

Найдем мгновенный спектр пространственных частот пространственно-временного сигнала (9) в апертуре АОЯ:

$$R_D(\Omega_z, z, Vt) = \frac{1}{\pi} e^{-\alpha(\Omega_z)z} \int_{-\infty}^{\infty} s(\Omega'_z) K_{АП}(\Omega'_z) \frac{\sin(\Omega_z - \Omega'_z)(\frac{D}{2})}{\Omega_z - \Omega'_z} e^{j(\Omega_z - \Omega'_z)Vt} d\Omega'_z, \quad (10)$$

причем:

$$p(z)s(Vt, z) = (1/2\pi) \int_{-\infty}^{\infty} R_D(\Omega_z, z, Vt) e^{-j\Omega_z(Vt-z)} d\Omega_z, \quad (11)$$

– пространственно-временной сигнал в апертуре АОЯ.

**Выводы.** Таким образом, следствием акустооптического взаимодействия в режимах дифракции Рамана-Ната, Брэгга и промежуточном является фазовая модуляция выходного лазерного излучения. Амплитудная характеристика акустооптического взаимодействия в режиме дифракции Рамана-Ната определяется выражением (2), в режиме Брэгга - соотношением (3).

Ширина полосы акустических частот (частотная характеристика) акустооптического взаимодействия, при которой реализуется режим дифракции Брэгга, соответствует выражению (7).

Индекс фазовой модуляции, информационной технологии акустооптического взаимодействия, численно равный  $\Delta\varphi \leq 1$ , может быть использован как критерий гомоморфности преобразований информации при слабом АОВ.

Информационная технология преобразования радиосигналов при акустооптическом взаимодействии, предполагает, что радиосигнал временной природы сначала преобразуется в акустический аналог пространственно-временной природы, а затем - в выходной световой аналог пространственно-временной природы (9), (10), причем существует жесткая связь между частотой радиосигнала и пространственной частотой его светового аналога (11).

## СПИСОК ЛИТЕРАТУРЫ

1. Lucas R. Optical properties of solids and liquids, under ultrasonic vibrations / R. Lucas, P. Biquard // Journ. Phys. Rad. – 1932. – Vol. 3. – P. 464-472.
2. Debye P. On the scattering of light by supersonic waves / P. Debye, F. Sears // Proc. Nat. Acad. Sciences. – 1932. – Vol. 18. – P. 409-413.
3. Raman C.V. The diffraction of light by high frequency sound waves / C.V. Raman, N.S.N. Nallh // Proc. Indian Acad. Sciences. – 1936. – Vol. 92. – P. 406-420.
4. Рытов С.М. Дифракция света на ультразвуковых волнах / С.М. Рытов // Изв. АН СССР. Сер. Физика. – № 2. – С. 222-242.
5. Klein W.R. Unified approach to ultrasonic light diffraction / W.R. Klein, B.D. Cook // IEEE Trans. – 1967. – SU-14. No. 3. – P. 123-134.
6. Балакший В.И. Физические основы акустооптики / В.И. Балакший, В.Н. Парыгин, Л.Е. Чирков. – М.: Радио и связь, 1985. – С. 224-240.
7. Данилов В.В. Классификационный анализ устройств управления лазерным пучком / В.В. Данилов // Технология и конструирование в электронной аппаратуре. – 2000. – № 2-3. – С. 52-57.
8. Анго А. Математика для электро- и радиоинженеров / А. Анго. – М.: Наука, 1965. – 772 с.
9. Ультразвуковые преобразователи для неразрушающего контроля / Под ред. И.Н. Ермолова. – М.: Машиностроение, 1986. – 356 с.
10. Ярив А. Введение в оптическую электронику / А. Ярив. – М.: Высш. школа, 1983. – 398 с.
11. Най Д. Физические свойства кристаллов и их описание при помощи тензоров и матриц / Д. Най. – М.: Мир, 1967. – 386 с.

