

# Выбор оптимальной стратегии комбинирования покадровых результатов распознавания символа в видеопотоке<sup>1</sup>

К.Б. Булатов

**Аннотация** Работа посвящена задаче комбинирования результатов классификации нескольких наблюдений одного и того же объекта. Задача рассматривается как частный случай задачи коллективного принятия решений группой экспертов с вычисляемыми оценками уровней их компетентности. Произведено исследование точности различных методов комбинирования результатов классификации в зависимости от моделей входных данных на примере задачи объединения результатов покадрового распознавания символов в видеопотоке. Экспериментально показано преимущество стратегии выбора единственного наиболее компетентного эксперта в случае модели входных данных, предполагающей отсутствие нерелевантных наблюдений (в контексте распознавания символов в видеопотоке — отсутствие ошибок локализации и сегментации символов). Одновременно с этим показано преимущество стратегии объединения результатов классификации нескольких наиболее компетентных экспертов, согласно правилу произведения оценок или голосования, в случае наличия нерелевантных наблюдений.

**Ключевые слова:** принятие решений, распознавание образов, распознавание в видеопотоке, комбинирование классификаторов.

## Введение

Системы, предполагающие использование мобильных устройств в задачах автоматического ввода документов, располагают рядом преимуществ как с точки зрения пользователя, так и с точки зрения постановки задачи распознавания образов. В системе автоматического ввода документов, которая использует цифровые камеры мобильных устройств в качестве сканирующего устройства, оцифровка документа производится путем видео- или фотосъемки оригинала. В случае видеосъемки документа цифровым представлением документа является не единственное изображение, а последовательность видеок кадров, содержащих образ од-

ного и того же документа (или его фрагмента), что обеспечивает возможность производить многократное распознавание одного и того же объекта в видеопотоке в реальном времени, увеличивая тем самым точность и надежность распознавания.

При этом использование мобильных устройств в задачах распознавания и автоматического ввода документов сопряжено с рядом проблем: помимо низких вычислительных мощностей мобильных устройств, к этим проблемам следует отнести широкий спектр искажений, возникающих вследствие особенностей оптической схемы малоформатных камер мобильных устройств [1], а также вследствие особенностей процесса съемки документа при

<sup>1</sup> Работа выполнена при частичной финансовой поддержке РФФИ в рамках научных проектов №№ 17-29-03170, 15-29-06083, 15-07-06520.

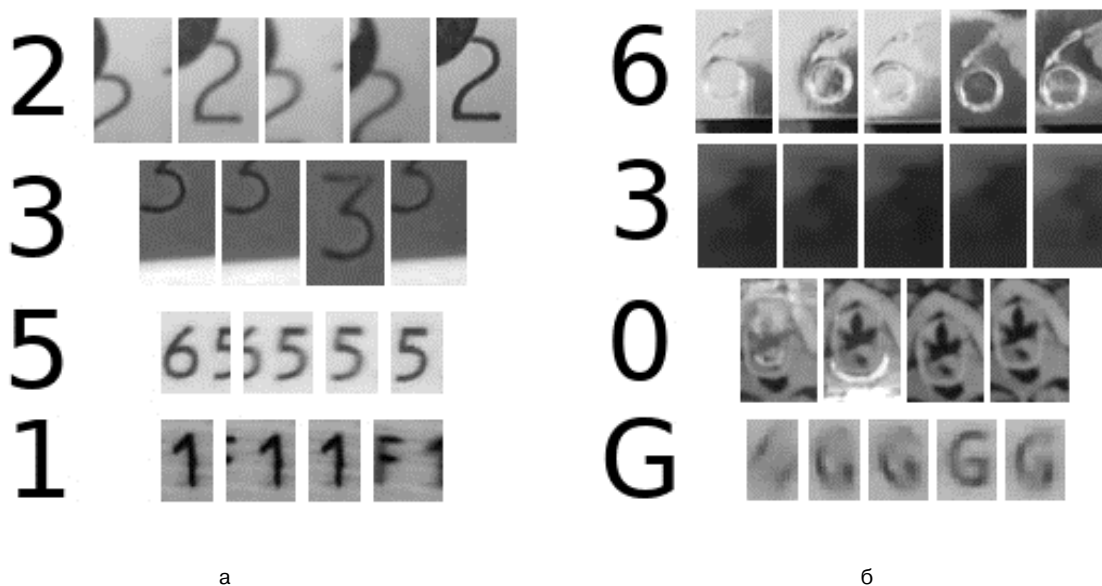


Рис. 1. Примеры видеопоследовательностей символов

(а) с ошибками локализации и сегментации символов

(б) без ошибок локализации и сегментации символов, но с другими типами искажений входных изображений

помощи мобильного устройства [2, 3]. К характерным искажениям, связанным с оптической схемой малоформатных цифровых камер, относятся аберрации, блики и отражения внутри оптической системы, а также цифровой шум. К искажениям, связанным с особенностями процесса съемки документа при помощи мобильного устройства, относятся неравномерная или недостаточная освещенность сцены, расфокусировка изображения и смазанность из-за движения оптического сенсора относительно носителя, блики от внешнего источника освещения, геометрические искажения, такие как проективные искажения изображения документа или нелинейные искажения, вызванные изгибами бумажного носителя, помехи, создаваемые голографическим защитным слоем документа и др. (Рис. 1 б). С точки зрения подсистемы оптического распознавания одиночного символа к перечисленным проблемам следует также отнести ошибки подсистем анализа изображения документа и предварительной подготовки изображения к распознаванию: ошибки поиска и локализации документа на входном изображении, ошибки анализа структуры и локализации текстовых строк документа, ошибки сегментации текстовых строк на отдельные символы [4] (Рис. 1 а). Часть ошибок

распознавания, связанных с этими проблемами, можно решить при помощи многократного распознавания одного и того же объекта, при этом естественным образом возникает задача выбора оптимальной стратегии комбинирования результатов покадрового распознавания. Данное исследование будет посвящено методам решения этой задачи.

