Модель системы распознавания объектов в видеопотоке мобильного устройства*

В.В. Арлазаров[I,III], К.Б. Булатов[II,III], А.В. Усков[II]

[I] Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, г. Москва, Россия
[II] Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук, г. Москва, Россия
[III] ООО «Смарт Энджинс Сервис», г. Москва, Россия

Аннотация. В работе исследована задача автоматического распознавания объектов с использованием видеопотока в качестве цифрового образа. Рассматриваются варианты формализации системы распознавания в видеопотоке, обсуждаются свойства динамической модели системы распознавания. Описываются задача интеграции результатов распознавания объекта и задача останова, возникающие в системе распознавания со временем, в отсутствии естественного ограничения на количество входных изображений. Представлены формальные постановки задач интеграции и останова в рамках модели взаимодействия системы распознавания с пользователем.

Ключевые слова: *распознавание образов, видеопоток, мобильные устройства, системы распознавания, OCR.*

DOI: 10.14357/20790279180508

Введение

Внедрение технологических, социальных и коммерческих процессов, основанных на использовании мобильных устройств и технологий, в условиях современного мира уже является обыденностью. Системы технического зрения с использованием мобильных технологий, к примеру, системы автоматического ввода и анализа документов на мобильных устройствах продолжают вытеснять традиционные стационарные системы. Развитие технологий технического зрения с применением мобильных устройств в условиях аппаратных ограничений, связанных с ними, становится все более актуальной задачей.

Классические системы распознавания и автоматического ввода предполагают использование сканированного изображения или фотографии объекта в качестве его оцифрованного представления. При использовании мобильных устройств для оцифровки образов распознаваемых объектов возникает дополнительная возможность использовать видеопоток цифровой камеры помимо отдельных фотографий или кадров. Процесс фотографии объекта при помощи современных мобильных устройств предполагает этап «наведения» оператором объектива камеры на объект с отображением кадров видеопотока на экране устройства в реальном времени для контроля оператора. В случае, если обработка изображения производится с одного изображения, информация, которая содержится в захваченных предварительных кадрах используется лишь косвенно (оператором). При рассмотрении цельного видеопотока в качестве цифрового образа объекта появляется возможность использовать гораздо больше визуальной информации [1].

Рис. 1. Процесс съемки идентификационного документа при помощи мобильного устройства (в качестве документа используется макет идентификационной карты Германии)

Использование видеопотока позволяет решать задачи, недоступные для решения при анализе одиночной фотографии. Внешние условия съемки могут привести к тому, что распознавае-

* Работа выполнена при частичной финансовой поддержке РФФИ (проекты №№ 17-29-03170, 17-29-03263).

мый объект сильно искажен на одиночном изображении [2]. Примером является блик от протяженного источника света, проявляющийся на глянцевой поверхности плоского объекта (см. рис. 1) Поскольку в видеопотоке геометрическое положение снимаемого объекта, как правило, меняется между кадрами, блик также «сдвигается», что позволяет получить информацию о скрываемом объекте на другом кадре видеопотока. Существуют также важный класс объектов, детектирование и распознавание которых невозможно на одиночных снимках – к примеру, голографические элементы защиты, которые на единичных изображениях могут быть неотличимы от бликов или рисунков [2].

В таких условиях возникает задача выбора оптимальной стратегии комбинирования результатов покадрового распознавания. Данная задача в литературе практически не описана, и наиболее близкий спектр методов касается задачи комбинирования результатов распознавания одного и того же объекта, но разными классификаторами [3-5]. Помимо базовых стратегий объединения оценок в работах, затрагивающих гетерогенные методы объединения результатов классификаторов, рассматриваются стратегии взвешивания уровней значимости классификаторов [6], методы обучения правил комбинирования, учитывающие статистические особенности объединяемых классификаторов [7, 8] и методы, не привязанные к статистическим особенностям классификаторов, но использующие аппарат мультимножеств для построения модели групповой классификации объектов [9, 10].

Главным отличием видеопотока как цифрового образа распознаваемого объекта является тот факт, что для одного и того же объекта рассматривается последовательность наблюдений, которые отличаются между собой. Рассмотрим причины, по которым результат распознавания объекта может быть ошибочным, исходя из предположения, что система действует всегда детерминировано, т.е. в любой момент времени и при любых внешних условиях результаты распознавания одного и того же набора входных данных всегда совпадают. Таким образом любая ошибка является следствием неспособности системы различить объект того или иного класса. Ошибки распознавания можно условно разделить на три группы:

