РАСПОЗНАВАНИЕ И ОЗВУЧИВАНИЕ ТЕКСТОВ ДЛЯ ОБЕСПЕЧЕНИЯ УЧЕБНОГО ПРОЦЕССА ЛЮДЕЙ С НАРУШЕНИЯМИ ЗРЕНИЯ

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

Григорий Александрович Юрьев,
аспирант факультета информационных технологий Московского городского психолого-педагогического университета

• распознавание образов • нейронные сети • вейвлет-преобразования • сети Хэмминга • восстановление изображений • цепи Маркова •

Лёгкий путь доступа к информации стал неотъемлемой частью нашей жизни. Для большинства людей получение информации не представляет труда. Однако слабовидящие люди лишены возможности в полном объёме пользоваться книгами и прессой. Проблема доступа к текстовой информации — одна из наиболее значимых в процессе адаптации[1] людей с нарушениями зрения в современной компьютеризированной среде. И хотя сейчас появляется большое количество художественной литературы в аудиоформате, ни публицистика, ни бо́льшая часть технической литературы в этом формате не выходят. У издательств в этом необходимости нет — аудитория подобных проектов была бы слишком мала. Однако возможность для слабовидящих людей в индивидуальном порядке «прослушать» ту или иную газету, журнал, техническую инструкцию может быть весьма ценной.

Представленная технология обработки текстов для людей с нарушениями зрения, интегрирующая средства сканирования, распознавания и озвучивания может быть использована для чтения литературы, изданной традиционным печатным способом, и работы за экраном компьютера (здесь и далее подразумевается любой жидкокристаллический дисплей) с обычным программным обеспечением, не предназначенным для незрячих пользователей. Решение проблемы чтения именно плоскопечатного текста обеспечивает доступ к образовательным ресурсам, повышает эффективность процесса обучения и позволяет реализовать профессиональные навыки.

Распространённые традиционные средства доступа, включая азбуку Брайля, мультимедийные книги и программное обеспечение для озвучивания представленных в электронной форме текстов лишь частично решают указанную проблему. Эти средства требуют значительных усилий, затрачиваемых на предварительную подготовку исходного материала и перевода его в доступную для восприятия брайлевскую или электронную форму представления. Технология обеспечивает возможность считывания текста с бумажных и других носителей при минимальных требованиях к аппаратной поддержке, что повышает мобильность и доступность системы. Разработанное программное обеспечение включает развитые средства устранения ошибок, обусловленных низким качеством обрабатываемого изображения.

Из средств воспроизведения текстов, предназначенных для людей с нару-

[1] **Богомолов А.М.** Личностный адаптационный потенциал в контексте системного анализа // Психологическая наука и образование. 2008. № 1. С. 67–73.

шениями зрения, в настоящее время доступны:

— системы оптического распознавания текстов FineReader[2], CuneiForm[3] и другие, позволяющие озвучивать занесённые в память компьютера отсканированные тексты с бумажных источников;
— системы невизуального доступа JAWS[4], Adriane Knoppix[5] и их аналоги, озвучивающие представленные в электронной форме фрагменты текстов и имеющие развитую систему навигации по ключевым словам;
— звуковые книги, подготовленные в специальных аудиоформатах.

Однако:
— система FineReader, предназначенная для работы с полученной сканированием информацией, не обеспечивает нужный темп и гибкость работы с текстами, напечатанными на бумаге, и совершенно не приспособлена к озвучиванию текстов, отображаемых на экране компьютерного монитора и других недоступных сканированию поверхностях;
— система JAWS не работает с печатными текстами и текстами, представленными в графических форматах в виде рисунка;
— звуковые книги требуют предварительной аудиозаписи речи в студии и её последующей трудоёмкой разметки.

Учитывая указанные ограничения, технология озвучивания плоскопечатных и компьютерных текстов, работающая в режиме, задаваемом самим пользователем, поддерживает возможности, не реализованные в настоящее время ни в одной из существующих программных систем.

Из известных аппаратных решений, близких по тематике, на рынке присутствует IRISPen. Это устройство (маркер по форме) подключается к компьютеру через USB интерфейс. С одной стороны устройства расположена камера, предполагается, что пользователь будет вести маркер по строчке и в процессе движения отсканированный текст транслируется на выбранный язык встроенной системой распознавания и перевода. Компания позиционирует устройство как предмет для перевода цитат информации с визитных карточек, т.е. работа с большими массивами данных не предусматривается. Очевидно, для использования этой системы необходимо видеть строки, такой вариант неприемлем для людей с нарушениями зрения.