### 1. Модель результата классификации одиночного объекта и базовые стратегии комбинирования

Рассмотрим задачу объединения результатов классификации множества образов одного и того же объекта, частным случаем которой является задача распознавания символа в видеопотоке, с целью максимизировать точность классификации за счет использования большего количества информации об объекте. Пусть задана последовательность наблюдений  $X = \{x_1, x_2, \dots, x_N\}$ , алфавит  $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_K\}$  и классификатор, действующий на множестве образов объекта. Классификатор реализует функцию

$$C_{\Sigma}(x_i) = \{\langle \sigma_1, q_1 \rangle, \dots, \langle \sigma_K, q_K \rangle\}, \quad (1)$$

которая ставит образу  $x_i$  в соответствие отображение  $\Sigma \rightarrow [0,1]$  оценок принадлежности образа к каждому из классов,  $\sum_{k=1}^K q_k = 1$ . Класс  $\sigma_C^*(x_i) = \operatorname{argmax}_{\sigma \in \Sigma} C_{\Sigma}(x_i)(\sigma)$ , соответствующий максимальной оценке принадлежности, является "выбором" классификатора при входе  $x_i$ .

В литературе задача объединения результатов классификации одиночных объектов обычно рассматривается в контексте методов получения более точной классификации путем объединения результатов нескольких разных классификаторов [5-7]. По природе комбинируемых классификаторов такие методы подразделяются на два класса: гомогенные и гетерогенные [8]. В гомогенных методах базовые классификаторы конструируются при помощи одного и того же алгоритма, но с использованием различных подмножеств обучающей выборки. В гетерогенных методах базовые классификаторы конструируются либо различными алгоритмами, либо с использованием различной признаковой базы.

В зависимости от используемой модели результата классификации объекта и от интерпретации оценок классификатора рассматриваются различные методы комбинирования. Для модели результата классификации в виде пары  $\langle \sigma, q \rangle$ , где  $\sigma$  — метка класса, а  $q$  — показатель уверенности классификатора (без альтернатив), используется схема голосования с выбором лучшего ответа с максимизацией некоторого количественного критерия. Обобщенный критерий приведен в работе [9]:

$$\operatorname{Score}(\langle \sigma, q \rangle) = \alpha \cdot \operatorname{Freq}(\sigma) + (1 - \alpha) \cdot q, \quad (2)$$

где  $\operatorname{Freq}(\sigma)$  — частота класса  $\sigma$  среди ответов комбинируемых классификаторов,  $\alpha$  — обучаемый параметр алгоритма. Распространенным методом, основывающимся на голосовании классификаторов и применяющийся в системах распознавания речи при помощи комбинирования нескольких алгоритмов, а также для комбинирования результатов распознавания текстовых строк [10, 11], является метод

ROVER (Recognizer Output Voting Error Reduction) [12].

В случае более общей модели (1) результата классификации, с интерпретацией оценок принадлежности  $q_k$  в рамках Байесовской модели (когда оценка принадлежности  $q_k$  есть апостериорная оценка условной вероятности принадлежности образа  $x_i$  классу  $\sigma_k$ ), описываются различные правила комбинации [5, 13]. Базовые правила описаны в работе [13], которая признается фундаментальным трудом, связанным с задачей объединения результатов различных классификаторов в рамках Байесовской модели. Опишем эти правила, но в применении к задаче объединения результатов классификации нескольких образов одним классификатором:

1) *Правило произведения*

$$\operatorname{Prod}(X)(\sigma) = P(\sigma|X) = \frac{1}{P(\sigma)^{N-1}} \prod_{i=1}^N C_{\Sigma}(x_i)(\sigma) \quad (3)$$

2) *Правило суммы:*

$$\operatorname{Sum}(X)(\sigma) = \frac{1}{N} \sum_{i=1}^N C_{\Sigma}(x_i)(\sigma) \quad (4)$$

3) *Правило минимума:*

$$\operatorname{Min}(X)(\sigma) = \frac{\min_{i=1}^N C_{\Sigma}(x_i)(\sigma)}{\sum_{k=1}^K \min_{i=1}^N C_{\Sigma}(x_i)(\sigma_k)} \quad (5)$$

4) *Правило максимума:*

$$\operatorname{Max}(X)(\sigma) = \frac{\max_{i=1}^N C_{\Sigma}(x_i)(\sigma)}{\sum_{k=1}^K \max_{i=1}^N C_{\Sigma}(x_i)(\sigma_k)} \quad (6)$$

5) *Правило медианы:*

$$\operatorname{Med}(X)(\sigma) = \frac{\operatorname{med}_{i=1}^N C_{\Sigma}(x_i)(\sigma)}{\sum_{k=1}^K \operatorname{med}_{i=1}^N C_{\Sigma}(x_i)(\sigma_k)} \quad (7)$$

В случае с интерпретацией оценок  $q_k$  как нечетких свидетельств принадлежности к классам либо абстрактных показателей уверенности, используются методы комбинирования основанные на теории Демпстера-Шафера [14, 15]. Также в работах, затрагивающих гетерогенные методы объединения результатов

классификаторов, рассматриваются стратегии взвешивания уровней значимости классификаторов [16] и методы обучения правил комбинирования, учитывающие статистические особенности объединяемых классификаторов [17-19].