1. Ошибки, обусловленные несовершенством алгоритма распознавания, т.е. ошибки, являющиеся «внутренними» с точки зрения системы распознавания объектов и которые могут проявляться даже при идеальном функционировании других подсистем. Данный класс ошибок является безусловным атрибутом любой системы распознавания, вне зависимости от модели входа.
2. Ошибки, обусловленные дефектами надсистемы. Система распознавания одиночного изображения, как правило, является одной из подсистем некоторого комплекса и изображения, подаваемые на вход системе распознавания формируются в результате действия других подсистем (см. рис. 2). Как следствие, могут возникнуть ошибки, связанные с несовершенством предшествующих подсистем. К примеру, пусть в результате разбиения изображения текстовой строки на изображения отдельных символов была допущена ошибка, в следствии которой положение правой границы изображения латинской буквы «P» было найдено некорректно, в результате чего на изображении буквы была утеряна перемычка между двумя горизонтальными штрихами. Изображение, полученное в результате, с точки зрения системы распознавания одиночного символа, может быть неотличимо от латинской буквы «F».
3. Ошибки, обусловленные шумом среды. Возникают такие ошибки в случае, если в условиях внешней среды, в которой находится распознаваемый объект, его изображение становится неотличимым от изображения объекта другого класса. К примеру, предположим, что производится съемка фотографии документа, удостоверяющего личность, содержащего поле «Имя» с истинным значением «HANNA». Данное поле начертано на белом фоне и документ покрыт защитной глянцевой поверхностью. В момент съемки на документе проявился блик от внешнего источника света, полностью закрывший букву «H» и оставивший изображения остальных букв неизменными. Таким образом, изображение данного поля будет неотличимо от изображение поля «ANNA» на аналогичном документе.

Рис. 2. Пример ошибочной сегментации текстовой строки на отдельные символы в условиях размытости изображения и дефектов, связанных с защитным голографическим слоем документа

По отношению к системе распознавания одиночного изображения ошибки, связанные с шумом среды либо с дефектами надсистемы, являются следствием искажения входного изображения. Обладая возможностью использовать несколько наблюдений объекта можно ожидать, что влияние шума среды и дефектов надсистемы на эти наблюдения будут различны. Однако даже при фиксировании системы

распознавания одиночного объекта, вне зависимости от дефектов надсистемы, остаются ошибки, обусловленные несовершенством модели классификации. Современные исследования показывают, что наиболее высокоэффективный метод распознавания изображений [11, 12], который в ряде отдельных задач показывает результаты, способные конкурировать с человеком [13], тем не менее может показывать неустойчивый результат при минимальных изменениях входного изображения [14, 15], даже если эти изменения касались всего лишь одного пикселя [16]. Так, даже используя наиболее точный метод распознавания, но обладая единственным входным изображением объекта, невозможно отделить полезный сигнал от шума, влияние которого может кардинальным образом поменять результат.

Таким образом, рассматривая в качестве цифрового образа объекта не одиночное изображение, а видеопоток, появляется возможность уменьшить влияние ошибок за счет вариативности шума применительно к отдельным кадрам видеопотока, которой не обладают классические системы распознавания объектов.

Одним из методов, позволяющих производить анализ множества изображений одной и той же сцены с целью уменьшить влияние шума оптической системы и дефектов, связанных с неконтролируемыми условиями съемками, является техника «супер-разрешения» – процесс получения изображения высокого разрешения из нескольких изображений того же объекта с более низким разрешением. Данной задаче уделялось большое внимание в литературе и предложено большое количество подходов, принимающих во внимание специфику финальной задачи обработки изображения и распознавания объекта или сцены [17, 18]. Однако как было отмечено ранее, дальнейшая обработка полученного единого изображения объекта остается подверженной ошибкам алгоритма распознавания, в частности, неустойчивости сверточных нейронных сетей.

Целью данной работы является построение динамической модели системы распознавания произвольного объекта в видеопотоке в более общем виде, постановка задачи интеграции результатов распознавания и задачи останова, возникающих в подобных системах и являющихся новыми в контексте систем распознавания.

1. Модель системы распознавания объектов в видеопотоке

Рассмотрим модель системы распознавания одиночного объекта x. Пусть задано множество, содержащее M классов $C = \{c_1, c_2, \ldots, c_M\}$. В случае распознавания символа множеством классов может выступать множество какой-либо фиксированный алфавит. Рассматривая задачу типизации страницы документа на изображении после локализации ее границ и проективного исправления, множеством классов может выступать коллекция типов страниц документов, доступных для дальнейшей обработки. Отдельно следует упомянуть, что иногда в задачах распознавания объектов и явлений допускается наличие «пустого класса», который должен быть ответом системы распознавания на входное изображение объекта, о котором системе не известно, либо на изображение, которое не содержит объекта.