12. Харкевич А.А. Спектры и анализ / А.А. Харкевич. – М.: Гостехиздат, 1953. – 236 с.
13. Данилов В.В. Моделирование акустически нагруженного пьезопреобразователя акустооптической ячейки / В.В. Данилов, С.В. Иванов // Радиотехника. – 2000. – № 115. – С. 66-70.

*Поступила в редакцию 18.04.2016 г.*

## OPTICAL INFORMATION TECHNOLOGIES. TRANSFORMATION OF RADIO SIGNALS

*I. Tretyakov, M. Pavlov, T. Tiebienieva, V. Danilov*

The article gives the results of optical information technology's research, in particular, transformation of radio signals in their laser analogies with minimum garbling by using acousto-optical interaction

**Keywords:** homomorphy, acousto-optical interaction, acousto-optic cell, diffraction, The Klein-Cook parameter.

**Третьяков Игорь Александрович**

Донецкий национальный университет, г. Донецк  
Аспирант  
es75i@mail.ru

**Павлов Максим Леонидович**

Донецкий национальный университет, г. Донецк  
Студент  
maksim-pavlov199294@mail.ru

**Тебенева Татьяна Сергеевна**

Донецкий национальный университет, г. Донецк  
Студентка  
tetasia19@gmail.com

**Данилов Владимир Васильевич**

Доктор технических наук, профессор  
ut5iv@mail.ru  
Контактный телефон: +38 066 605 67 62

**Tretyakov Igor**

Donetsk National University, Donetsk  
Postgraduate

**Pavlov Maxim**

Donetsk National University, Donetsk  
Student

**Tiebienieva Tetiana**

Donetsk National University, Donetsk  
Student

**Danilov Vladimir**

Donetsk National University, Donetsk  
Doctor of Engineering Sciences, Full Professor

## РАСПОЗНАВАНИЕ ТЕКСТА В ЗАШУМЛЕННЫХ ИЗОБРАЖЕНИЯХ СКАНИРОВАННЫХ ДОКУМЕНТОВ

© 2016. *Т.В. Шарий, Р.О. Лялин, А.Е. Гукай, В.Н. Котенко*

В настоящее время по-прежнему актуальна проблема перевода сканированных документов относительно низкого качества в текстовый формат. В статье предлагается робастный метод распознавания текстовых символов в зашумленных изображениях. Описаны алгоритмы, применяемые для очистки изображения от шумов, выделения контуров текстовых областей, сегментации текста по буквам. Приведены результаты экспериментов по распознаванию текста на основе статистической модели машин опорных векторов.

**Ключевые слова:** OCR, машинное обучение, медианный фильтр, бинаризация, сегментация.

**Введение.** Оптическое распознавание символов (Optical Character Recognition, OCR) [1–4] остается на протяжении многих лет актуальной задачей с точки зрения как бизнеса, так и развития научно-технического направления компьютерного зрения. Суть OCR-процесса состоит, прежде всего, в переводе цифровых изображений в другие цифровые форматы, более подходящие для редактирования, поиска, реферирования информации, а именно: непосредственно текст и метаданные о документе (количество слов, абзацев, таблиц, подписей и т.д.). С каждым годом число изображений на локальных компьютерах и в сети интернет стремительно растет, и все больше документов нуждается в оцифровке и постобработке. В данном контексте OCR является элементом передового направления разработок в современной IT-сфере – «больших данных» (Big Data and Data Science), в рамках которого большие объемы данных должны эффективно храниться и подготавливаться для оперативного анализа. Также к вариантам применения OCR можно отнести: автоматический ввод данных из бланков разного рода в компьютер, автоматическое распознавание автомобильных номеров, программы-помощники для лиц с нарушениями зрения и др.

В настоящее время существуют программные решения для распознавания текста в изображениях и преобразования форматов файлов, среди которых следует отметить коммерческие приложения [5]: OmniPage, Adobe Acrobat, ABBYY FineReader, ReadIris, PowerPDF, SodaPDF. Точность распознавания текста в лучших из них достигает 97–99 % для качественных сканов документов. Тем не менее, ряд проблем значительно ухудшает качество OCR даже самых эффективных программ. Эти проблемы вызваны искажением изображений фоновым шумом (пятна от чашек с кофе, завернутые уголки, следы от скрепок и т.д.), возможным поворотом текста, вариабельностью цветов и начертаний шрифтов, языков. Таким образом, предварительная обработка зашумленного сканированного изображения выходит на первый план в современных информационных технологиях хранения и обработки цифровых документов.

