

# Распознавание и озвучивание текстов для облегчения обучения людей с нарушениями зрения

**Куравский Л. С.\***,

доктор технических наук, профессор,  
декан факультета информационных  
технологий, заведующий кафедрой  
прикладной информатики Московского  
городского психолого-педагогического  
университета

**Юрьев Г. А.\*\***,

студент кафедры прикладной информа-  
тики факультета информационных тех-  
нологий Московского городского пси-  
холого-педагогического университета

В статье представлена новая технология обработки текстов для людей с нарушениями зрения, интегрирующая средства сканирования, распознавания и озвучивания. Авторы отмечают, что, несмотря на наличие огромного количества электронных источников, пока существует ситуация, при которой людям с нарушениями зрения доступен только бумажный вариант текста. Именно поэтому разработка программно-аппаратных средств, обеспечивающих для незрячих людей доступ к этим источникам, является в настоящее время чрезвычайно актуальной. Материалы статьи показывают, что разработанная технология имеет значимые преимущества перед существующими в настоящее время средствами озвучивания текстов для людей с нарушениями зрения.

**Ключевые слова:** распознавание образов, нейронные сети, вейвлет-преобразования, сети Хэмминга, восстановление изображений.

## Основные компоненты технологии распознавания

Рассматриваемая технология распознавания и озвучивания текстов представлена на рис. 1. Она содержит три основных компонента:

– предварительную обработку изображения,

– распознавание символов и  
– озвучивание распознанного текста.  
Предварительная обработка изображения включает:  
– ввод изображения в одном из стандартных графических форматов,  
– преобразование изображения к монохромному представлению,