Пусть задано изображение объекта $I(x)$ из некоторого множества всевозможных изображений $\mathbb{I}$ и в рамках модели взаимодействия системы распознавания с надсистемой или с пользователем/оператором существует класс $c^* \in C$, к которому принадлежит объект x. Задача распознавания изображения одиночного объекта состоит в определении этого класса. Результат работы системы распознавания в общем виде представим как всюду определенное отображение из множества классов C в множество оценок принадлежности: $\boldsymbol{r} : C \to \mathbb{R}$. Учитывая, что множество классов C содержит ровно M элементов:

$$\boldsymbol{r}\left(I(x)\right) = \left\{\left(c_1, q_1\right), \left(c_2, q_2\right), \ldots, \left(c_M, q_M\right)\right\}, \quad (1)$$

где $q_i \in \mathbb{R}$, $i \in \{1, \ldots, M\}$ – вещественные оценки принадлежности объекта x к классу $c_i \in C$ при условии, что наблюдается изображение объекта $I(x)$. В качестве окончательного решения классификации принимается класс $c^*\left(I(x)\right) = \operatorname{argmax} \boldsymbol{r}\left(I(x)\right)$. Тривиальная схема системы распознавания объекта в рамках описанной модели представлена на рис. 3.

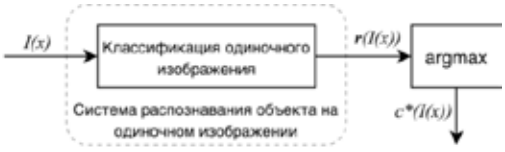

Рис. 3. Тривиальная схема системы распознавания одиночного объекта

Если исключить из рассмотрения процесс валидации результатов распознавания и процесс обучения параметров системы распознавания (в случае, если для решения задачи классификации используются методы машинного обучения, к примеру, искусственные нейронные сети), и рассматривать непосредственно процесс распознавания, то такая система распознавания является статической и не предполагает обратных связей.

Рассмотрим теперь задачу распознавания объекта x в видеопотоке. Источником видеопотока является некоторое захватывающее устройство, предоставляющее последовательность различных кадров, каждый из которых является независимым изображением объекта x. В условиях фиксированного количества кадров можно рассматривать задачу распознавания объекта в видеопотоке как статическую систему, аналогичную представленной на рис. 3, но с более сложной моделью входа. Тогда последовательность из N кадров можно рассматривать как множество изображений объекта x: $\mathbf{I}(x) = \{I_1(x), I_2(x), \ldots, I_N(x)\} \subset \mathbb{I}$. При этом модель выхода системы остается неизменной.

В зависимости от подхода к интеграции данных такая системы может быть выражена по-разному. Возможно тривиальное рассмотрение процесса классификации как «черного ящика», обрабатывающего сразу множество изображений (рис. 4а). Другие варианты частично или полностью используют методы распознавания одиночных изображений объекта и осуществляют интеграцию либо на уровне входных изображений (рис. 4б), либо на уровне результатов распознавания каждого отдельного изображения (рис. 4в).

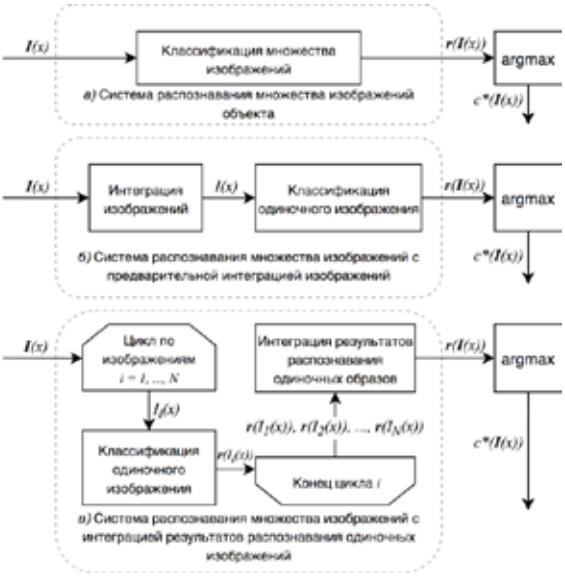

Рис. 4. Варианты статических систем распознавания множества изображений объекта

Однако представленные статические модели системы распознавания объекта в видеопотоке не в полной мере отражают сценарий распознавания при помощи мобильного устройства – поскольку данные модели предполагают в качестве входа лишь множество кадров и не предполагают изменения состояния системы в процессе съемки. Также в условиях аппаратных ограничений мобильных устройств хранение и обработка множества изображений может быть нецелесообразна или невозможна. Для того, чтобы более точно соответствовать процессу распознавания объекта в видеопотоке мобильного устройства предлагается рассмотреть динамическую модель с дискретным временем.

Для целей формализации представим видеопоток как генерирующаяся во времени последовательность изображений объекта. Таким образом, задано дискретное время $t = 0, 1, 2, \ldots$ и видеопоток, содержащий изображения наблюдаемого объекта $I_t(x) \in \mathbb{I}$. Подобная дискретная модель видеопотока соответствует принципам представления кодированного видеопотока в программных системах [19].