Аналогичным по функциональности является устройство Intel Reader, но оно обладает сравнительно высокой стоимостью, на данный момент не поддерживает русский язык и позиционируется, скорее, как разработка позволяющая реализовать чтение книжек вслух для людей без ограничений по зрению, что в свою очередь сказывается на управлении прибором.

ОСНОВНЫЕ КОМПОНЕНТЫ ТЕХНОЛОГИИ РАСПОЗНАВАНИЯ

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

- предварительную обработку изображения;
- распознавание символов;
- озвучивание распознанного текста.

Предварительная обработка изображения включает:
- ввод изображения в одном из стандартных графических форматов;
- преобразование изображения к монохромному представлению;
- 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 надёжности, распознавание символов производится независимо тремя различными способами:

- с помощью функций для обработки изображений из библиотеки IMAQ Vision, входящей в среду графического программирования LabVIEW[6] (этот встроенный набор

[2] FineReader (http://www.abbyy.ru/upload/files/FineReader_9.0_Reviewer%27s_Guide_Russian.pdf).

[3] CuneiForm (ftp://ftp.dol.ru/pub/users/cgntv/download/cuneiform/eng/cunei_e.pdf).

[4] JAWS(http://www.cardiff.ac.uk/accessibility/technicalinformation/guidetojaws/index.html).

[5] Adriane Knoppix (http://www.knopper.net/knoppix-adriane/index-en.html).

[6] LabVIEW tutorial for Windows. National Instruments Corp., 2004–2007.