\*kuravsky@yahoo.com

\*\*grinch89@mail.ru

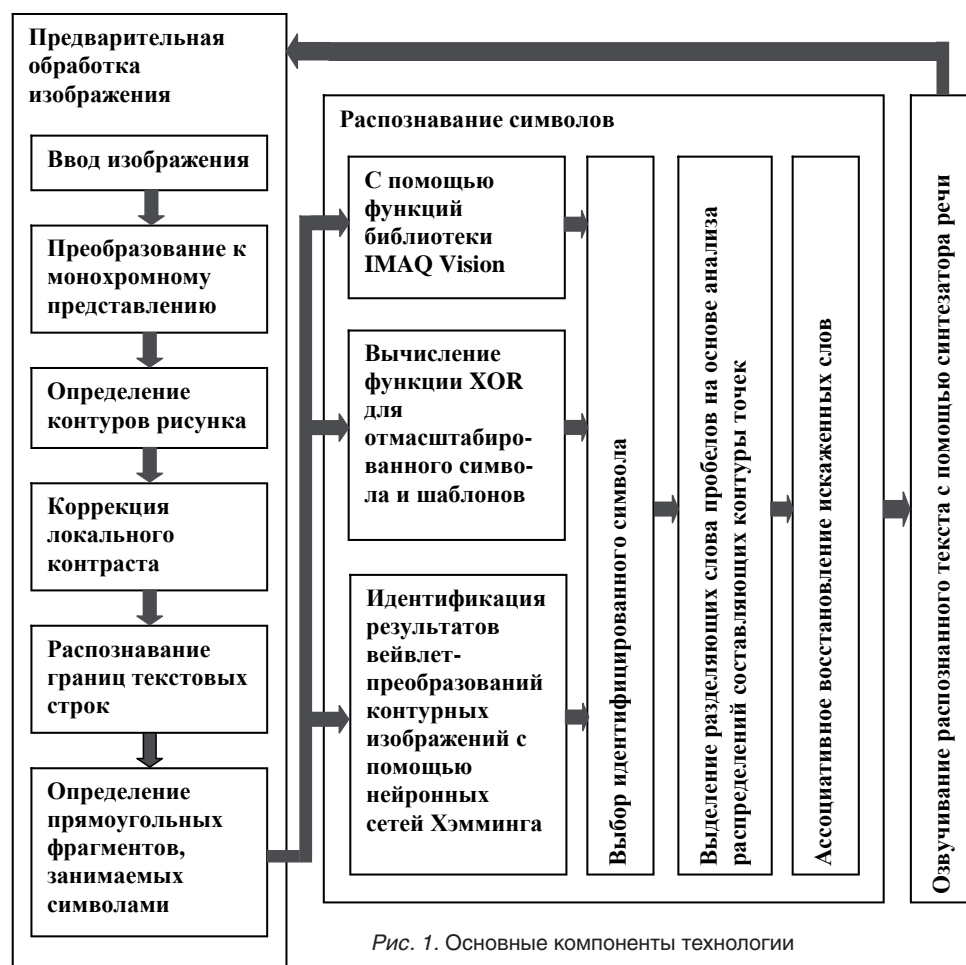


Рис. 1. Основные компоненты технологии

- определение контуров рисунка (преобразование к бинарному представлению),
- распознавание границ текстовых строк,
- определение прямоугольных фрагментов изображения, занимаемых символами в строке, включая составление списка их геометрических характеристик.

Для повышения надежности распознавание символов производится независимо тремя различными способами:

1) с помощью функций для обработки изображений из библиотеки IMAQ Vision, входящей в среду графического программирования LabVIEW;

2) путем вычисления функции «исключающего ИЛИ» для отмасштабированного символа и имеющихся шаблонов;

3) путем идентификации результатов вейвлет-преобразований контурных символьных изображений с помощью релаксационных нейронных сетей Хэмминга [2–4].

Важными подзадачами, решаемыми в процессе распознавания текстов, являются выделение разделяющих слова пробелов на основе анализа распределений составляющих контуры точек и ассоциативное восстановление слов, искаженных вследствие ошибок при идентификации символов.

**Особенности программной реализации и практического использования**

Программная реализация выполнена в среде графического программирования LabVIEW с использованием стандартных виртуальных инструментов для анализа данных, функций для обработки изображений из библиотеки IMAQ Vision, а также ряда процедур обработки изображений, выполненных в среде Borland Delphi и интегрированных в LabVIEW в форме динамически подключаемых библиотек.

Стандартный вариант применения рассматриваемой технологии предполагает сканирование текста с помощью веб-камеры, инициализация которой происходит в наилучшем из доступных режимов<sup>1</sup>. Использование такого инструмента в качестве универсального считывающего устройства обусловлено его доступностью, компактностью, простотой в управлении, поддержкой на большей части портативных компьютеров, а также однотипностью в управлении на программном уровне.

Введенное изображение преобразуется в монохромное черно-белое представле-

ние с последующим переводом в контурный рисунок с помощью непрореженного вейвлет-преобразования на базе биортонанального вейвлета (рис. 2).

Важной операцией, необходимой для корректной реконструкции связного текста, является выделение из изображения символьных строк. Задача усложняется тем, что камера, как правило, удерживается в руке и вследствие этого строки попадают в кадр повернутыми. С целью повышения надежности распознавания при достаточно больших углах поворота необходимо восстанавливать горизонтальное положение строк. Для этого на основе анализа плотностей черных и белых точек на нескольких вертикальных срезах изображения определяют местоположение начал и концов строк, что позволяет оценить угол, на который следует повернуть графическое представление для получения удобного для анализа состояния (рис. 3, 4). После такого поворота границы текстовых строк определяются повторно. На основе информации об этих границах фиксируются занимаемые символами прямоугольные фрагменты изображения (рис. 5).