**Постановка задачи.** Целью данной работы является разработка и исследование робастного метода распознавания текста в изображениях отсканированных документов, содержащих шумы. Для визуализации и автоматизации проведения исследований необходимо создать соответствующий программный инструментарий. В статье акцент делается на предварительной цифровой обработке изображения. Этап распознавания также рассмотрен, однако подробный анализ и разработка статистических моделей машинного обучения, применяемых на этом этапе, является предметом дальнейших исследований.

**Цифровая обработка зашумленного сканированного изображения.** В статье за основу взята типовая схема процесса OCR, включающая блок предварительной обработки скана документа и блок распознавания символов. В первом блоке решаются следующие задачи (рис.1):

- 1) фильтрация шумов и удаление фона;
- 2) бинаризация изображения;
- 3) идентификация текстовой области документа;
- 4) поворот текстовой области при необходимости (выравнивание документа);
- 5) сегментация (горизонтальная сегментация по строкам текста и вертикальная сегментация по символам).

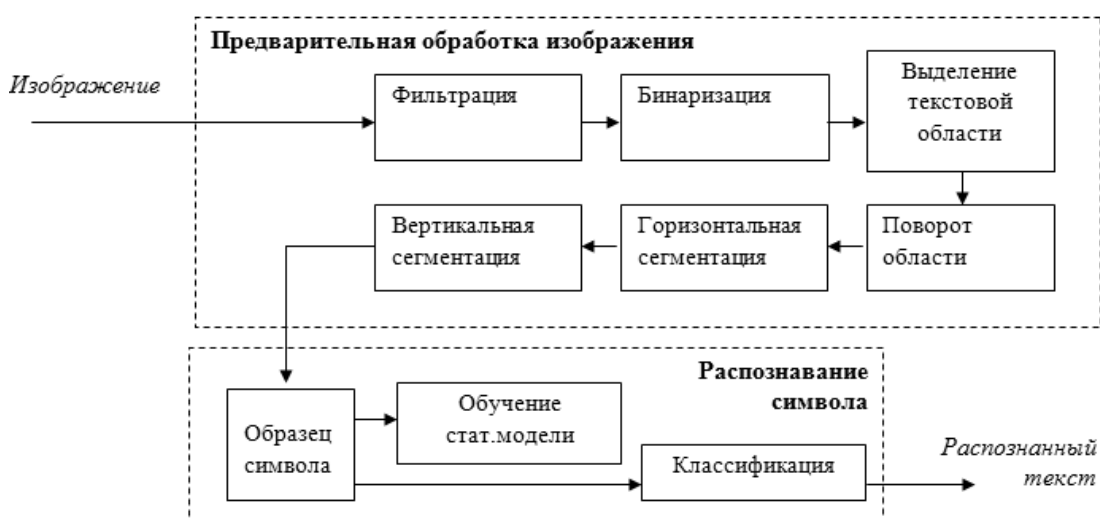


Рис. 1. Схема OCR

Вся описываемая далее цифровая обработка производится над изображением документа в шкале серого цвета с нормированными значениями пикселей. На первом этапе в работе производится медианная фильтрация. Медианный фильтр является распространенным простым средством удаления импульсных шумов в сигнале. Данный фильтр представляет собой нелинейный низкочастотный фильтр, заменяющий значение каждого пикселя изображения медианным значением пикселей из его окрестности, подавляя, таким образом, выбросы:

$$y[m, n] = \text{median} \{x[i, j], (i, j) \in \omega_N\},$$

где  $x$  – исходное изображение;  $y$  – отфильтрованное изображение;  $m, n$  – координаты текущего центрального пикселя;  $\omega_N$  – окрестность пикселя размера  $N$ .