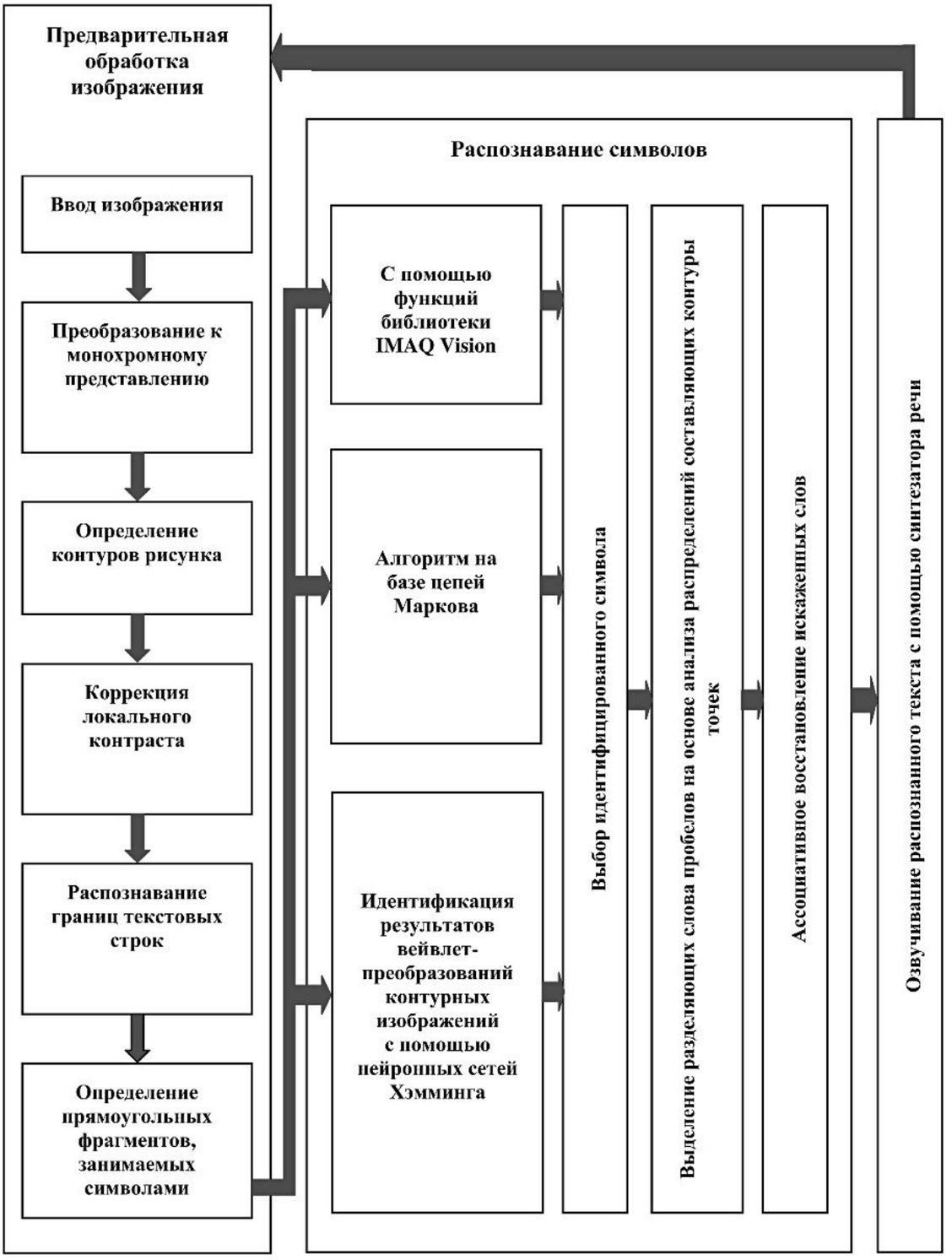

Рис. 1. Основные компоненты технологии распознавания и озвучивания плоскопечатных текстов

функций в основном ориентирован на распознавание символов, включённых в штрихкоды, что делает механизм недостаточно эффективным при использовании на изображениях с низким разрешением и высоким уровнем помех);

- с применением нового алгоритма использующего возможности Марковских цепей, нейронных сетей Хэмминга[7] (этот метод показал наивысшую эффективность при работе с сильно зашумленными изображениями, что связано, как с особенностью вэйвлет преобразований, так и с вероятностной природой релаксационных нейронных сетей).

Подобное голосование уменьшает вероятность ложного распознавания. Символы, которые будут признаны нераспознанными, позже восстановит специальный механизм, опирающийся на словарь, что является стандартным методом в подобном случае.

Важными подзадачами, решаемыми в процессе распознавания текстов, являются:

— выделение пробелов;
— ассоциативное восстановление слов.

Выделение разделяющих слова пробелов происходит на основе анализа распределений составляющих контуры точек. Восстанавливать искажённые слова приходится вследствие ошибок при идентификации символов.

ОСОБЕННОСТИ ПРОГРАММНОЙ РЕАЛИЗАЦИИ И ПРАКТИЧЕСКОГО ИСПОЛЬЗОВАНИЯ

Программная реализация выполнена в среде графического программирования LabVIEW. Использовались стандартные виртуальные инструменты для анализа данных, функции для обработки изображений из библиотеки IMAQ Vision, а также ряд 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[8]. Использование веб-камеры в качестве универсального считывающего устройства обусловлено её доступностью, компактностью, простотой и однотипностью в управлении на программном уровне. В нашей жизни всё большее распространение получают так называемые субноуты. Они обладают низким электропотреблением, которое обеспечивает большее время работы, и малыми габаритами, что делает их ещё более мобильными по сравнению с другими портативными компьютерами. Очевидно, что их технические характеристики полностью удовлетворяют требованиям рассматриваемой системы, что делает её очень мобильной и доступной простому пользователю.

Введённое изображение преобразуется в монохромное чёрно-белое представление. Именно в таком виде наиболее удобны любые операции распознавания и анализа изображений (рис. **2**).

Важной операцией, необходимой для корректной реконструкции связанного текста, является выделение из изображения символьных строк. Задача усложняется тем, что камера, как правило, удерживается в руке и, вследствие этого, строки попадают в кадр искажёнными. Для повышения надёжности распознавания при достаточно больших углах поворота необходимо восстанавливать горизонтальное положение строк (рис. **3**). Поскольку для всех вычислений используются средние значения, метод не позволяет сделать строки абсолютно параллельными границам изображения, но для работы остальных алгоритмов результат вполне приемлем.

Ещё одним минусом подобной технологии является то, что в кадр должно обязательно попасть минимум две строки, если это условие не выполняется, изображение ос-

[7] **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. С. 4–14; **Kuravsky L.S., Baranov S.N.** Wavelet transforms and relaxation neural networks as promising technology components of technical and medical diagnostics and monitoring // Proc. 2nd World Congress on Engineering Asset Management and 4th International Conference on Condition Monitoring, Harrogate, United Kingdom, June 2007, pp. 1117–1132; **Kuravsky L.S., Baranov S.N.** Technical diagnostics and monitoring based on capabilities of wavelet transforms and relaxation neural networks // Insight, Vol. 50, No 3, March 2008, pp. 127–132.