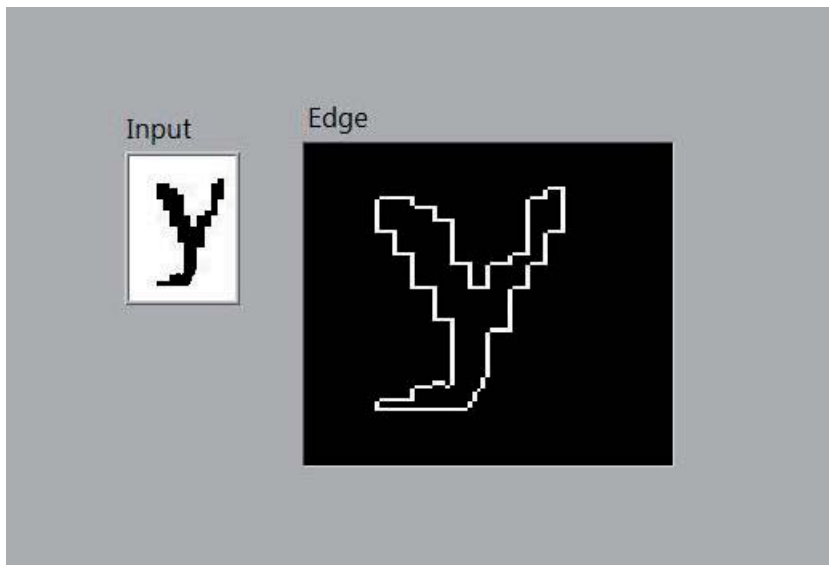


Рис. 2. Преобразование монохромного представления символа в контурный рисунок

<sup>1</sup> Как правило, распознавание проводилось с разрешением 320 x 240 и глубиной цвета 24 bit