Результатом медианной фильтрации является изображение, которое можно считать фоном, т.к. данная операция сохраняет медленно меняющиеся признаки и удаляет высокочастотные компоненты изображения, представленные самим текстом. Таким образом, для получения текста можно вычесть изображение фона из исходного изображения скана или оставить только те пиксели, цвет которых темнее фона. При этом, в работе, по аналогии с предложенным в [6] алгоритмом, используется также некоторый порог для уменьшения влияния шумов и отбрасывания незначущих для текста пикселей:

$$t[m, n] = \begin{cases} x[m, n], & \text{если } x[m, n] < y[m, n] - \delta_S \\ 1.0, & \text{иначе} \end{cases},$$

где  $t$  – результирующее изображение, содержащее текст;  $m, n$  – координаты текущего пикселя;  $\delta_s$  – порог вычитания. Экспериментально подобраны значения параметров  $N = 7$  и  $\delta_s = 0.1$ , при которых текст на тестовых сканах документов субъективно воспринимается наиболее отчетливо.

Результаты для двух изображений с текстом приведены на рис. 2. На рис. 2, а и 2, б показаны сканы документов, на рис. 2, в и 2, д – фон как результат работы медианного фильтра, на рис. 2, е и 2, е – разность изображения и фона, являющаяся искомым текстом.