## 2. Комбинирование результатов классификации как задача коллективного принятия решения

Введем понятие априорного предиктора достоверности результата классификатора как вещественнозначную функцию  $f_C(x_i)$  на множестве образов, отражающую степень уверенности в том, что результат классификации образа  $x$  будет верным. Пусть область значений предиктора ограничена сверху величиной  $f_{\text{sup}}$ . Тогда устойчивым предиктором достоверности будем считать такой предиктор, что функция

$$P_{f_C}(t) = P(\sigma_C^*(x_i) - \text{верный} | f_C(x_i) \geq t) \quad (8)$$

монотонно возрастает в окрестности  $f_{\text{sup}}$ . Функция  $P_{f_C}(t)$  является оценкой вероятности правильного распознавания образа  $x_i$  для всех образов со значением априорного предиктора достоверности классификатора не менее  $t$ . В качестве априорных предикторов имеет смысл использовать вычислимые характеристики изображений, заведомо влияющие на точность классификации [20], такие как оценка смазывания и уровня фокусировки [21], оценка уровня шума, артефактов оцифровки [22] и пр. Аналогичным образом введем понятие устойчивого апостериорного предиктора достоверности, как функцию на множестве результатов классификации (отображений из алфавита в множество оценок). Апостериорные предикторы связаны с понятием надежности распознавания [23, 24].

При заданном устойчивом предикторе (или наборе предикторов) достоверности задача комбинирования результатов распознавания множества образов одного и того же объекта является задачей коллективного принятия решения с  $N$  экспертами, оценки уровней компетентности которых являются функциями от значений соответствующих предикторов достоверности. Стоит заметить, что уровни компе-

тентности экспертов в данной модели являются отражением входных данных, т.к. именно характеристики отдельных наблюдений (элементов последовательности  $X$ ) необходимы для оценки значимости экспертов.

Важным вопросом в рамках этой задачи является вопрос о целесообразности использования голосования нескольких экспертов вместо использования мнения самого компетентного эксперта [25, 26]. Переходя к частной задаче, этот вопрос формулируется следующим образом: при каких моделях входных данных в задаче комбинирования результатов распознавания следует выбирать ту или иную стратегию комбинирования?

## 3. Сравнение базовых стратегий комбинирования

Одной из целей настоящей работы является исследование влияния структуры множества образов одного и того же объекта на стратегию комбинирования результатов их классификации. Важным классом ошибок, компенсация которых позволяет значительно увеличить точность распознавания символов в видеопотоке, являются ошибки поиска и локализации текстовых строк на входном изображении, а также ошибки сегментации текстовых строк на отдельные символы. При наличии подобных ошибок, даже с использованием методов выравнивания текстовых строк [12], среди последовательности образов  $X$  могут встретиться как фрагменты текстовой строки, не целиком содержащие образ распознаваемого символа, так и образы других символов.

Для проведения численных экспериментов подготовлены четыре набора данных, характеристики которых приведены в Табл. 1. Наборы данных MRZ-MSEGM и MRZ-CLEAN содержат видеопоследовательности результатов распознавания символов машиночитаемой зоны международных документов [3]. Наборы данных ICN-MSEGM и ICN-CLEAN содержат видеопоследовательности результатов распознавания символов поля "Номер" платежных банковских карт, выполненного при помощи индент-печати. Изображениям символов в рассматриваемых тестовых наборах свойственен широкий спектр искажений: неравномерная

Табл. 1. Характеристики тестовых наборов данных

Характеристика тестового набора данных	MRZ-MSEGM	MRZ-CLEAN
Размер алфавита $\Sigma$	37	
Общее количество образов символов	637874	631530
Точность распознавания отдельных образов символов, %	96,7357	96,8994
Количество видеопоследовательностей	7581	7508
Минимальная длина видеопоследовательности	3	
Максимальная длина видеопоследовательности	223	
Средняя длина видеопоследовательности	21	
Характеристика тестового набора данных	ICN-MSEGM	ICN-CLEAN
Размер алфавита $\Sigma$	10	
Общее количество образов символов	31580	29166
Точность распознавания отдельных образов символов, %	90,9816	96,8936
Количество видеопоследовательностей	1898	1748
Минимальная длина видеопоследовательности	3	
Максимальная длина видеопоследовательности	25	
Средняя длина видеопоследовательности	12	

или недостаточная освещенность, цифровой шум, расфокусировка и "смазанность" ввиду движения оптического сенсора относительно носителя, блики от внешнего источника света и помехи, создаваемые голографическим защитным слоем документа и др. Результат распознавания каждого отдельного образа символа получен при помощи сверточных нейронных сетей, обученных отдельно для символов машиночитаемой зоны и для символов поля "Номер" платежных банковских карт, на отдельных обучающих наборах изображений с применением метода аугментации данных [27]. Наборы данных MRZ-MSEGM и ICN-MSEGM содержат ошибки, вызванные некорректной или недостаточно точной работой алгоритмов локализации документа и алгоритмов сегментации текстовых строк. Наборы MRZ-CLEAN и ICN-CLEAN являются подмножествами соответствующих наборов MRZ-MSEGM и ICN-MSEGM, не содержащими подобных ошибок. Таким образом, в наборах данных MRZ-CLEAN и ICN-CLEAN каждая видеопоследовательность содержит образы строго одного и того же символа, без каких-либо дефектов сегментации.

