_1 - c_2. \quad (17)$$

При обработке входных потоков четвертого типа векторы альтернатив обычно подвергаются интеграции по одному из предложенных выше методов. Точнее, для каждой буквы l интегрируется соответствующая ей вероятность c . Таким образом, на каждом шаге мы имеем интегрированный вектор альтернатив, для которого можно посчитать достоверность, например, по формуле (17).

Теперь достаточно подобрать пороговое значение c_{thr} для интегрированного c_{lir}^{ig} , и исключить из рассмотрения кадры, где $c_{lir}^{ig} < c_{thr}$. Для определения c_{thr} воспользуемся следующим неформальным принципом: текущий интегрированный результат распознавания следует считать достоверным, если он гарантированно не изменится при добавлении к интегрированной последовательности новых кадров того же документа. Рассмотрим, в каком случае мы можем гарантировать указанную неизменность.

Допустим, интеграция вектора альтернатив происходит методом усреднения с экспоненциально затухающей весовой функцией. Предположим идеальный случай: на вход OCR-системы поступает изображение одного и того же документа, а на выходе система дает идентичные для каждого кадра векторы альтернатив. Поскольку векторы идентичны, c_{lir} для каждого кадра постоянно. Как будет меняться при этом c_{lir}^{ig} ? Учитывая то, что весовая функция является экспонентой, c_{lir}^{ig} будет меняться по закону суммы членов геометрической прогрессии. Как известно, сумма членов бесконечно затухающей геометрической прогрессии равна

$$S_{\infty} = \frac{b_1}{1-q},$$

где b_1 — первый член прогрессии, q — ее знаменатель. В случае весовой функции с коэффициентом затухания a , для интегральной достоверности распознавания получим:

$$c_{lir}^{ig} \rightarrow \frac{c_{lir}}{1-a}.$$

Это и есть тот максимум, к которому будет стремиться интегральная достоверность. Поэтому достоверным можно считать всякий результат, для которого c_{lir}^{ig} будет достаточно близко к указанному числу:

$$c_{thr} = \frac{c_{lir}^{typ}}{1-a} \cdot f, \quad (18)$$

где c_{lir}^{typ} — типичное значение достоверности для отдельного кадра, f — некий коэффициент, близкий к единице, строго меньший единицы.

c_{lir}^{typ} можно вычислить, интегрируя c_{lir} одним из предложенных выше способов. Отличие между c_{lir}^{typ} и c_{lir}^{ig} состоит в том, что c_{lir}^{typ} представляет собой интеграцию c_{lir} , вычисленных для каждого отдельного неинтегрированного вектора альтернатив, а c_{lir}^{ig} — вычислено на побуквенно интегрированном векторе альтернатив (т. е. в первом случае сначала считаем, потом интегрируем, а во втором — сначала побуквенно интегрируем, потом считаем).

В некоторых случаях в системах распознавания нет требования максимальной надежности в определении достоверности распознавания, зато есть требование к скорости системы: максимальное число кадров, после которого система должна дать результат. В этом случае порог c_{thr} можно понизить, введя дополнительный параметр n_f — число кадров, достаточных для срабатывания системы. Принцип для определения c_{thr} тогда будет формулироваться следующим образом: текущий интегрированный результат распознавания следует считать достоверным, если он гарантированно не изменится при добавлении к интегрированной последовательности не более n_f кадров того же документа.

Воспользовавшись формулой частичной суммы геометрической прогрессии, получим:

$$c_{thr} = \frac{c_{lir}^{typ} \cdot (1 - a^{n_f})}{1 - a}. \quad (19)$$

Итоги

Вышеописанные алгоритмы были применены в реальных системах оптического распознавания документов, в частности, в системе автоматизированного ввода машинно-читаемых паспортов.

Результаты работы накопительных контекстов на стенде из около 6 тысяч изображений проиллюстрированы в следующей сводной таблице:

	Интеграция отключена	Интеграция включена, экспозиция одного документа 3 кадра	Интеграция включена, экспозиция одного документа 5 кадров	Интеграция включена, экспозиция одного документа 10 кадров
Доля правильно распознанных символов	99,01 %	99,65 %	99,73 %	99,81 %
Доля правильно распознанных документов	58,76 %	81,80 %	86,08 %	90,33 %

Как видно, качество распознавания делает резкий скачок вверх уже при малом числе кадров экспозиции одного документа, и далее заметно растет при увеличении экспозиции. В целом, применение накопительных контекстов позволило уменьшить число ошибок в 5–6 раз.