Для определения системы распознавания объекта в видеопотоке, который генерируется независимо, необходимо определить модель обслуживания, которая бы являлась промежуточным слоем между видеопотоком и непосредственным потоком обрабатываемых системой распознавания изображений. Наиболее тривиальной является схема обслуживания, при которой изображения, генерируемые во время обработки системой распознавания предыдущих изображения, сбрасываются. В случае, если возможно хранение коллекции изображений альтернативной моделью является схема обслуживания с буфером, позволяющим накапливать входящие изображения и выдавать их по запросу системы в произвольный момент времени, без ограничений, связанных с дискретизацией генерации изображений источником. С точки зрения непосредственно системы распознавания последовательности изображений набор методов и алгоритмов распознавания и интеграции результатов не зависят от схемы обслуживания, поэтому в рамках данной работы в дальнейшем будет предполагаться тривиальная схема со сбрасыванием изображений в периоды загрузки системы.

Система распознавания поддерживает некоторое внутреннее состояние $s_t \in S$, изменяющееся во времени. Время Δ_t, необходимое для получения обновленного результата после ввода очередного образа $I_t(x)$, в общем случае является функцией от изображения и внутреннего состояния системы: $\Delta_t = \Delta(I_t(x), s_t)$, которая может быть невычислима в момент времени t. Результат распознавания, учитывающий информацию, содержащуюся в изображении, которое было захвачено в момент времени t, может быть доступен только в момент времени $T(t) = t + \Delta_t$.

В начальный момент времени $t = 0$ инициализировано внутреннее состояние системы s_0. В каждый момент времени t происходит захват

изображения I_t(x) с камеры устройства и происходит проверка, находится ли в данный момент времени какое-либо изображение в обработке. Аналитически данное условие можно записать как $t < T\left(t_{prev}\right)$, где t_{prev} – индекс последнего изображения, поступившего в обработку. Однако поскольку это условие может быть невычислимо в момент времени t, для целей описания модели можно считать, что внутренне состояние системы s_t хранит информацию о том, находится ли какое-либо изображение в обработке в момент времени t. Если в момент t система уже обрабатывает какое-либо изображение, то вновь полученное изображение $I_t(x)$ сбрасывается (в рамках тривиальной схемы обслуживания). В противном случае изображение $I_t(x)$ поступает на классификацию. Результаты классификации интегрируются с накопленными к текущему моменту результатами (которые хранятся как часть текущей системы в s_t) и становятся доступны для вывода в момент времени $T(t)$. В моменты времени $t \in \{0,\ldots,T(0)-1\}$ результат распознавания объекта не определен. Результат распознавания, учитывающий информацию, которая содержится в N различных (последовательно захваченных) изображениях, может быть получен в момент времени $T^N(0)$. При этом индексы изображений, поступающих в обработку, равны, соответственно, $0, T^1(0), T^2(0), \ldots T^{N-1}(0)$ (под надстрочным знаком функции $T(t)$ подразумевается не возведение в степень, а множественная композиция функции). Схема описанной системы представлена на рис. 5.

Рис. 5. Схема системы распознавания объекта в видеопотоке с тривиальной моделью обслуживания, преобразующей видеопоток в последовательность обрабатываемых изображений

Методы выделения признаков и классификации объектов, применимые в статических системах (см. рис. 4) также применимы и в динамической модели, однако динамическая модель системы распознавания объекта в видеопотоке обладает рядом специфических свойств. В первую очередь необходимо отметить усиленное влияние производительности алгоритмов распознавания одиночного изображения на выход системы. Действительно, уменьшение времени Δ_t, необходимого для распознавания одного изображения $I_t(x)$, позволяет обработать большее 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, применительно к динамической системе распознавания объекта в видеопотоке, вне зависимости от схемы обслуживания, преобразующей входной видеопоток в поток обрабатываемых изображений объекта, возникают задачи, нетипичные для традиционных систем распознавания объектов на изображениях:

1. Задача интеграции результатов распознавания одиночных объектов.
2. Задача останова.

Блок-схема работы системы распознавания объекта в видеопотоке мобильного устройства в рамках описанной модели с интеграцией результатов распознавания одиночных изображений и с условием останова, представлена на рис. 6.

2. Задача интеграции результатов распознавания одиночных объектов

Основной задачей традиционных систем распознавания объектов является максимизация точности распознавания (т.е. максимизация доли «правильных» классификаций объектов). Задача интеграции результатов распознавания одиночных объектов состоит в максимизации точности результата распознавания множества различных изображений одного и того же объекта при заданных результатах распознавания одиночных изображений.