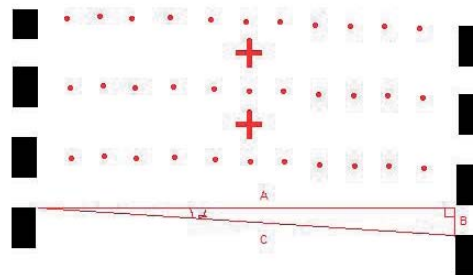


Рис. 3. Оценка расположения границ и угла поворота текстовых строк

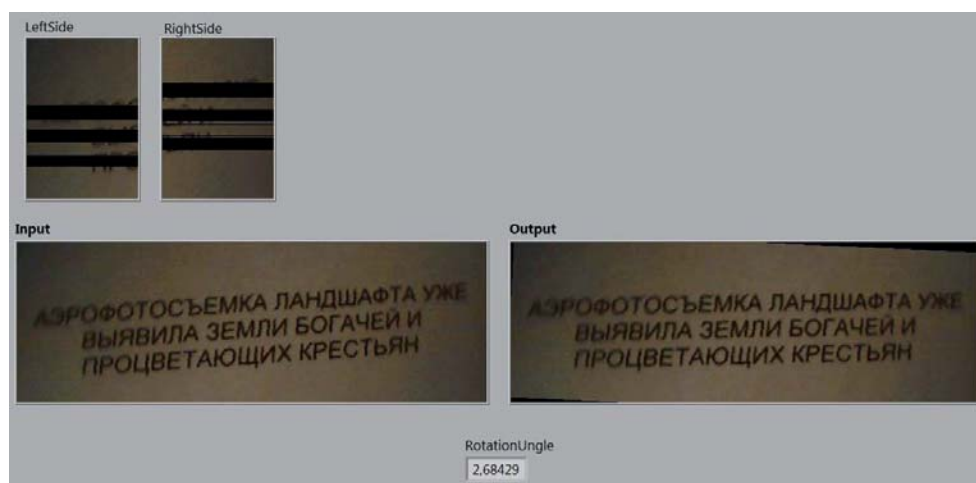


Рис. 4. Поворот текстовых строк

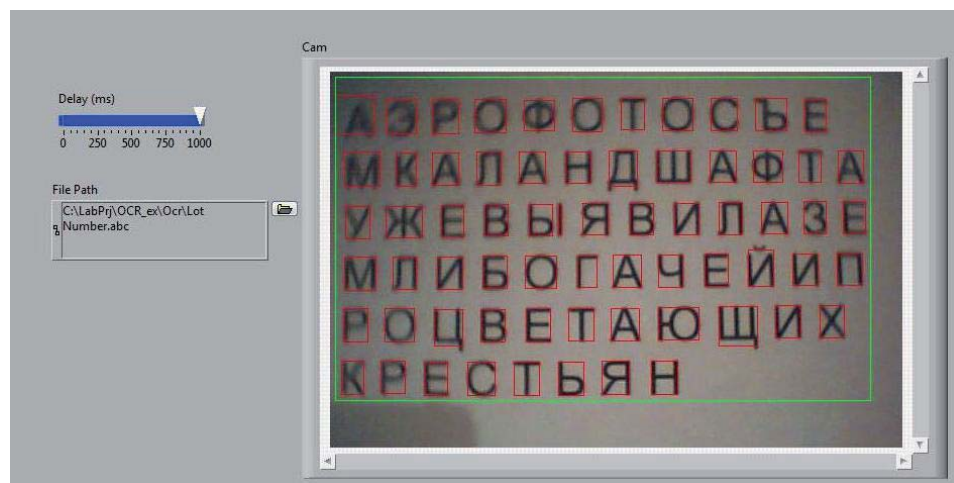


Рис. 5. Фиксация прямоугольных фрагментов изображения

По окончании выделения указанных прямоугольных фрагментов для всех символов в строке соответствующий ей участок вырезается из кадра, просматриваемого веб-камерой. При этом малые фрагменты, не превышающие заданный порог, удаляются из строки. Оставшиеся фрагменты рассматриваются как области, содержащие распознаваемые символы произвольного размера. Указанная процедура повторяется для всех строк из кадра, просматриваемого веб-камерой, причем информация о геометрических характеристиках как строк, так и прямоугольных фрагментов изображения, включающих в себя символы, сохраняется для последующего распознавания в порядке расположения фрагментов в кадре.

Алгоритм коррекции локального контраста ослабляет искажения в анализируемой части изображения. В выделенных прямоугольных фрагментах перебираются все черные точки, из числа которых удаляются слабосвязанные (имеющие мало одноцветных соседей). В то же время области, в которых предполагается исчезновение черных точек из-за нарушения освещенности или иных условий сканирования, заливаются черным цветом<sup>2</sup>. Указанный прием улучшает качество распознавания примерно на 10 %, практически не влияя на скорость работы программы.

Изображение, вводимое с веб-камеры, обычно затемнено с одного или нескольких краев. Этот эффект обусловлен непараллельностью поверхности текста и линзы, а также оптическими искажениями и недостаточным освещением. В случае незначительных искажений анализируется только центральная часть графического представления.

Один из реализованных способов распознавания символов опирается на возможности функций для обработки изображений, входящих в библиотеку IMAQ Vision среды графического программирования LabVIEW, позволяя, после предварительного обучения на образцах пользователя, проводить идентификацию содержимого

сформированной ранее последовательности прямоугольных фрагментов изображения. Результатом анализа является строка текста без пробелов с обозначенными нераспознанными позициями.

Для обеспечения надежности распознавания текста в достаточно неблагоприятных условиях, характерных для рассматриваемой задачи, потребовалась реализация двух дополнительных независимых способов распознавания. Первый из них прост и опирается на анализ результатов вычисления логической функции «*исключающего ИЛИ*» (XOR) для отмасштабированного символа и имеющихся буквенных шаблонов, с последующим выбором шаблона, имеющего наибольший процент совпадающих позиций. Второй способ реализует новый алгоритм [2; 4], использующий возможности вейвлет-преобразований и релаксационных нейронных сетей и способный эффективно работать после обучения на ограниченном числе образцов (рис. 6). Символ считается идентифицированным, если он выдается в качестве результата не менее чем двумя используемыми способами распознавания.

На первом этапе нейросетевого метода распознавания прямоугольные фрагменты с контурными изображениями преобразуются в векторы, описывающие распределения составляющих контуры точек вдоль осей абсцисс и ординат. Эти векторы далее подвергаются быстрому вейвлет-преобразованию, результаты которого подаются на вход сети Хэмминга с радиальными базисными элементами и экспоненциальными функциями активации [Там же]. Весовые коэффициенты радиальных элементов этих сетей вычисляются в соответствии с имеющимися эталонными образцами распознаваемых символов. После циклических вычислений нейронная сеть сходится к номеру ближайшего эталона. Таким образом, сходство с одним из заданных эталонных символов определяется нахождением в соответствующей области притяжения в пространстве допустимых представлений входного фраг-

<sup>2</sup>Примером может служить одна белая точка, все соседи которой в некоторой окрестности являются черными.