Таким образом, статистика стендирования полностью подтверждает высокую эффективность накопительных контекстов для задач распознавания видеопотоков изображений.

Литература

1. *Xian Tong, David A. Evans.* A Statistical Approach to Automatic OCR Error Correction In Context // *Proceedings of the Fourth Workshop on Very Large Corpora (WVLC-4 (1996))*.
2. *Youssef Bassil; Mohammad Alwani.* OCR Context-Sensitive Error Correction Based on Google Web 1T 5-Gram Data Set // *American Journal of Scientific Research*, Issue. 50, February 2012.
3. *Бузикашвили Н. Е., Крылова Г. А., Самойлов Д. В.* N-граммы в лингвистике // *Методы и средства работы с документами: Сборник трудов Института системного анализа Российской академии наук / Под ред. д-ра техн. наук, проф. В. Л. Арлазарова и д-ра техн. наук, проф. Н. Е. Емельянова. М.: URSS, 2000.*
4. *Шоломов Д. Л.* Синтаксический подход к пост-обработке распознанного текста // *Документооборот. Концепции и инструментарий. М.: URSS, 2004.*
5. *Шоломов Д. Л., Постников В. В., Марченко А. А., Усков А. В.* Пост-обработка результатов OCR распознавания, использующая частично определенный синтаксис // *Интеллектуальные информационные технологии. Концепции и инструментарий: Т. 16. М.: КомКнига/URSS, 2005.*
6. *Бузикашвили Н. Е.* Стохастические грамматики с отсечением. // *Методы работы с документами: Сборник трудов Института системного анализа Российской академии наук / Под ред. д-ра техн. наук, проф. В. Л. Арлазарова и д-ра техн. наук, проф. Н. Е. Емельянова. М.: URSS, 2000.*
7. *Michael Wick, Michael G. Ross, Erik G. Learned-miller.* Context-Sensitive Error Correction: Using Topic Models to Improve OCR // *International Conference on Document Analysis and Recognition — ICDAR, 2007.*
8. *Ray Smith, Chris Newton, Phil Cheatle.* Adaptive thresholding for OCR: a significant test. *Hewlett-Packard Laboratories, 1993.*
9. *Кляцкин В. М., Котович Н. В., Славин О. А.* Многопроходная схема распознавания документов с обучением. // *Управление информационными потоками / Сборник трудов Института системного анализа Российской академии наук. Под ред. д. т. н., проф. В. Л. Арлазарова, д. т. н., проф. Н. Е. Емельянова. М.: URSS, 2002.*
10. *Титов Ю. В.* Об искажении символов при сканировании. // *Системный подход к управлению информацией: Т. 23. М.: КомКнига/URSS, 2006.*
11. *Славин О. А., Федоров Г. О.* Вопросы распознавания текста, оцифрованного с помощью видеокамер // *Управление информационными потоками / Сборник трудов Института системного анализа Российской академии наук. Под ред. д. т. н., проф. В. Л. Арлазарова, д. т. н., проф. Н. Е. Емельянова. М.: URSS, 2002.*
12. *Титов Ю. В.* О восстановлении идеального прообраза по коллекции образов. // *Системный подход к управлению информацией: Т. 23. М.: КомКнига/URSS, 2006.*
13. *Славин О. А.* О способах суммирования бинарных образов // *Обработка информационных и графических ресурсов / Под ред. В. Л. Арлазарова. М.: Красанд/URSS, 2010.*
14. *Арлазаров В. В., Кляцкин В. М.* Определение достоверности результатов распознавания символа в системе Cognitive Forms // *Документооборот. Концепции и инструментарий. М.: URSS, 2004.*

Арлазаров Владимир Львович. Зав. лабораторией ИСА РАН. Чл.-корр. РАН, д. т. н. Окончил в 1961 г. МГУ. Кол-во печатных работ: более 200. Область научных интересов: методы и системы интеллектуального анализа данных и распознавание образов. E-mail: arl@cs.isa.ru

Марченко Алексей Евгеньевич. Разработчик ООО «Когнитивные технологии». Окончил в 2002 году МФТИ. Кол-во печатных работ: 5. Область научных интересов: распознавание образов. E-mail: alexey@cognitive.ru

Шоломов Дмитрий Львович. С. н. с. ИСА РАН. К. т. н. Окончил в 1997 году МГУ. Количество печатных работ: 16. Область научных интересов: распознавание образов. E-mail: sholomov@list.ru