Для формализации постановки задачи интеграции положим, что задан набор объектов $X = \{x_1, x_2, \ldots x_K\}$ мощности K и набор видеопоследовательностей $B = \left\{\mathbf{I}_1(x_{b_1}), \mathbf{I}_2(x_{b_2}), \ldots, \mathbf{I}_H(x_{b_H})\right\}$ мощности H, где b_h – индекс объекта из множества X для каждого $h \in \{1,2,\ldots H\}$, и каждая видеопоследовательность $\mathbf{I}_h(x_{b_h}) = \left\{I_{h1}(x_{b_h}), I_{h2}(x_{b_h}), \ldots I_{hN_h}(x_{b_h})\right\}$ – последовательность изображений объекта $x_{b_h} \in X$, которые могут быть подвержены шумам среды и надсистемы. Задано множество классов $C = \{c_1, c_2, \ldots, c_M\}$ и информация об идеальной принадлежности каждого объекта к соответствующему классу $\nu : X \to C$. В общем случае задачу распознавания объекта в видеопотоке можно

Рис. 6. Блок-схема работы системы распознавания объекта в видеопотоке с интеграцией результатов распознавания одиночных изображений и с условием останова

сформулировать как поиск классифицирующей функции $F : \mathbb{I}^* \to C$, максимизирующей точность распознавания [20]:

$$V_F(\mathrm{B}) = \frac{1}{H}\left(\sum_{h=1}^{H}\left[F\left(\mathbf{I}_h(x_{b_h})\right) = \nu(x_{b_h})\right]\right) \to \max_F \quad (2)$$

Более частная задача интеграции результатов распознавания одиночных объектов предполагает функцию интегрирования результатов распознавания $\tilde{F} : \left(\mathbb{R}^C\right) \to \mathbb{R}^C$, преобразующую последовательность результатов распознавания одиночных изображений в единый результат распознавания видеопоследовательностей. Поскольку финальным ответом распознавания видеопоследовательности является класс, соответствующий максимальной оценке в результате распознавания $F(\mathbf{I}) = \operatorname{argmax} \tilde{F}\left(\boldsymbol{r}(\mathbf{I})\right)$, постановка задачи интеграции строится на основе (2) и приобретает вид:

$$V_{\tilde{F}}(\mathrm{B}) = \frac{1}{H}\left(\sum_{h=1}^{H}\left[\operatorname{argmax} \tilde{F}\left(\boldsymbol{r}\left(\mathbf{I}_h(x_{b_h})\right)\right) = \nu(x_{b_h})\right]\right) \to \max_{\tilde{F}} \quad (3)$$

В идеальном случае классифицирующая функция F или функция интегрирования результатов $\tilde{F}$ должна обладать возможностью фильтровать выбросы, появляющиеся во входном потоке данных из-за шума среды или дефектов надсистемы, и обладать возможностью проводить фильтрацию шума классификатора, нивелируя случайные внутренние ошибки.

3. Задача останова

Модель системы распознавания объекта в видеопотоке (см. рис. 5) не предполагает ограничения на количество входных изображений, а поскольку основной целью системы распознавания объектов является автоматизация ввода, важным параметром является абсолютное время (т.е. время с точки зрения оператора), необходимое для получения окончательного результата распознавания. В отличие от процесса съемки фотографии, видеопоток естественным образом не ограничен во времени. Отсюда следует задача останова, которая заключается в принятии решения о том, что вновь полученный результат $\boldsymbol{r}\left(\left\{I_0(x), I_{T^1(0)}(x), I_{T^2(0)}(x), \ldots, I_t(x)\right\}\right)$ в момент времени $T(t)$ можно считать окончательным и цикл захвата изображений можно прекратить. При распознавании сложных объектов, которые состоят из множества независимо распознаваемых объектов, решение об останове распознавания отдельных объектов влияет на время Δ_t, необходимое для распознавания составного объекта, а значит и на количество информации, обрабатываемой в рамках общей системы. Таким образом, задача останова (тесно связанная с задачей интеграции) является важным аспектом системы распознавания в видеопотоке, в особенности в рамках надсистемы, объектом распознавания которой является составной объект, такой как текстовое поле или документ в целом. Правило останова в общем виде формально можно представить в виде предиката, действующего на видеопоследовательности $P : \mathbb{I}^* \to \{0,1\}$. Истинность предиката влечет остановку процесса захвата и распознавания изображений:

$$P\left(\left\{I_1(x), I_2(x), \ldots, I_n(x)\right\}\right) = \begin{cases} 1: \text{решение по обстановке} \\ 0: \text{продолжение работы} \end{cases} \quad (4)$$

Пусть $\mathbf{I}(\mathrm{x}) = \left\{I_1(x), I_2(x), \ldots, I_N(x)\right\}$ – последовательность изображений объекта $x \in X$, а $\mathbf{I}^{(n)}(x) = \left\{I_1(x), I_2(x), \ldots, I_n(x)\right\} \subseteq \mathbf{I}(x)$ – префикс этой последовательности, имеющий длину $n \le N$. Обозначим через $D_P\left(\mathbf{I}(x)\right)$ количество изображений, которые будут обработаны системой распознавания до срабатывания правила останова (4):

$$D_P\left(\mathbf{I}(x)\right) = \min\left[N, \min\left\{\left|\mathbf{I}^{(n)}(x)\right| n \in \{1,2,\ldots,N\} \wedge P\left(\mathbf{I}^{(n)}(x)\right)\right\}\right] \quad (5)$$

С учетом правила останова при обработке

видеопоследовательности $\mathbf{I}(\mathrm{x})$ на распознавание подаются только изображения из подпоследовательности $\mathbf{I}^{(P)}(x)=\mathbf{I}^{(D_P(\mathbf{I}(x)))}(x)$, и исходный набор видеопоследовательностей принимает вид $B^{(P)}=\left\{\mathbf{I}_1^{(P)}(x_{b_1}),\mathbf{I}_2^{(P)}(x_{b_2}),\ldots,\mathbf{I}_H^{(P)}(x_{b_H})\right\}$.