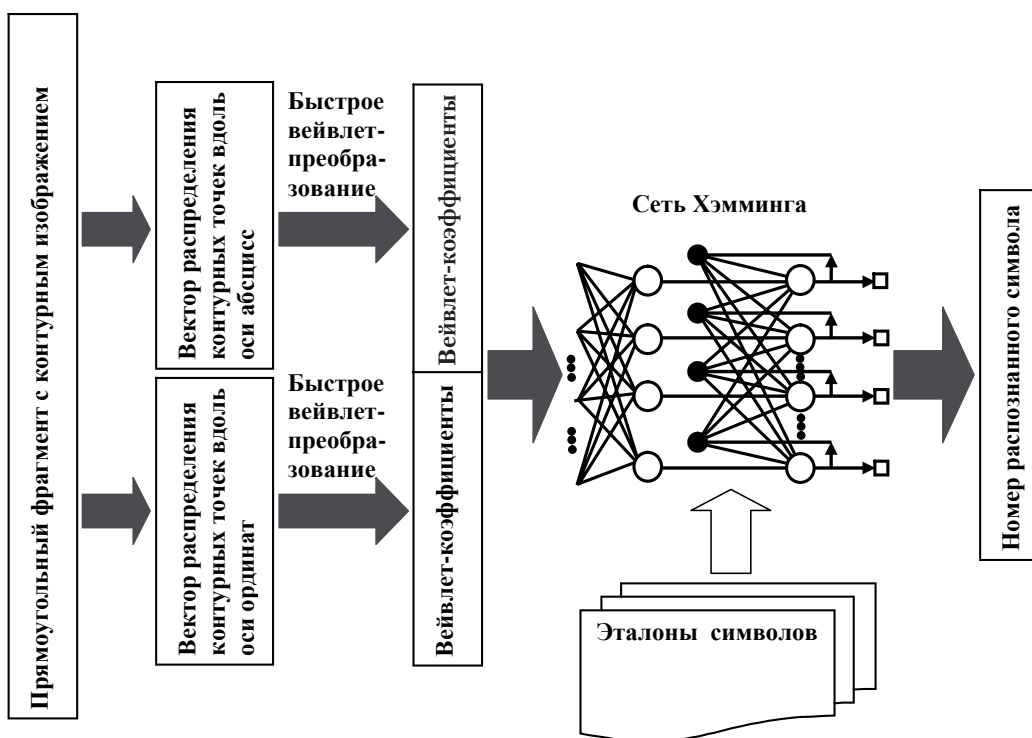


Рис. 6. Распознавание с использованием сетей Хэмминга с радиальными базисными элементами и экспоненциальными функциями активации

мента. Последовательность обработки данных при данном методе распознавания см. на рис. 4.

Если распознанная строка символов не будет содержать информацию о разделяющих слова пробелах, синтезатор речи не сможет корректно воспроизвести полученный текст. Для выявления пробелов производится анализ распределений составляющих контуры точек, аналогичный тому, что выполняется при нахождении строк. Данный анализ производится после нахождения прямоугольных фрагментов изображения, содержащих символы, поэтому можно легко определить позиции, после которых следует вставлять пробелы. Пробелы вставляются в текстовую строку перед заключительной коррекцией полученного текста, которая проводится на последнем

этапе распознавания. Эта коррекция выполняется путем ассоциативного восстановления слов, искаженных вследствие ошибок при идентификации символов либо содержащих нераспознанные элементы, с помощью встроенного словаря.

Озвучивание распознанного текста производится с помощью стандартного синтезатора речи. В частности, в среде операционной системы Windows для синтеза речи может быть использован интерфейс прикладного программирования Microsoft SAPI (Speech Application Programming Interface).

Очередной кадр изображения обрабатывается после озвучивания предыдущего. Темп обработки можно регулировать по желанию пользователя.

Текущая версия программной реализации эффективно распознает и озвучивает

произвольный плоскочечатный текст или текст на экране компьютерного монитора при достаточном освещении. Проведенное тестирование позволяет говорить об отсутствии русских букв, плохо идентифицируемых алгоритмом.