На представленных тестовых наборах данных проведено сравнение базовых стратегий комбинирования классификаторов (3)-(7). Точность распознавания видеопоследовательности символа есть относительная доля видеопоследовательностей, для которых идеальный ответ совпадает с классом, получившим максимальную оценку согласно тому или иному правилу комбинирования. Дополнительно проведено сравнение базовых правил комбинирования с методом голосования (2), обобщенным следующим образом:

$$\begin{aligned} \text{Vote}(\alpha)(X)(\sigma) &= \\ &= \alpha \cdot \frac{1}{N} \sum_{i=1}^N 1_{X_{\sigma}^C}(x_i) + (1 - \alpha) \cdot \max_{i=1}^N \left( 1_{X_{\sigma}^C}(x_i) \cdot f_C(x_i) \right), \end{aligned} \tag{9}$$

где  $X_{\sigma}^C = \{x \in X \mid \sigma_C^*(x) = \sigma\}$  — подмножество элементов видеопоследовательности, для которых выбором классификатора является класс  $\sigma$ ,  $1_{X_{\sigma}^C}(x_i)$  — индикаторная функция принадлежности образа  $x_i$  к подмножеству  $X_{\sigma}^C$ ,  $f_C(x_i)$  — предиктор достоверности. В каче-

стве предиктора достоверности использовался апостериорный предиктор "правило первой альтернативы" [24]:

$$f_C(x_i) = \max_{k=1}^K C_{\Sigma}(x_i)(\sigma_k). \quad (10)$$

На Рис. 2 представлены сравнительные значения точности распознавания видеопоследовательностей с использованием правил комбинирования (3)-(7) и (9) на тестовых наборах данных MRZ-MSEGM, MRZ-CLEAN, ICN-MSEGM и ICN-CLEAN. Горизонтальная ось графиков соответствует значениям параметра  $\alpha$  правила комбинирования (9). Точность распознавания с использованием правил (3)-(7) представлены горизонтальными линиями.

На Рис. 2 продемонстрирована значительная разница в оптимальном выборе стратегии комбинирования в зависимости от модели входных

данных: на тестовых наборах, в которых встречаются ошибки локализации и сегментации символов, более высокую точность распознавания видеопоследовательностей обеспечивают правило произведения (3), голосование (9) и правило суммы (4) (Рис. 2а, в). При этом на тестовых наборах, в которых такого типа ошибки были исключены (Рис. 2б, г), более высокую точность распознавания обеспечивает правило максимума (6). Другими словами, при рассмотрении данной задачи как задачи коллективного принятия решения, в случае более строгой модели входных данных (с отсутствием ошибок локализации и сегментации символов) выгоднее доверять единственному максимально компетентному эксперту, нежели коллективному мнению нескольких экспертов.

При наличии ошибок локализации и сегментации символов устойчивость предикторов до-

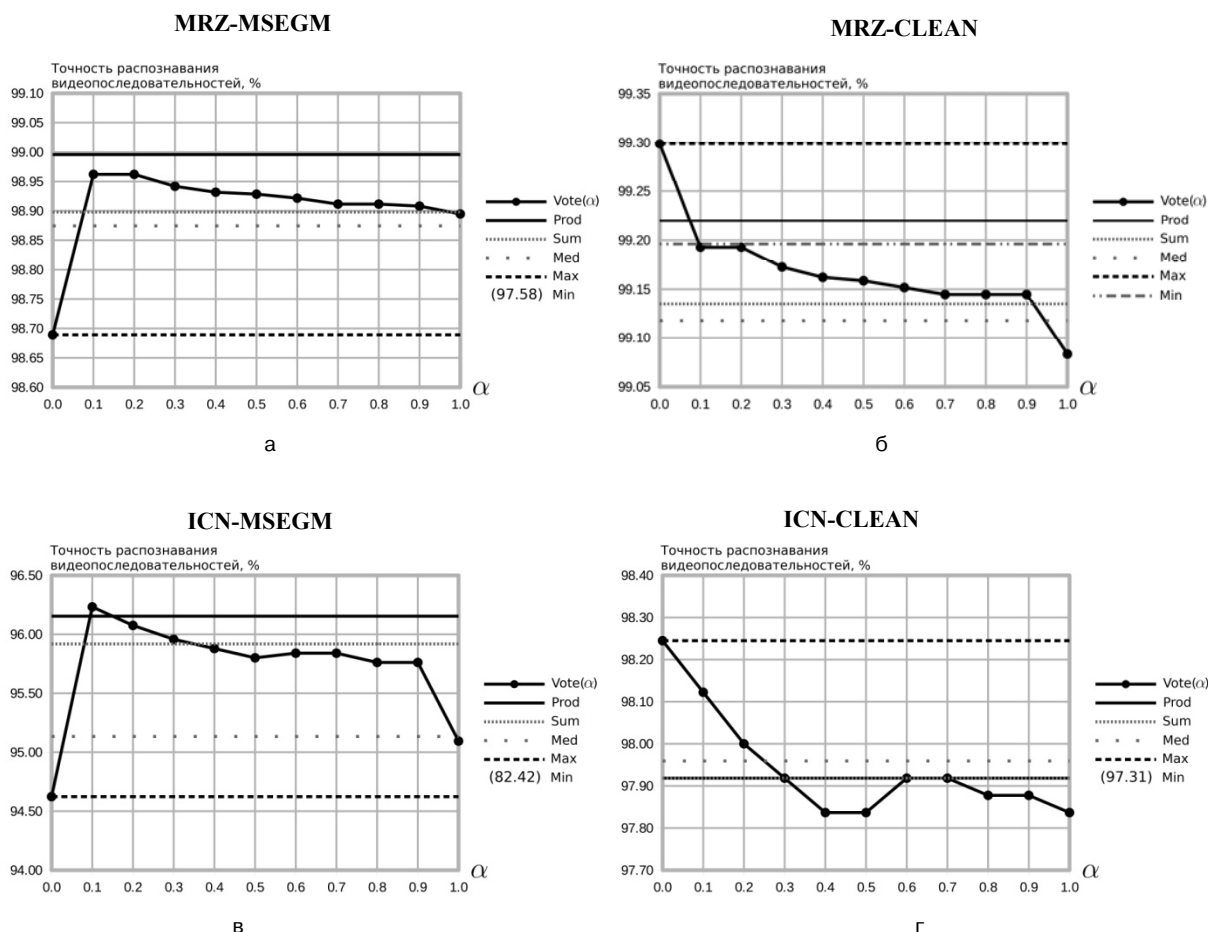


Рис. 2. Сравнение точности распознавания видеопоследовательностей символов с использованием базовых стратегий комбинирования