[8] Как правило, распознавание проводилось с разрешением 320 x 240 и глубиной цвета 24bit.

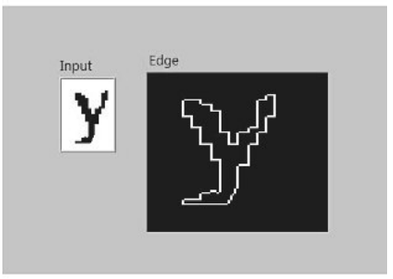

Рис. 2. Преобразование монохромного представления символа в контурный рисунок

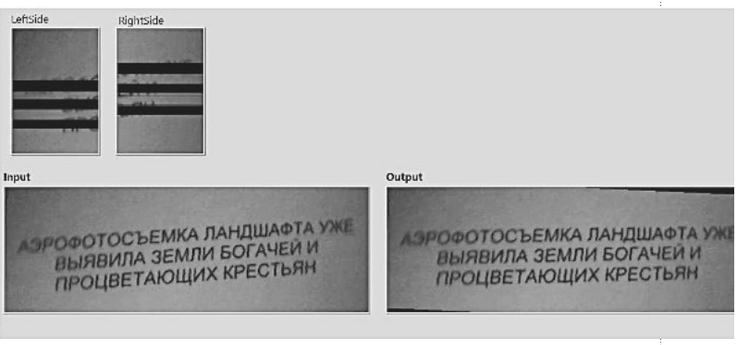

Рис. 3. Поворот тестовых строк

таётся неизменным. Стоит отметить, что, если строка обрывается на середине, это не влияет на вычисление в целом, в силу того, что срезы делаются более чем на двух участках изображения. После такого поворота границы текстовых строк определяются повторно. На основе информации об этих границах фиксируются занимаемые символами прямоугольные фрагменты изображения (рис. **4**).

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.

Ослаблению искажения в анализируемой части изображения способствует алгоритм коррекции локального контраста. Этот прием позволяет уменьшить эффект, возникающий в связи с неравномерным освещением. При переводе в чёрно-белый формат невозможно установить общий уровень контраста для всего изображения, так как либо часть символов распадётся, и распознать их будет невозможно, либо некоторые окажутся закрашенными чёрным целиком. Суть метода в следующем: на выделенных прямоугольных фрагментах перебираются все чёрные точки, из числа которых удаляются слабосвязанные (имеющие малое число одноцветных соседей). В то же время области, в которых предполагается исчезновение чёрных точек из-за нарушения освещенности или иных условий сканирования, заливаются чёрным цветом[9].

Рис. 4. Фиксация прямоугольных фрагментов изображения, занимаемых символами

Изображение, вводимое с веб-камеры, обычно затемнено с одного или нескольких краёв, и, несмотря на все использованные алгоритмы, с таким уровнем затемнения бороться бессмысленно. Эффект обусловлен непараллельностью поверхности текста и линзы, а также оптическими искажениями и недостаточным освещением. В случае значительных 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. Результатом анализа является строка текста без пробелов, с обозначенными нераспознанными позициями.

Для обеспечения надёжности распознавания текста в особенно неблагоприятных условиях, характерных для рассматриваемой задачи, потребовалась реализа-

[9] Примером может служить одна белая точка, все соседи которой в некоторой окрестности являются чёрными.

ция двух дополнительных независимых способов распознавания.

Первый опирается на возможности Марковских цепей. Каждый прямоугольник с символом переводится в контурный рисунок, прямоугольные фрагменты с контурными изображениями последовательно преобразуются в числовые векторы, которые подвергаются быстрому вейвлет преобразованию. Полученные вейвлет-коэффициенты рассматриваются как представления символов и анализируются с помощью распознающей цепи Маркова, структура которой представлена на рис. 5. Процедура распознавания основана на анализе знаков вейвлет-коэффициентов из указанной последовательности.

Распознавание поступающих на вход системы новых символов возможно после завершения обучения. Тип символа, который соответствует цепи Маркова, выдаётся как результат распознавания.