#### Основные результаты и выводы

Разработана и программно реализована технология обработки текстов, особенностями которой являются:

1) интеграция в одном программном продукте средств сканирования, распознавания и озвучивания;

2) наличие развитых средств устранения ошибок, обусловленных низким качеством сканированного изображения;

3) применение нового алгоритма распознавания символов, опирающегося на возможности вейвлет-преобразований и релаксационных нейронных сетей и способного эффективно работать после обучения на ограниченном числе образцов;

4) использование веб-камеры для ска-

нирования озвучиваемых изображений.

Разработанная технология играет важную роль в процессе адаптации [1] людей с нарушением зрения и имеет значимые преимущества перед существующими в настоящее время средствами озвучивания текстов. Эти преимущества представлены:

– мобильностью аппаратных средств;

– высокой скоростью и гибкостью воспроизведения информации в удобном для пользователя режиме;

– возможностью работы с текстами, представленными не в электронной форме;

– способностью работать с изображениями на экране компьютерного монитора.

Представленные здесь средства могут быть использованы:

– для чтения литературы, изданной традиционным плоскочечатным способом;

– работы за компьютером с обычным программным обеспечением, не предназначенным для незрячих пользователей.

#### Литература

1. Богомолов А. М. Личностный адаптационный потенциал в контексте системного анализа // Психологическая наука и образование. 2008. № 1.

2. Куравский Л. С., Баранов С. Н., Буланова О. Е., Кравчук Т. Е. Нейросетевая технология диагностики патологических состояний по аномалиям электроэнцефалограмм // Нейрокомпьютеры: разработка и применение. 2007. № 4.

3. Kuravsky L. S., Baranov S. N. Wavelet transforms and relaxation neural networks as promising technology components of technical and medical diagnostics and monitoring. In: Proc. 2nd World Congress on Engineering Asset Management and 4th International Conference on Condition Monitoring. Harrogate. United Kingdom. June 2007.

4. Kuravsky L. S., Baranov S. N. Technical diagnostics and monitoring based on capabilities of wavelet transforms and relaxation neural networks // Insight. March 2008. Vol. 50. № 3.



## Document recognition and vocalization technology for facilitation of learning process of people with vision impairments

**L. S. Kouravskiy,**

*PhD in engineering, Professor, Dean of information technologies faculty, head of applied informatics department, the Moscow State University of Psychology and Education*

**G. A. Yuriev,**

*student, applied informatics department, information technologies faculty, the Moscow State University of Psychology and Education*

Access to documented information is one of the biggest challenges for adaptation of people with vision impairments in the current computerized society. The main source of awareness in this so-called information space is just text matter. Despite of a big deal of e-sources, it is still widespread situation that texts are available in printed version only. One can by no means always obtain required information from books in Braille writing. Besides, usual printed texts and a part of browseable ones are unavailable for people with vision impairments, though those texts provide access to the vast majority of text data. Therefore, design of soft and hardware enabling blind people to get access to those sources is highly topical nowadays. The given piece of work presents cutting-edge technology of text proceeding for people with vision impairment, which integrates scanning, recognition and vocalization tools.

**Keywords:** image recognition, neural networks, wavelet-conversion, Hamming networks, image recovery.

### References

1. Bogomolov A.M. *Lichnostnyj adaptacionnyj potencial v kontekste sistemnogo analiza // Psihologicheskaja nauka i obrazovanie. 2008. № 1.*
2. Kuravskij L. S., Baranov S. N., Bulanova O. E., Kravchuk T. E. *Nejrosetevaja tehnologija diagnostiki patologicheskikh sostojanij po anomalijam jelektrojencefalogramm // Nejrokomp'jutery: razrabotka i primenenie. 2007. № 4.*
3. L. S. Kuravsky, S. N. Baranov. *Wavelet transforms and relaxation neural networks as promising technology components of technical and medical diagnostics and monitoring. In: Proc. 2-nd World Congress on Engineering Asset Management and 4th International Conference on Condition Monitoring. Harrogate. United Kingdom. June 2007.*
4. L. S. Kuravsky, Baranov S. N. *Technical diagnostics and monitoring based on capabilities of wavelet transforms and relaxation neural networks // Insight. March 2008. Vol. 50. № 3.*