стоверности уменьшается, что в свою очередь увеличивает разницу между оценками компетентности экспертов (которые конструируются на основе значений предикторов) и действительными значениями компетентности экспертов (которые соответствуют апостериорным вероятностям принятия правильного решения). В таком случае выбор эксперта с максимальным уровнем компетентности чаще бывает ошибочным и таким образом разница между уровнем действительной компетентности выбранного эксперта и уровнями компетентности остальных экспертов сокращается. Таким образом, оптимальность выбора наилучшего (с точки зрения устойчивого предиктора достоверности) кадрового результата в случае, когда в видеопоследовательности отсутствуют ошибки локализации и сегментации символов, соответствует более широкому положению теории коллективного принятия решения [25, 28], согласно которому нарушение первой части утверждения Кондорсе (при увеличении количества экспертов вероятность коллективного принятия правильного решения увеличивается, если для каждого эксперта вероятность принятия правильного индивидуального решения выше, чем вероятность принятия неправильного решения) происходит при увеличении разницы между уровнями компетентности максимально компетентного эксперта и остальных.

#### 4. Принятие решения подмножеством наиболее компетентных экспертов

В работе [16, 18] отмечается преимущество методов взвешенных комбинаций классификаторов, в рамках которых комбинирование гетерогенных классификаторов производится с учетом параметров значимости каждого из классификаторов. При рассмотрении задачи построения оптимальной стратегии комбинирования результатов кадрового распознавания символов как задачи коллективного принятия решения с  $N$  экспертами представляет интерес сокращение количества экспертов до  $T$  наиболее компетентных.

Пусть  $\tilde{X}_{fc}^T = \{\tilde{x}_1, \dots, \tilde{x}_T\} \subseteq X$ ,  $1 \leq T \leq N$  — последовательность наблюдений такая, что

$$i < j \Rightarrow f_c(\tilde{x}_i) \geq f_c(\tilde{x}_j) \quad \text{и}$$

$\forall i: T < i \leq N \Rightarrow f_c(\tilde{x}_T) \geq f_c(x_i)$ . Другими словами,  $\tilde{X}_{fc}^T$  является подмножеством входных наблюдений, частично упорядоченным по убыванию предиктора достоверности (по убыванию оценки компетентности экспертов) и содержащим  $T$  наблюдений с наибольшим значением предиктора достоверности (с наибольшей оценкой компетентности экспертов). В качестве стратегии комбинирования результатов распознавания образов из  $X$  предлагается применять описанные выше правила комбинирования, учитывая только наблюдения из  $\tilde{X}_{fc}^T$ . На Рис. 3 представлены сравнительные значения точности распознавания видеопоследовательностей с использованием правил комбинирования (3)-(7) на тестовых наборах данных MRZ-MSEGM, MRZ-CLEAN, ICN-MSEGM и ICN-CLEAN с использованием предлагаемой стратегии. Горизонтальная ось графиков соответствует значению  $T$ . Значение  $\alpha = 0,1$  для правила комбинирования  $\text{Vote}(\alpha)$  (9) было выбрано как обеспечивающее в среднем наилучший результат обобщенного голосования (Рис. 2).

На Рис. 3 продемонстрировано падение точности распознавания видеопоследовательностей с использованием правил произведения (3), суммы (4), минимума (5), медианы (7) и обобщенного голосования (9) по сравнению с правилом максимума (6) на тестовых наборах MRZ-CLEAN и ICN-CLEAN, на которых отсутствуют ошибки локализации и сегментации символов (Рис. 3б, 3г), при увеличении количества наиболее компетентных экспертов  $T$ . При этом на тестовых наборах MRZ-MSEGM и ICN-MSEGM, содержащих подобные ошибки, правила (3), (4), (7) и (9) показывают более точные результаты по сравнению с правилом максимума (6). На тестовом наборе ICN-MSEGM (Рис. 3в) можно наблюдать падение точности распознавания после достижения максимума при  $T = 3$ , что соответствует преимуществу использования подмножества наиболее компетентных экспертов для этого тестового набора данных.