Для формализации задачи останова воспользуемся общей моделью взаимодействия системы распознавания с пользователем, которая используется в задачах определения достоверности результата распознавания объекта [21, 22] и для оценки эффективности работы системы использует функционал, описанный в экономических терминах. Пусть W_c – стоимость ввода корректного результата распознавания объекта, W_e – стоимость ввода ошибочного результата, W_f – стоимость распознавания одного изображения объекта. Тогда функция эффективности правила останова может быть записана в виде средней стоимости работы системы:

$$W_{F,P}(B)=W_c\cdot V_F\left(B^{(P)}\right)+W_e\cdot\left(1-V_F\left(B^{(P)}\right)\right)+ \\ +W_f\cdot\frac{1}{H}\left(\sum_{h=1}^{H}D_P\left(\mathbf{I}(x)\right)\right), \quad (6)$$

где $V_F\left(B^{(P)}\right)$ – точность распознавания видеопоследовательностей с учетом останова по правилу P (4), вычисляемая согласно (2) (аналогично в случае интеграции результатов распознавания одиночных объектов точность вычисляется согласно (3)).

Упрощая выражение (6) и принимая во внимание константность W_e приходим к общей постановке задачи останова как к задаче поиска правила останова, оптимизирующего функционал эффективности:

$$W_{F,P}(B)=V_F\left(B^{(P)}\right)\cdot\left(W_c-W_e\right)+W_f\cdot\frac{1}{H}\left(\sum_{h=1}^{H}D_P\left(\mathbf{I}(x)\right)\right)\to\min_P \quad (7)$$

Аналогичный функционал эффективности строится с учетом функционала точности (3) в рамках задачи интеграции результатов распознавания одиночных объектов. Как видно из постановок задач (2), (3) и (7), а также из того, что стоимость ввода ошибочного результата W_e всегда превышает стоимость ввода корректного результата W_c, задача останова и задача максимизации точности распознавания объекта в видеопотоке не конфликтуют между собой и имеет смысл их рассматривать в совокупности.

Заключение

В работе были показаны свойства задачи распознавания объекта в видеопотоке. Представлены различные способы формализации системы распознавания в видеопотоке и описана модель динамической системы как наиболее полно отражающей процесс видеосъемки объекта. Были показаны свойства динамической системы распознавания объектов в видеопотоке и предложены формальные постановки задач интеграции результатов распознавания и останова. Постановки могут быть использованы в дальнейшем для разработки и исследования методов повышения точности и производительности систем распознавания на мобильных устройствах.

В рамках дальнейшей работы планируется построение модели интегратора результатов распознавания объектов в видеопотоке в рамках модели динамической системы распознавания и провести исследование методов построения оптимального предиката останова.

Литература

1. *Bulatov K., Arlazarov V.V., Chernov T., Slavin O., Nikolaev D.* “Smart IDReader: Document Recognition in Video Stream” // 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). – 2017. –V. 6, – P. 39-44.
2. *Арлазаров В.В., Жуковский А., Кривцов В., Николаев Д., Полевой Д.* Анализ особенностей использования стационарных и мобильных малоразмерных цифровых видео камер для распознавания документов // Информационные технологии и вычислительные системы. – 2014. – № 3. – С. 71-78.
3. *Wemhoener D., Yalniz I.Z., Manmatha R.* “Creating an Improved Version Using Noisy OCR from Multiple Editions” // 12th IAPR International Conference on Document Analysis and Recognition (ICDAR). – 2013. – P. 160-164.
4. *Rokach L.* “Ensemble-based classifiers” // Artificial Intelligence Review. – 2010. – Vol. 33, No. 1. – P. 1-39.
5. *Kittler et al.* “On Combining Classifiers” // IEEE Trans. Pattern Analysis and Machine Intelligence. – 1998. – Vol. 20, No. 3. – P. 226-239.
6. *Ting K.M., Witten I.H.* “Issues in Stacked Generalization” // Journal of Artificial Intelligence Research. – 1999. – Vol. 10, No. 1. – P. 271-289.
7. *Kuncheva L.I., Bezdek J.C., Duin R.P.* “Decision templates for multiple classifier fusion: an experimental comparison” // Pattern Recognition. – 2001. – Vol. 34, No. 2. – P. 299-314.
8. *Nguyen T.T. et al.* “A Novel Combining Classifier Method Based on Variational Inference” // Pattern Recognition. – 2016. – Vol. 49, No. C. – P. 198-212.