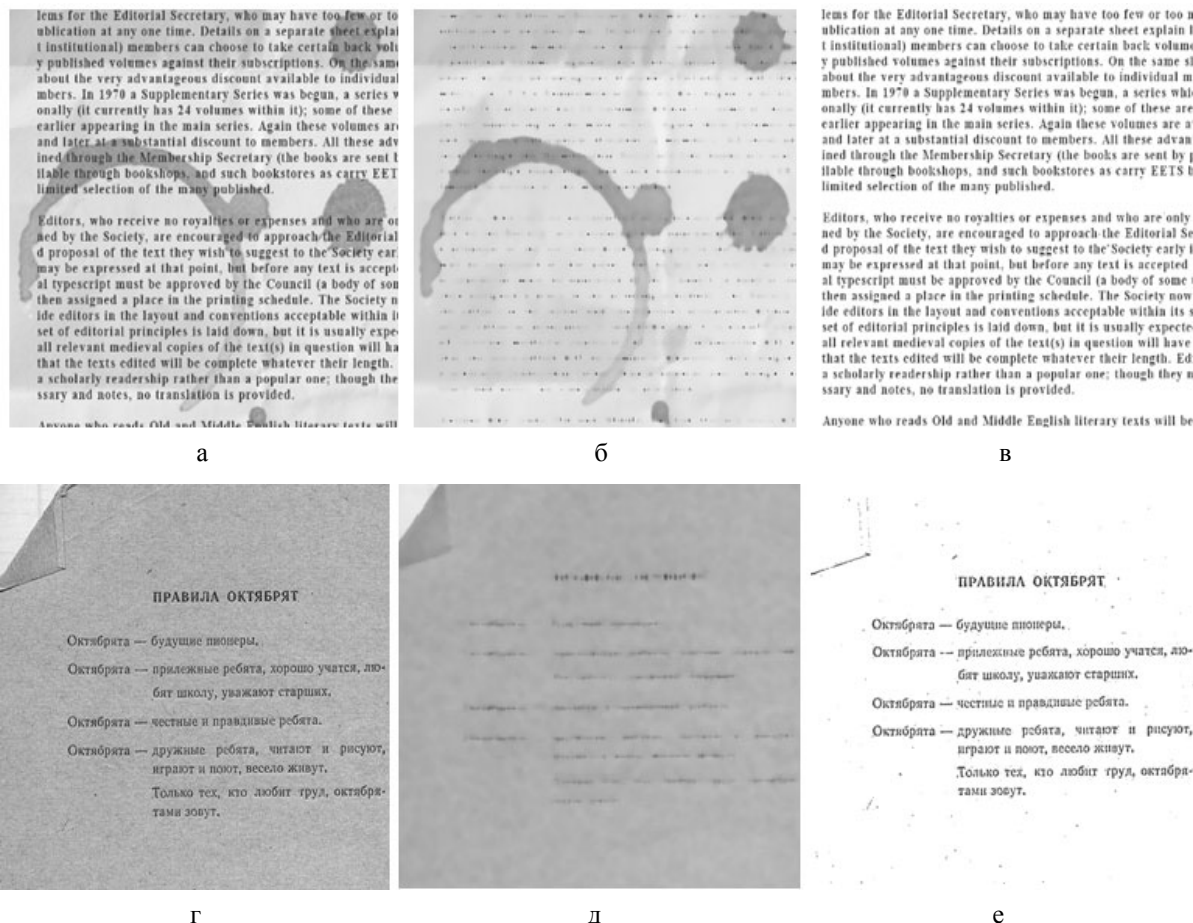


Рис. 2. Примеры результатов медианной фильтрации сканированных изображений:  
а) тестовое изображение №1; б) выделенный фон в изображении №1; в) выделенный текст в изображении №1; г) тестовое изображение №2; д) выделенный фон в изображении №2; е) выделенный текст в изображении №2

Следующим шагом при цифровой обработке изображения документа является его бинаризация. В работе используется следующий алгоритм бинаризации:

1. Изображение фильтруется с помощью гауссовского фильтра с ядром размером  $L = 11$  пикселей и стандартным отклонением  $\sigma = 2$ :

$$G[m, n] = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{m^2 + n^2}{2\sigma^2}\right).$$

2. В отфильтрованном изображении рассчитывается порог для бинаризации на основе алгоритма Оцу [7]. Алгоритм позволяет вычислить на основе гистограммы

изображения оптимальный глобальный порог для разделения всех пикселей на бинарные группы.

- По результатам сравнения с порогом, вычисленным на шаге 2, пикселям изображения присваивается значение 0 или 1 (черный или белый цвет).

Приведенный алгоритм позволяет улучшить качество бинаризации по сравнению со стандартным глобальным алгоритмом Оцу за счет удаления лишних теней.

Далее в бинарном изображении осуществляется идентификация области, содержащей текст. В большинстве случаев этой областью является абзац. Для этих целей применяется алгоритм выделения контуров, предложенный в [8]. Алгоритм относится к классу алгоритмов следования границам (Border following), позволяет отслеживать как внешние, так и внутренние границы, и может быть использован для широкого круга задач топологического анализа изображений. Об эффективности алгоритма говорит, в частности, тот факт, что он реализован в функции `findContours` популярной библиотеки компьютерного зрения OpenCV [9]. Пример результата выделения границ текстовых областей приведен на рис. 3.

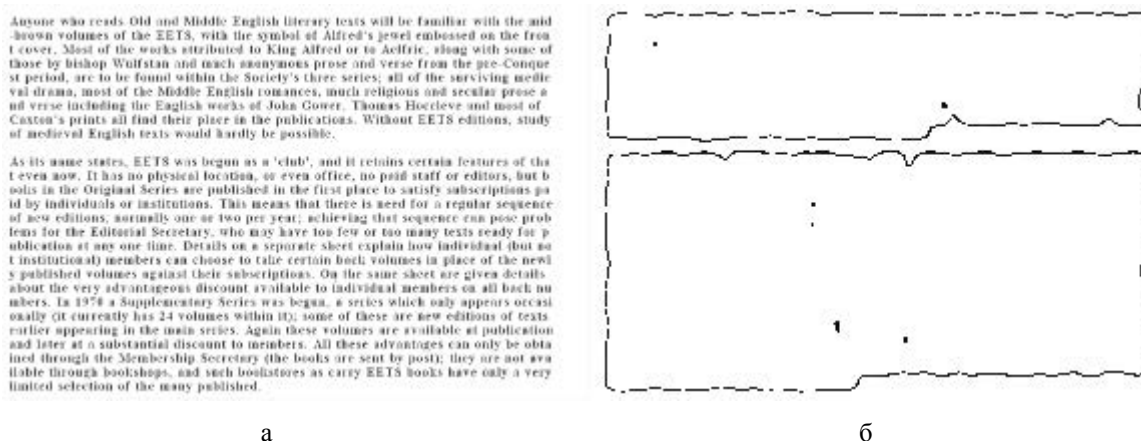


Рис. 3. Пример результатов выделения текстовых областей:

а) исходное изображение; б) выделенные контуры текстовых областей

Важной частью предлагаемого метода распознавания текста в зашумленных сканах документов является обнаружение наклона и поворот выделенной текстовой области. Для этого в работе на основе координат левого верхнего, правого верхнего, левого нижнего и правого нижнего пикселей контура текстовой области определяется угол наклона и производится поворот области на найденный угол (преобразование на основе стандартной матрицы поворота).