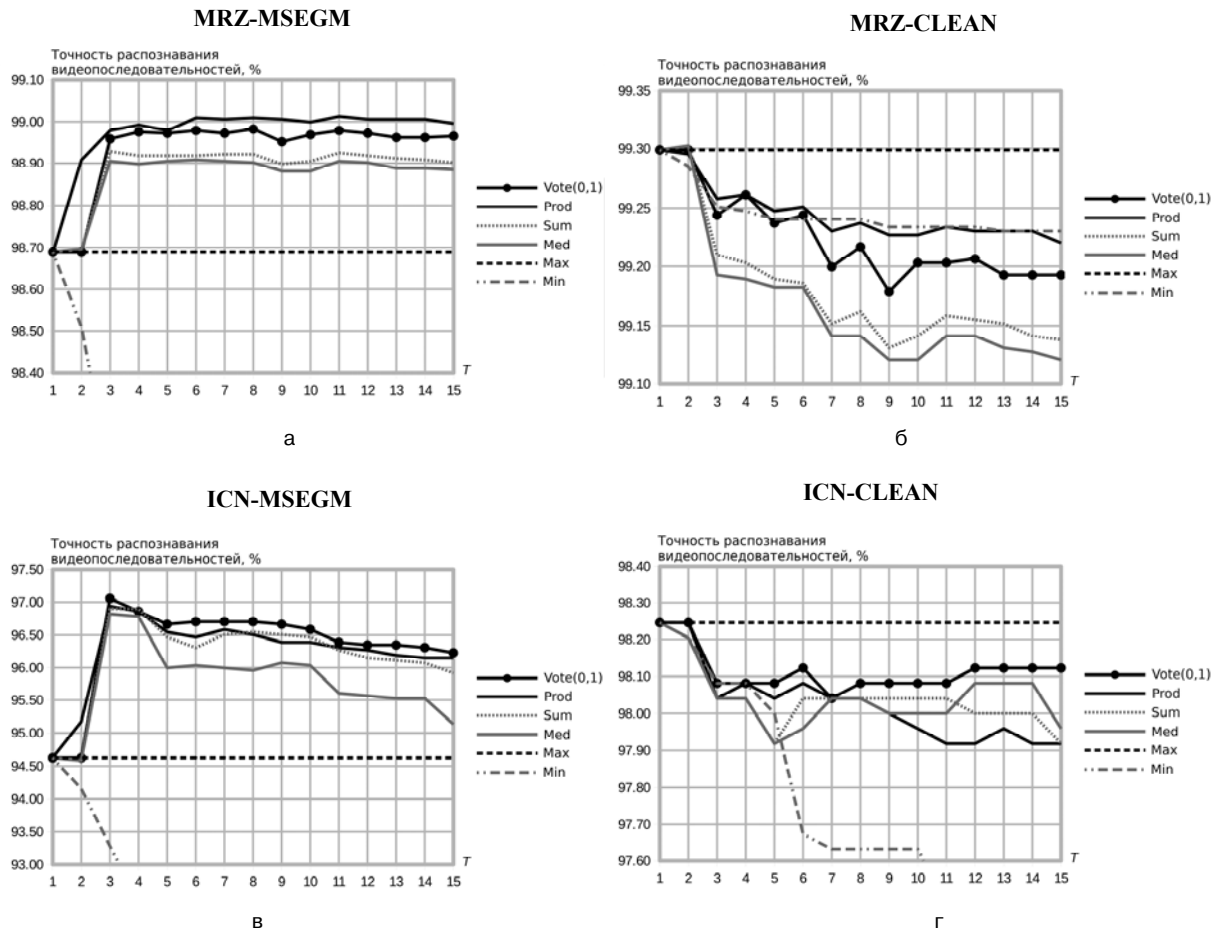


Рис. 3. Сравнение точности распознавания видеопоследовательностей символов с использованием стратегий комбинирования подмножеств  $\tilde{X}_{fC}^T$

### Заключение

В работе рассматривалась задача выбора оптимальной стратегии комбинирования результатов покадрового распознавания для достижения наилучшей точности распознавания видеопоследовательностей символов. Комбинирование результатов покадрового распознавания символов рассматривалась как задача коллективного принятия решения с множеством экспертов, уровни компетентности которых оценивались при помощи предиктора достоверности результата распознавания. Рассматривались правила комбинирования результатов классификаторов в Байесовской интерпретации: правило произведения (3), суммы (4), минимума (5), максимума (6), медианы (7) [13] и обобщенного голосования (9) [9, 12]. Было показано, что при наличии ошибок сегмента-

ции и локализации символа, которые приводят к появлению нерелевантных покадровых результатов в рамках видеопоследовательности, наилучшие результаты комбинирования могут быть достигнуты при помощи применения правил произведения и обобщенного голосования для подмножества нескольких самых компетентных экспертов. При этом в случае отсутствия ошибок сегментации и локализации символа наилучшую точность показывает правило максимума.

В рамках дальнейшей работы предполагается рассмотреть правила комбинирования с интерпретацией оценок одиночных классификаторов как нечетких свидетельств принадлежности к классам либо как абстрактных показателей уверенности, с применением аппарата теории Демпстера-Шафера [14, 15], а также рассмотреть процесс распознавания видеопоследовательности как



динамический процесс, в рамках которого, помимо задачи оптимального комбинирования результатов покadroвого распознавания, возникает задача останова.

