

МАШИННОЕ ЗРЕНИЕ: ОТ ТЕОРИИ РАСПОЗНАВАНИЯ К ПРАКТИЧЕСКОЙ РЕАЛИЗАЦИИ

Компания "Инженерные Системы"

Введено понятие целостности восприятия. Рассмотрены три подхода к распознаванию символов – шаблонный, структурный и признаковый, их преимущества и недостатки. Предложен новый подход к распознаванию символов, основанный на применении структурно-пятенного эталона. Приведен пример практической реализации системы распознавания в области сортировки зерновых культур.

Ключевые слова: целостность восприятия, распознавание символов, структурно-пятенный эталон.

О существовании специальных систем, которые "автоматически вводят в компьютер текст", знают даже начинающие пользователи. Со стороны все выглядит довольно просто и логично. На отсканированном изображении система находит фрагменты, в которых "узнает" буквы, а затем заменяет эти изображения настоящими буквами, или, по-другому, их машинными кодами. Так осуществляется переход от изображения текста к "настоящему" тексту, с которым можно работать в текстовом редакторе. Как этого добиться?

Целостность восприятия

В основе фонетического преобразования лежит принцип целостности. В соответствии с ним любой воспринимаемый объект рассматривается как целое, состоящее из частей, связанных между собой определенными отношениями. Так, например, печатная страница состоит из статей, статья – из заголовка и колонок, колонка – из абзацев и т.д. При этом все перечисленные элементы текста связаны между собой определенными пространствами и языковыми отношениями.

Для выделения целого требуется определить его части. Части же, в свою очередь, можно рассматривать только в составе целого. Поэтому целостный процесс восприятия может происходить только в рамках гипотезы о воспринимаемом объекте – целом. После того как выдвинуто предположение о воспринимаемом объекте, выделяются и интерпретируются его части. Затем предпринимается попытка "собрать" из них целое, чтобы проверить правильность исходной гипотезы. Разумеется, воспринимаемый объект может интерпретироваться в рамках более крупного целого. Так, читая предложение, человек узнает буквы, воспринимает слова, связывает их в синтаксические конструкции и понимает смысл.

В технических системах любое решение при распознавании текста принимается неоднозначно, а путем последовательного выдвижения и проверки гипотез и привлечения как знаний о самом исследуемом объекте, так и общего контекста. Целостное описание класса объектов восприятия отвечает двум условиям: все объекты данного класса удовлетворяют этому описанию; ни один объект другого класса не удовлетворяют описанию. Например, класс изображений буквы "К" должен быть описан так, чтобы любое изображение буквы "К" в него попадало, а изображение всех других букв – нет. Такое описание обла-

дает свойством отображаемости, то есть обеспечивает воспроизведение описываемых объектов: эталон буквы для системы OCR позволяет визуально воспроизвести букву, эталон слова для распознавания речи позволяет произнести слово, а описание структуры предложения в синтаксическом анализаторе позволяет синтезировать правильное предложение. С практической точки зрения отображаемость играет огромную роль, поскольку позволяет эффективно контролировать качество описаний.

Существует два вида целостного описания: шаблонное и структурное. В первом случае описание представляет собой изображение в растровом или векторном представлении, и задан класс преобразований (например, повтор, масштабирование и пр.). Во втором случае описание представляется в виде графа, узлами которого являются составляющие элементы входного объекта, а дугами – пространственные отношения между ними. В свою очередь элементы могут оказаться сложными (то есть иметь свое описание).

Конечно, шаблонное описание проще в реализации, чем структурное. Однако оно не может использоваться для описания объектов с высокой степенью изменчивости. Шаблонное описание, например, может приниматься для распознавания только печатных символов, а структурное – еще и для рукописных.

Целостность восприятия предлагает два важных архитектурных решения. Во-первых, все источники знания должны работать по возможности одновременно. Нельзя, например, сначала распознать страницу, а затем подвергнуть ее словарной и контекстной обработке, поскольку в этом случае невозможно будет осуществить обратную связь от контекстной обработки к распознаванию. Во-вторых, исследуемый объект должен представляться и обрабатываться по возможности целиком.

Первый шаг восприятия – это формирование гипотезы о воспринимаемом объекте. Гипотеза может формироваться как на основе априорной модели объекта, контекста и результатов проверки предыдущих гипотез (процесс "сверху – вниз"), так и на основе предварительного анализа объекта ("снизу – вверх"). Второй шаг – уточнение восприятия (проверка гипотезы), при котором производится дополнительный анализ объекта в рамках выдвинутой гипотезы и в полную силу привлекается контекст.