Завершающим шагом цифровой обработки изображения сканированного документа перед распознаванием символов является сегментация. В качестве единицы распознавания выбрана буква как алфавитная единица любого естественного языка, а также двухбуквенные сочетания, которые в данной статье не рассматриваются и являются предметом дальнейших исследований. В связи с этим, необходимо разделять абзацы сначала на строки (горизонтальная сегментация), а затем на буквы (вертикальная сегментация). При горизонтальной сегментации в каждой строке пикселей изображения вычисляется среднее значение пикселей, и если оно превышает пороговое значение, то полагается, что текущая линия принадлежит пустой области, разделяющей строки текста. Термин «строка пикселей» необходимо отличать от термина «строка текста». Первый относится к изображению, второй – к распознаваемому тексту на изображении. На рис. 4 приведен при-

мер результата алгоритма построчной сегментации (слева расположены графики средних значений пикселей по строкам пикселей; белый цвет имеет максимальное значение).

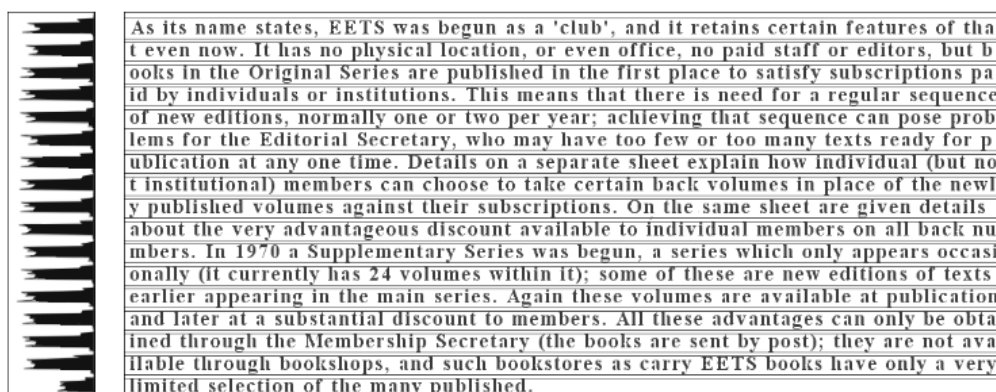


Рис. 4. Пример результата алгоритма построчной сегментации

При вертикальной (побуквенной) сегментации вычисляется среднее значение пикселей столбцов в каждом столбце участка строки текста, выделенной ранее. На рис. 5 приведен пример результата такой побуквенной сегментации (вверху указан график средних значений пикселей по столбцам пикселей).

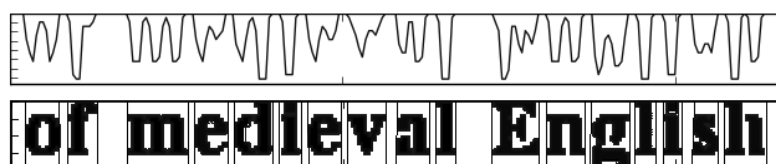


Рис. 5. Пример результата алгоритма побуквенной сегментации

В данном случае также осуществляется сравнение с порогом, которого в ряде случаев оказывается достаточно. Однако при артефактах изображения, не удаленных на предыдущих этапах предобработки, алгоритм может давать ложные срабатывания. В связи с этим, в работе вводится еще адаптивный порог разности между соседними темными пикселями в столбцах, благодаря которому отсеиваются случайные темные пиксели в выделенной строке, а сохраняются только относительно плотно прилегающие друг к другу символы.