## Литература

1. Doermann D., Liang J., Huiping L. Progress in camera-based document image analysis // Proceedings of Seventh International Conference on Document Analysis and Recognition, Vol. 1, 2003, pp. 606-616.
2. В.В. Арлазаров, А. Жуковский, В. Кривцов, Д. Николаев, Д. Полевой. Анализ особенностей использования стационарных и мобильных малоразмерных цифровых видео камер для распознавания документов // Информационные технологии и вычислительные системы, № 3, 2014, С. 71-78.
3. К.Б. Булатов, Д.А. Ильин, Д.В. Полевой, Ю.С. Чернышова. Проблемы распознавания машиночитаемых зон с использованием малоформатных цифровых камер мобильных устройств // Труды ИСА РАН, Том 65/3, 2015, С. 85-93.
4. Т.С. Чернов, Д.А. Ильин, П.В. Безматерных, И.А. Фараджев, С.М. Карпенко. Исследование методов сегментации изображений текстовых блоков документов с помощью алгоритмов структурного анализа и машинного обучения // Вестник РФФИ. Обработка изображений и распознавание образов, № 4 (92), 2016, С. 55-71.
5. Rokach L. Ensemble-based classifiers. Artificial Intelligence Review, V. 33, Issue 1-2, 2010, pp. 1-39, doi:10.1007/s10462-009-9124-7.
6. Fumera G., Roli F. Linear combiners for classifier fusion: some theoretical and experimental results // Proceedings of the 4th International Conference on Multiple Classifier Systems, 2003, pp. 74-83.
7. Schwenk H., Gauvain J.-L. Combining multiple speech recognizers using voting and language model information // IEEE International Conference on Speech and Language Processing, 2000, pp. 915-918.
8. Zhang C.-X., Duin R. An experimental study of one- and two- level classifier fusion for different sample sizes // Pattern Recognition Letters 32, 2011, pp. 1756-1767, doi:10.1016/j.patrec.2011.07.009.
9. Chen D. Text detection and recognition in images and video sequences. Thesis №2863, Lausanne, EPFL, 2003, 141 pages.
10. Wemhoener D., Yalniz I.Z., Manmatha R. Creating an improved version using noisy OCR from multiple editions // Proceedings of the 12th International Conference on Document Analysis and Recognition, 2013, pp. 160-164.
11. Lopresti D., Zhou J. Using consensus sequence voting to correct OCR errors // Computer Vision and Image Understanding, V. 67(1), 1997, pp. 39-47.
12. Fiscus J.G. A post-processing system to yield reduced word error rates: Recognizer Output Voting Error Reduction (ROVER) // Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding, 1997, pp. 347-354, doi:10.1109/ASRU.1997.659110.
13. Kittler J. On combining classifiers // IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, № 3, 1998, pp. 226-239.
14. Rogova G. Combining the results of several neural network classifiers // Neural Networks, Vol. 7, № 5, pp. 777-781, 1994.
15. Quost B., Masson M.-H., Denoeux T. Classifier fusion in the Dempster-Shafer framework using optimized t-norm based combination rules // International Journal of Approximate Reasoning, Vol. 52, Issue 3, 2011, pp. 353-374.
16. Ting K.M., Witten I.H. Issues in stacked generalization // Journal of Artificial Intelligence Research, Vol. 10, 1999, pp. 271-289.
17. Kuncheva L.I., Bezdek J.C., Duin R.P.W. Decision templates for multiple classifier fusion: an experimental comparison // Pattern Recognition, Vol. 34, 2001, pp. 299-314.
18. Merz C. Using correspondence analysis to combine classifiers // Machine Learning, Vol. 36, 1999, pp. 33-58.
19. Nguyen T.T., Nguyen T.T.T., Pham X.C., Liew A.W.-C. A novel combining classifier method based on variational inference // Pattern Recognition, Vol. 49, 2016, pp. 198-212.
20. Ye P., Doermann D. Document image quality assessment: a brief survey // Proceedings of 12th International Conference on Document Analysis and Recognition, 2013, pp. 723-727.
21. Bulatov K., Polevoy D. Reducing overconfidence in neural networks by dynamic variation of recognizer relevance // Proceedings of 29th European Conference on Modelling and Simulation, 2015, pp. 488-491.
22. Д.П. Николаев, Д.В. Полевой, Т.С. Чернов. Метод автоматической оценки качества цветовой сегментации в задаче упаковки изображений печатных документов // Труды ИСА РАН, Том 63/3, 2013, С. 78-84.
23. Арлазаров В.Л., Емельянов Н.Е. (Ред.) Документооборот. Прикладные аспекты. М.: Едиториал УРСС, 2005, 184 с.
24. В.В. Арлазаров, К.Б. Булатов, С.М. Карпенко. Метод определения надежности распознавания в задаче распознавания тисненых символов // Труды ИСА РАН, Том 63/3, 2013, С. 117-122.
25. Berend D., Kontorovich A. Consistency of weighted majority votes // Advances in Neural Information Processing Systems, Vol. 27, Issue 2, 2014, pp. 3446-3454.
26. Dzeroski S., Zenko B. Is combining classifiers better than selecting the best one? // Machine Learning, Vol. 45, Issue 3, 2004, pp. 255-273.
27. Ilin D., Krivtsov V. Creating training datasets for OCR in mobile video stream // Proceedings of 29th European Conference on Modelling and Simulation, 2015, pp. 516-520.
28. Nitzan S., Paroush J. Collective decision-making and jury theorems // The Oxford Handbook of Law and Economics: Volume 1: Methodology and Concepts, 2017, doi:10.1093/oxfordhb/9780199684267.013.035.

**Булатов Константин Булатович.** Младший научный сотрудник ФИЦ ИУ РАН. Аспирант Национального исследовательского технологического университета «МИСиС». Окончил НИТУ «МИСиС» в 2013 году. Количество печатных работ: 10. Область научных интересов: распознавание образов, машинное обучение, информационные системы. E-mail: hpbuko@gmail.com

## Selecting optimal strategy for combining per-frame character recognition results in video stream

K.B. Bulatov

**Abstract.** This paper considers a problem of combining classification results from several observations of the same object. The task is seen as a case of collective decision making by a group of experts with estimated competence levels. Precision of different classification result combination methods is analyzed with different input data model, having per-frame character recognition results combination problem in video stream as an example. Experiments show that the strategy which selects a single most competent expert performs better with input data model without any non-relevant observations (in the context of character recognition in video stream — without characters location and segmentation errors). At the same time experiments show that strategies which combine several most competent experts using product rule or voting procedure outperform single-expert strategy with input data containing non-relevant observations.