Оценка вероятности корректного распознавания символов с помощью описанного алгоритма, полученная по результатам 1000 проб на базе асимптотической аппроксимации биномиального распределения, равна 0,791, причём нижняя и верхняя границы 95% доверительного интервала есть 0,76 и 0,83.

Оказалось, что достаточно использовать последовательности **W**, состоящие из двенадцати вейвлет-коэффициентов (***N***=12), начиная с третьего коэффициента разложения распределений контурных точек. Расчёты выявили неинформативность первых двух вейвлет-коэффициентов этого разложения для решения задачи распознавания.

Второй способ реализует новый алгоритм, использующий возможности вейвлет-преобразований и релаксационных нейронных сетей. Он способен эффективно работать после обучения на ограниченном числе образцов (рис. 6).

В этом методе используется сеть Хэмминга. Суть данной структуры заключается в поиске хэммингова расстояния (расстоянием Хэмминга называется число отличающихся битов в двух бинарных векторах) от представленного образца до образца из обучающей выборки. Образец, до которого такое расстояние окажется наименьшим, признаётся искомым. Помимо этого, существует ряд особенностей, которые позволяют сделать пространство поиска более «контрастным» и, как следствие, улучшить распознавание в сложных случаях.

На первом этапе нейросетевого метода распознавания прямоугольные фрагменты с контурными изображениями последовательно преобразуются в числовой вектор. Полученный вектор подвергается быстрому вейвлет-преобразованию, результаты которого подаются на вход сети Хэмминга с радиальными базисными элементами и экспоненциальными функциями активации. После циклических вычислений нейронная сеть сходится к номеру ближайшего эталона. Последовательность обработки данных, в этом методе распознавания представлена на рис. 6.

Символ считается идентифицированным, если он выдаётся в качестве результата не менее чем двумя используемыми способами.