**Распознавание текстовых символов.** На этапе распознавания текстовых символов традиционно применяются статистические модели и методы машинного обучения. Данные модели сначала обучаются на большом количестве изображений образцов символов, представляющих обучающую выборку, после чего могут быть использованы для решения задач классификации и регрессии, в частности, для распознавания текста. Также в последнее время наблюдается тенденция использования моделей машинного обучения и на этапе цифровой обработки сигнала. На текущем этапе работы детальный анализ моделей машинного обучения не производится. Тем не менее, проведены предварительные эксперименты по распознаванию текста на основе машин опорных векторов (Support Vector Machine, SVM) [10, 11].

Алгоритм обучения SVM находит среди элементов обучающей выборки векторы, лежащие на границе двух разделяемых подмножеств и строит между этими векторами гиперплоскость, максимально разделяющую входные образы в пространстве признаков. В терминах SVM это опорные векторы.

Решающая функция классификатора SVM задана формулой:

$$h(x) = \text{sgn}\left(\sum_{i=1}^m \lambda_i y_i K(\mathbf{x}_i, \mathbf{x}) - \omega_0\right),$$

где  $\lambda_i$  и  $\omega_0$  – коэффициенты;  $\mathbf{x}_i$  и  $y_i$  – входной вектор и соответствующее ему значение (0 или 1) из обучающей выборки, соответственно;  $K(\mathbf{x}_i, \mathbf{x}) = \varphi(\mathbf{x}_i, \mathbf{x})^T \varphi(\mathbf{x}_i, \mathbf{x})$  – ядро. Функция ядра служит для того, чтобы отображать входной вектор в пространство более высокой размерности, в котором, согласно теореме Ковера [11], вероятность разделимости образов повышается. В работе в качестве ядра применяются так называемые радиально-базисные функции:

$$\varphi(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2), \quad \gamma > 0 \quad (1)$$

Обучение SVM заключается в нахождении коэффициентов  $\lambda_i$  и  $\omega_0$ . Для этого решается задача квадратичной оптимизации с ограничениями:

$$\begin{aligned} \sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \rightarrow \max_{\lambda}, \\ \begin{cases} \sum_{i=1}^m \lambda_i y_i = 0, \\ 0 \leq \lambda_i \leq C; \quad C > 0, \quad i = 1, 2, \dots, m. \end{cases} \end{aligned} \quad (2)$$

Константы  $\gamma$  и  $C$  в формулах (1) и (2), соответственно, являются свободными параметрами модели и задаются пользователем. Наиболее часто параметр  $\gamma$  в случае радиально-базисных функций устанавливается в диапазоне  $[0.0001, 0.1]$ , параметр  $C$  – в диапазоне  $[1, 100]$ .

Отдельная SVM решает задачу бинарной классификации, т.е. способна разделить образы по принципу «свой-чужой». В случае текстовых символов число выходных образов соответствует числу букв в тестовом алфавите, поэтому в работе модель SVM расширяется на случай мультиклассовой классификации путем композиций автономных SVM, принимающих для каждого символа решение по типу «один против остальных».

Для решения задачи, рассматриваемой в статье, SVM обучаются и тестируются на буквах из алфавита английского языка. В тестовых сканах документов используются шрифты со сходным начертанием символов. При формировании вектора признаков для SVM двумерный образ сегментированного символа масштабируется к размеру 20x20 пикселей, центрируется и преобразовывается в одномерный вектор размерности 400 пикселей. Отметим, что это простейший способ представления символа. В дальнейшем планируется исследовать более эффективные и компактные способы описания символа на основе специальных дескрипторов изображения: гистограмм ориентированных градиентов (HOG), признаков SURF и др.

**Подготовка эксперимента и анализ результатов.** Информационная технология распознавания текста в зашумленных сканах документов основывается на методе, описанном в статье, и специально разработанном инструментальном комплексе TextRecognizer. Программный комплекс представляет собой пакет модулей на языке Python: denoising.py (очистка изображения от шумов), segmentation.py (сегментация), train.py