**Keywords:** decision theory, pattern recognition, recognition in video stream, ensemble classifiers.

### References

1. Doermann D., Liang J., Huiping L. Proceedings of Seventh International Conference on Document Analysis and Recognition, Vol. 1, 2003, pp. 606-616.
2. V.V. Arlazarov, A. Zhukovsky, D. Krivtsov, D. Nikolaev, D. Polevoy. Analysis of using stationary and mobile small-scale digital cameras for documents recognition // Information technologies and computational systems, № 3, 2014, pp. 71-78.
3. K.B. Bulatov, D.A. Ilin, D.V. Polevoy, Yu.S. Chernyshova. Problems of machine-readable zones recognition using small-scale digital cameras of mobile devices // Proc. ISA RAS, Vol. 65/3, 2015, pp. 85-93.
4. T.S. Chernov, D.A. Ilin, P.V. Bezmaternykh, I.A. Faradjev, S.M. Karpenko. Research of methods for segmentation of document text block images using algorithms of structure analysis and machine learning // Proc. RFBR. Image processing and pattern recognition, № 4 (92), 2016, pp. 55-71.
5. Rokach L. Ensemble-based classifiers. Artificial Intelligence Review, V. 33, Issue 1-2, 2010, pp. 1-39, doi:10.1007/s10462-009-9124-7.
6. Fumera G., Roli F. Linear combiners for classifier fusion: some theoretical and experimental results // Proceedings of the 4th International Conference on Multiple Classifier Systems, 2003, pp. 74-83.
7. Schwenk H., Gauvain J.-L. Combining multiple speech recognizers using voting and language model information // IEEE International Conference on Speech and Language Processing, 2000, pp. 915-918.
8. Zhang C.-X., Duin R. An experimental study of one- and two- level classifier fusion for different sample sizes // Pattern Recognition Letters 32, 2011, pp. 1756-1767, doi:10.1016/j.patrec.2011.07.009.
9. Chen D. Text detection and recognition in images and video sequences. Thesis №2863, Lausanne, EPFL, 2003, 141 pages.
10. Wemhoener D., Yalniz I.Z., Manmatha R. Creating an improved version using noisy OCR from multiple editions // Proceedings of the 12th International Conference on Document Analysis and Recognition, 2013, pp. 160-164.
11. Lopresti D., Zhou J. Using consensus sequence voting to correct OCR errors // Computer Vision and Image Understanding, V. 67(1), 1997, pp. 39-47.
12. Fiscus J.G. A post-processing system to yield reduced word error rates: Recognizer Output Voting Error Reduction (ROVER) // Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding, 1997, pp. 347-354, doi:10.1109/ASRU.1997.659110.
13. Kittler J. On combining classifiers // IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, № 3, 1998, pp. 226-239.
14. Rogova G. Combining the results of several neural network classifiers // Neural Networks, Vol. 7, № 5, pp. 777-781, 1994.
15. Quost B., Masson M.-H., Denooux T. Classifier fusion in the Dempster-Shafer framework using optimized t-norm based combination rules // International Journal of Approximate Reasoning, Vol. 52, Issue 3, 2011, pp. 353-374.
16. Ting K.M., Witten I.H. Issues in stacked generalization // Journal of Artificial Intelligence Research, Vol. 10, 1999, pp. 271-289.
17. Kuncheva L.I., Bezdek J.C., Duin R.P.W. Decision templates for multiple classifier fusion: an experimental comparison // Pattern Recognition, Vol. 34, 2001, pp. 299-314.
18. Merz C. Using correspondence analysis to combine classifiers // Machine Learning, Vol. 36, 1999, pp. 33-58.
19. Nguyen T.T., Nguyen T.T.T., Pham X.C., Liew A.W.-C. A novel combining classifier method based on variational inference // Pattern Recognition, Vol. 49, 2016, pp. 198-212.
20. Ye P., Doermann D. Document image quality assessment: a brief survey // Proceedings of 12th International Conference on Document Analysis and Recognition, 2013, pp. 723-727.
21. Bulatov K., Polevoy D. Reducing overconfidence in neural networks by dynamic variation of recognizer relevance // Proceedings of 29th European Conference on Modelling and Simulation, 2015, pp. 488-491.

22. D.P. Nikolaev, D.V. Polevoy, T.S. Chernov. Method for automatic quality estimation of color segmentation within a problem of compressing images of printed documents // Trudy ISA RAN [Proc. ISA RAS], Vol. 63/3, 2013, pp. 78-84.
23. Arlazarov V.L., Emelyanov N.E. (Ed.) Documents flow. Applied aspects. Moscow, Editorial URSS, 2005, 184 p.
24. V.V. Arlazarov, K.B. Bulatov, S.M. Karpenko. Method for determining recognition confidence in the problem of embossed characters recognition // Proc. ISA RAS, Vol. 63/3, 2013, pp. 117-122.
25. Berend D., Kontorovich A. Consistency of weighted majority votes // Advances in Neural Information Processing Systems, Vol. 27, Issue 2, 2014, pp. 3446-3454.
26. Dzeroski S., Zenko B. Is combining classifiers better than selecting the best one? // Machine Learning, Vol. 45, Issue 3, 2004, pp. 255-273.
27. Ilin D., Krivtsov V. Creating training datasets for OCR in mobile video stream // Proceedings of 29th European Conference on Modelling and Simulation, 2015, pp. 516-520.
28. Nitzan S., Paroush J. Collective decision-making and jury theorems // The Oxford Handbook of Law and Economics: Volume 1: Methodology and Concepts, 2017, doi:10.1093/oxfordhb/9780199684267.013.035.

**K. B. Bulatov.** Junior researcher at Federal Research Center "Computer Science and Control" of Russian Academy of Sciences. Post-graduate at National University of Science and Technology "MISIS", graduated from NUST "MISIS" in 2013. Number of publications: 10. Fields of interest: pattern recognition, machine learning, information systems. E-mail: hpbuko@gmail.com