Для удобства восприятия необходимо провести предварительную обработку объекта, не потеряв при

этом существенной информации о нем. Обычно предварительная обработка сводится к преобразованию входного объекта в представление, удобное для дальнейшей работы (например, векторизация изображения), или получение всевозможных вариантов сегментации входного объекта, из которого путем выдвижения и проверки гипотез выбирается правильный. Процесс выдвижения и проверки гипотез должен быть явно отражен в архитектуре программы. Каждая гипотеза должна быть объектом, который можно было бы оценить или сравнить с другими. Поэтому обычно гипотезы выдвигаются последовательно, а затем объединяются в список и сортируются на основе предварительной оценки. Для окончательного же выбора гипотезы активно используется контекст и другие дополнительные источники знаний.

Ныне одним из лидеров в области генетического программирования является группа исследователей из Стэнфордского университета, работающая под руководством профессора Д. Коза. Генетическое программирование вдохнуло новую жизнь в хорошо уже подзабытый язык LISP (List Processing), который создавался группой Д. Маккарти (того самого, кто в 60-е годы XX века ввел в обиход термин "искусственный интеллект") как раз для обработки списков и функционального программирования. Кстати, именно этот язык в США был и остается одним из наиболее распространенных языков программирования для задач искусственного интеллекта.

Распознавание символов

Сегодня известно три подхода к распознаванию символов — шаблонный, структурный и признаковый. Но принципу целостности отвечают лишь первые два. Шаблонное описание проще в реализации, однако в отличие от структурного оно не позволяет описывать сложные объекты с большим разнообразием форм. Именно поэтому шаблонное описание применяется для распознавания лишь печатных символов, в то время как структурное — для рукописных, имеющих, естественно, гораздо больше вариантов начертания.

Шаблонные системы преобразуют изображение отдельного символа в растровое, сравнивают его со всеми шаблонами, имеющимися в базе и выбирают шаблон с наименьшим числом точек, отличных от входного изображения. Шаблонные системы довольно устойчивы к дефектам изображения и имеют высокую скорость обработки входных данных, но надежно распознают только те шрифты, шаблоны которых им "известны". И если распознаваемый шрифт хоть немного отличается от эталонного, шаблонные системы могут делать ошибки даже при обработке очень качественных изображений.

Структурные системы описывают объект как граф, узлами которого являются элементы входного объекта, а дугами — пространственные отношения между ними. Системы, реализующие подобный подход, обычно работают с векторными изображениями.

Структурными элементами являются составляющие символ линии. Так, для буквы "р" — это вертикальный отрезок и дуга. К недостаткам структурных систем следует отнести их высокую чувствительность к дефектам изображения, нарушающим составляющие элементы. Также векторизация может добавить дополнительные дефекты. Кроме того, для этих систем до сих пор не созданы эффективные автоматизированные процедуры обучения. Поэтому в большинстве продуктов структурные описания приходится создавать вручную.

Признаковые системы представляют усредненное изображение каждого символа как объект в n -мерном пространстве признаков. Здесь выбирается алфавит признаков, значения которых вычисляются при распознавании входного изображения. Полученный n -мерный вектор сравнивается с эталонными, и изображение относится к классу, с эталоном которого выявлено наибольшее совпадение. Признаковые системы не отвечают принципу целостности. Необходимое, но недостаточное условие целостности описания класса объектов (в нашем случае это класс изображений, представляющих один символ) состоит в том, что описанию должны удовлетворять все объекты данного класса и ни один из объектов других классов. Но поскольку при вычислении признаков теряется существенная часть информации, трудно гарантировать, что к данному классу удастся отнести только "родные" объекты.

Структурно-пятенный эталон

Фонтанное преобразование совмещает достоинства шаблонной и структурной систем и позволяет избежать недостатков, присущих каждой из них по отдельности. В основе этой технологии лежит использование структурно-пятенного эталона. Он позволяет представить изображения в виде набора пятен, связанных между собой n -арными отношениями, задающими структуру символа. Эти отношения (то есть расположение пятен друг относительно друга) образуют структурные элементы, составляющие символ. Так, например, отрезок — это один тип n -арных отношений между пятнами, эллипс — другой, дуга — третий. Другие отношения задают пространственное расположение образующих символ элементов.

В эталоне задаются: имя; обязательные, запрещающие и необязательные структурные элементы; отношения между структурными элементами; отношения, связывающие структурные элементы с описываемым прямоугольником символа; атрибуты, используемые для выделения структурных элементов; атрибуты, используемые для проверки отношений между элементами; атрибуты, используемые для оценки качества элементов и отношений; позиция, с которой начинается выделение элемента (отношения локализации элементов).