Если распознанная строка символов не будет содержать информацию о разделя-

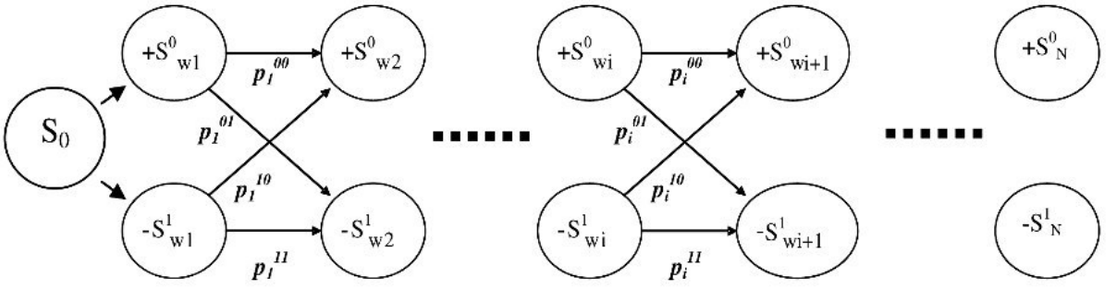

Рис. *5. Структура цепи Маркова*

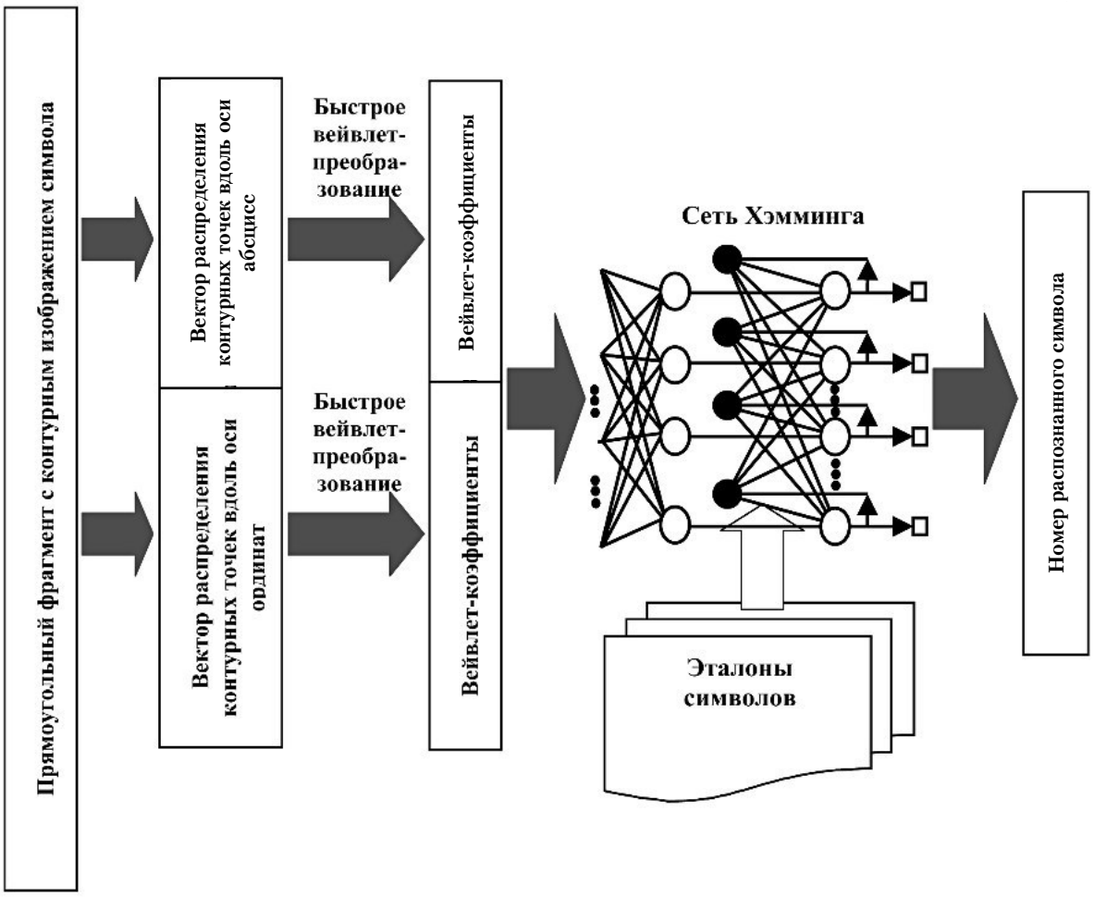

Рис. 6. Распознавание с использованием сетей Хэмминга с радиальными базисными элементами и экспоненциальными функциями активации

ющих слова пробелах, то синтезатор речи не сможет корректно воспроизвести полученный текст. Выявление пробелов в тексте необходимо для обеспечения работы синтезатора речи. Поиск пробелов происходит на основании частотного анализа. Расстояния между буквами обычно меньше, чем между словами, в полтора-два раза. Этот факт позволяет говорить о весьма устойчивом разделении слов текста.

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 путём ассоциативного восстановления слов (искажённых вследствие ошибок при идентификации символов, либо содержащих нераспознанные элементы) с помощью встроенного словаря.

Для того чтобы текст был озвучен, необходимо подготовить строку, содержащую последовательно все символы с пробелами. Эта строка передается заранее настроенному синтезатору речи. Так же программа осуществляет контроль за тем, завершился процесс воспроизведения или нет. Если процесс не завершился, новая порция текста не передается, так как передача вызовет заполнение буфера и воспроизведение длительной последовательности звуков уже после завершения сеанса работы с программой. Для озвучивания распознанного текста используется стандартный синтезатор речи. 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. Это вполне разумно, так как для восприятия различных по уровню информационной насыщенности текстов требуется, как правило, разное время. Не следует забывать и об индивидуальных различиях в скорости восприятия информации. Так же установка заниженного темпа может быть весьма полезна на начальных этапах работы с системой, пока пользователь не привык к новому способу восприятия информации, по сути требующим координированной работы рук, слуха и одновременной интерпретации услышанного.

Выводы и результаты

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

- интеграция в одном программном продукте средств сканирования, распознавания и озвучивания;
- наличие развитых средств устранения ошибок, обусловленных низким качеством сканированного изображения;
- применение нового алгоритма распознавания символов, опирающегося на возможности вейвлет-преобразований и релаксационных нейронных сетей и способного эффективно работать после обучения на ограниченном числе образцов;
- разработка и реализация нового алгоритма распознавания символов на базе Цепей Маркова, определение формальных оценок его эффективности;
- использование веб-камеры для сканирования озвучиваемых изображений.

Разработанная технология имеет значимые преимущества перед существующими в настоящее время средствами озвучивания текстов для людей с нарушениями зрения, связанные с:

— 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 средства могут быть использованы:

— для чтения литературы, изданной традиционным плоскопечатным способом;
— для работы за компьютером с обычным программным обеспечением, не предназначенным для незрячих пользователей. ❑