9. *Петровский А.Б.* Методы групповой классификации многопризнаковых объектов (часть 1) // Искусственный интеллект и принятие решений. – 2009. – № 3. – С. 3-14.
10. *Петровский А.Б.* Методы групповой классификации многопризнаковых объектов (часть 2) // Искусственный интеллект и принятие решений. – 2009. – № 4. – С. 3-14.
11. *LeCun Y. et al.* "Gradient-Based Learning Applied to Document Recognition" // Proceedings of the IEEE. – 1998.
12. *Krizhevsky A., Sutskever I., Hinton G.E.* "ImageNet Classification with Deep Convolutional Neural Networks" // Advances in Neural Information Processing Systems 25 / ed. by F. Pereira [et al.]. – Curran Associates, Inc., 2012. – P. 1097-1105.
13. *Taigman Y. et al.* "DeepFace: Closing the Gap to Human-Level Performance in Face Verification" // IEEE Conference on Computer Vision and Pattern Recognition. – 2014. – P. 1701-1708.
14. *Moosavi-Dezfooli S., Fawzi A., Frossard P.* "DeepFool: a simple and accurate method to fool deep neural networks" // CoRR. – 2015. – Vol abs/1511.04599.
15. *Papernot N. et al.* "The Limitations of Deep Learning in Adversarial Settings" // CoRR. – 2015. – Vol. abs/1511.07528.
16. *Su J., Vargas D.V., Sakurai K.* "One pixel attack for fooling deep neural networks" // CoRR. – 2017. – Vol. abs/1710.08864.
17. *Sung Cheol Park, Min Kyu Park, Moon Gi Kang.* "Super-resolution image reconstruction: a technical overview" // IEEE Signal Processing Magazine. – 2003. – V.20. – N. 3. – P. 21-36.
18. *Semwal A., Chamoli A., Mukesh C.A., Salman A.* "A Survey: The Methods & Techniques of Super-Resolution Image Reconstruction" // International Journal for Scientific Research & Development. – 2017. – V. 4. – I. 12. – P. 243-249.
19. *International standard ISO/IEC* 14496-12 "Information technology – Coding of audio-visual objects – Part 12: ISO base media file format". ISO/IEC. – 2005. – 94 p.
20. *Arlazarov V.L., Loginov A.S., Slavin O.A.* "Characteristics of Optical Text Recognition Programs" // Programming and Computer Software. – 2002. – Vol. 28, No. 3. – P. 148-161.
21. *Арлазаров В.В., Кляцкин В.М.* Решение задачи определения достоверности результатов распознавания символа в системе Cognitive Forms // Документооборот. Концепции и инструментарий. Сборник трудов Института системного анализа РАН. – 2004. – 208 с.
22. *Kimura S. et al.* "A Man-Machine Cooperating System Based on the Generalized Reject Model" // 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). – 2017. – V. 1. – P. 1324-1329.

Арлазаров Владимир Викторович. Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, г. Москва, Россия. Ведущий научный сотрудник. Кандидат технических наук. Количество печатных работ: 30. Область научных интересов: искусственный интеллект, машинное обучение, системы распознавания, информационные технологии. E-mail: vva777@gmail.com

Булатов Константин Булатович. Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук, г. Москва, Россия. Программист 1-ой категории. Количество печатных работ: 14. Область научных интересов: машинное обучение, компьютерное зрение, системы распознавания, информационные технологии. E-mail: hpbuko@gmail.com

Усков Анатолий Васильевич. Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук, г. Москва, Россия.. Зав. лабораторией. Кандидат физико-математических наук. Количество печатных работ: более 50 (в том числе 3 монографии). Область научных интересов: искусственный интеллект и системное программирование.
E-mail: uskov@isa.ru

A model of object recognition system in video stream of a mobile device

V.V. Arlazarov[I,III], K.B. Bulatov[II,III], A.V. Uskov[II]

[I] Institute for information transmission problems (Kharkevich Institute) RAS, Moscow, Russia

[II] Institute for Systems Analysis, Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

[III] LLC "Smart Engines Service", Moscow, Russia

Abstract. This paper describes a problem of automatic objects recognition using video stream as digital object representation. Several variants of video stream system formulation are described, properties of dynamic recognition system model are discussed. Recognition results integration problem and stopping problem are described, which occur in recognition system with time parameters and without natural restriction on the number of input frames. Formal statements of both problems are presented in scope of a general integration model of the recognition system and its user.