Структурные элементы, выделяемые для класса изображений, могут быть исходными и составными. Исходные структурные элементы — это пятна, со-

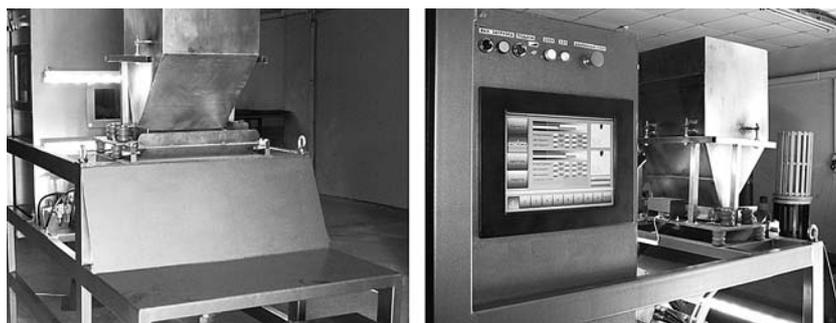


Рис. 1. Visual Sorter 2.0. grain для сортировки зерновых культур

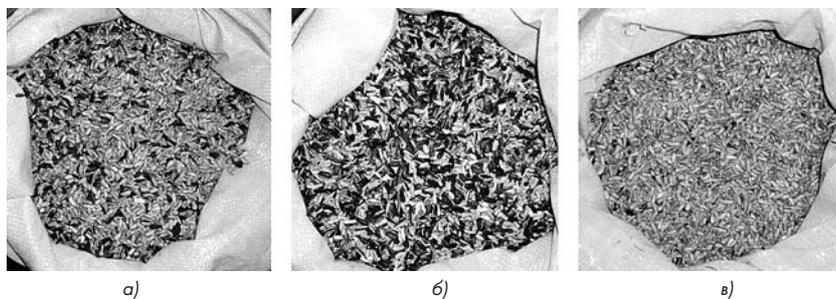


Рис. 2. Результат работы Visual Sorter 2.0.grain на ядре подсолнечника, где а) исходное сырье, б) отходы сортировки, в) результат обработки

ставные — отрезок, дуга, кольцо, точка. В качестве составных структурных элементов могут быть взяты любые объекты, описанные в эталоне. Кроме того, они могут быть описаны как через исходные, так и через другие составные структурные элементы. Например, для распознавания корейских иероглифов (слоговое письмо) составными элементами для описания слога являются описания отдельных букв (но не отдельные элементы букв). В итоге, использование составных структурных элементов позволяет строить иерархические описания классов распознаваемых объектов.

В качестве отношений используются связи между структурными элементами, которые определяются либо метрическими характеристиками этих элементов (например, "длина больше"), либо их взаимным расположением на изображении (например, "правее", "соприкасается").

При задании структурных элементов и отношений используются конкретизирующие параметры, позволяющие доопределить структурный элемент или отношение при использовании этого элемента в эталоне конкретного класса. Для структурных элементов конкретизирующими могут являться, например, параметры, задающие диапазон допустимой ориентации отрезка, а для отношений — параметры, задающие предельное допустимое расстояние между характерными точками структурных элементов в отношении "соприкасается".

Конкретизирующие параметры используются также для вычисления "качества" конкретного структурного элемента изображения и "качества" выполнения данного отношения.

Построение и тестирование структурно-пятенных эталонов для классов распознаваемых объектов — процесс сложный и трудоемкий. База изображений, которая используется для отладки описаний, должна содержать примеры хороших и плохих (предельно допустимых) изображений для каждой графемы, а изображения базы разделяются на обучающее и контрольное множества.

Разработчик описания предварительно задает набор структурных элементов (разбиение на пятна) и отношения между ними. Система обучения по базе изображений автоматически вычисляет параметры элементов и отношений. Полученный эталон проверяется и корректируется по контрольной выборке изображений данной графемы. По контрольной же выборке проверяется результат распознавания, то есть оценивается качество подтверждения гипотез.

Распознавание с использованием структурно-пятенного эталона происходит следующим образом. Эталон накладывается на изображение, и отношения между выделенными на изображении пятнами сравниваются с отношениями пятен в эталоне. Если выделенные на изображении пятна и отношения между ними удовлетворяют эталону некоторого символа, то данный символ добавляется в список гипотез о результате распознавания входного изображения.

Практическая реализация

Описанный подход реализован в программно-техническом комплексе Visual Sorter Grain (рис. 1), предназначенном для оптической сортировки зерновых культур, например, для сортировки кофе, орехов, семян подсолнечника. Система способна в реальном времени определять параметры каждой гранулы, ее цвет, размер и посторонние цветочные включения. На основе этой информации производится разделение продукции по заранее описанным правилам при помощи пневматического исполнительного механизма. На рис. 2 представлен результат работы системы на ядре подсолнечника.

В настоящее время, максимальная производительность сортировочного комплекса составляет около 1 т/ч. Система масштабируема, что позволяет создавать комплексы с большей общей производительностью. Особенности автоматической линии сортировки зерновых культур в том, что она при сравнительно компактных размерах и низкой стоимости позволяет достичь высоких результатов.

Контактный телефон (4942) 37-30-80.
<http://www.es-prof.com>