Keywords: *pattern recognition, video stream, mobile devices, recognition systems, OCR.*
DOI: 10.14357/20790279180508

References

1. *Bulatov K., Arlazarov V.V., Chernov T., Slavin O., Nikolaev D.* "Smart IDReader: Document Recognition in Video Stream" // 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). – 2017. –V. 6, – P. 39-44.
2. *Arlazarov V.V., Zhukovsky A., Krivtsov V., Nikolaev D., Polevoy D.* "Analysis of using stationary and mobile small-scale digital cameras for document recognition " // Information technologies and computation systems. – 2014. – № 3. – P. 71-78.
3. *Wemhoener D., Yalniz I.Z., Manmatha R.* "Creating an Improved Version Using Noisy OCR from Multiple Editions" // 12th IAPR International Conference on Document Analysis and Recognition (ICDAR). – 2013. – P. 160-164.
4. *Rokach L.* "Ensemble-based classifiers" // Artificial Intelligence Review. – 2010. – Vol. 33, No. 1. – P. 1-39.
5. *Kittler et al.* "On Combining Classifiers" // IEEE Trans. Pattern Analysis and Machine Intelligence. – 1998. – Vol. 20, No. 3. – P. 226-239.
6. *Ting K.M., Witten I.H.* "Issues in Stacked Generalization" // Journal of Artificial Intelligence Research. – 1999. – Vol. 10, No. 1. – P. 271-289.
7. *Kuncheva L.I., Bezdek J.C., Duin R.P.* "Decision templates for multiple classifier fusion: an experimental comparison" // Pattern Recognition. – 2001. – Vol. 34, No. 2. – P. 299-314.
8. *Nguyen T.T. et al.* "A Novel Combining Classifier Method Based on Variational Inference" // Pattern Recognition. – 2016. – Vol. 49, No. C. – P. 198-212.
9. *Petrovsky A.B.* "Methods of group classification of multi-feature objects (part 1)" // Artificial intelligence and decision theory. – 2009. – № 3. – P. 3-14.
10. *Petrovsky A.B.* "Methods of group classification of multi-feature objects (part 2)" // Artificial intelligence and decision theory. – 2009. – № 4. – P. 3-14.
11. *LeCun Y. et al.* "Gradient-Based Learning Applied to Document Recognition" // Proceedings of the IEEE. – 1998.
12. *Krizhevsky A., Sutskever I., Hinton G.E.* "ImageNet Classification with Deep Convolutional Neural Networks" // Advances in Neural Information Processing Systems 25 / ed. by F. Pereira [et al.]. – Curran Associates, Inc., 2012. – P. 1097-1105.
13. *Taigman Y. et al.* "DeepFace: Closing the Gap to Human-Level Performance in Face Verification" // IEEE Conference on Computer Vision and Pattern Recognition. – 2014. – P. 1701-1708.
14. *Moosavi-Dezfooli S., Fawzi A., Frossard P.* "DeepFool: a simple and accurate method to fool deep neural networks" // CoRR. – 2015. – Vol abs/1511.04599.
15. *Papernot N. et al.* "The Limitations of Deep Learning in Adversarial Settings" // CoRR. – 2015. – Vol. abs/1511.07528.
16. *Su J., Vargas D.V., Sakurai K.* "One pixel attack for fooling deep neural networks" // CoRR. – 2017. – Vol. abs/1710.08864.
17. *Sung Cheol Park, Min Kyu Park, Moon Gi Kang.* "Super-resolution image reconstruction: a technical overview" // IEEE Signal Processing Magazine. – 2003. – V.20. – N. 3. – P. 21-36.
18. *Semwal A., Chamoli A., Mukesh C.A., Salman A.* "A Survey: The Methods & Techniques of Super-Resolution Image Reconstruction" // International Journal for Scientific Research & Development. – 2017. – V. 4. – I. 12. – P. 243-249.

19. *International standard ISO/IEC* 14496-12 “Information technology – Coding of audio-visual objects – Part 12: ISO base media file format”. ISO/IEC. – 2005. – 94 p.
20. *Arlazarov V.L., Loginov A.S., Slavin O.A.* “Characteristics of Optical Text Recognition Programs” // Programming and Computer Software. – 2002. – Vol. 28, No. 3. – P. 148-161.
21. *Arlazarov V.V., Kliatsine V.M.* “Solving the problem of confidence determination for symbol recognition result in Cognitive Forms system “ // Document processing. Concepts and instruments. Proceedings of ISA RAS. – 2004. – 208 p.
22. *Kimura S. et al.* “A Man-Machine Cooperating System Based on the Generalized Reject Model” // 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). – 2017. – V. 1. – P. 1324-1329.

V. V. Arlazarov. Institute of information transmission problems (Kharkevich institute) of Russian academy of sciences, Moscow, Russia. Lead researcher, PhD. Number of publications: 30. Scientific interests: artificial intelligence, machine learning, recognition systems, information technology.

K. B. Bulatov. Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia. I-st category programmer. Number of publications: 14. Scientific interests: machine learning, computer vision, recognition systems, information technologies. E-mail: hpbuko@gmail.com (Corresponding author).

A. V. Uskov. Institute for Systems Analysis, Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia, head of laboratory. Ph.D. in Physics and Mathematics. Number of publications: more than 50 (including 3 monographs). Research interests: artificial intelligence and system programming. E-mail: uskov@isa.ru