

Распознавание рукописного текста по восстановленному следу пера с помощью медиального представления

С.П. Арсеев¹, Л.М. Местецкий¹

¹Московский государственный университет имени М.В. Ломоносова, Ленинские Горы 1, Москва, Россия, 119991

Аннотация. Статья посвящена сведению задачи offline-распознавания рукописного текста (по изображению текста) к задаче online-распознавания (по следу пера). Предложен метод восстановления следа пера из изображения рукописного текста с использованием медиального представления (скелета). Проведено экспериментальное сравнение качества распознавания символов по изображению, по истинному следу пера и по восстановленному следу пера.

1. Введение

Распознавание рукописного текста является активно исследуемой задачей компьютерного зрения, находящей своё применение во многих практических областях, например, в автоматической обработке документов или в работе с архивами. Существует два основных подхода к решению этой задачи: offline- и online-распознавание.

Offline-распознавание представляет собой задачу распознавания написанного текста по его изображению. Эта задача чаще всего встречается на практике, и способы её решения разнятся от использования классических методов компьютерного зрения с выделением признаков и их классификацией до использования свёрточных нейронных сетей, показывающих сейчас наилучшие результаты в большинстве областей компьютерного зрения.

Online-распознавание представляет собой задачу распознавания текста по записанному в процессе его написания следу пера. В настоящее время в решении этой задачи хорошо себя показывают рекуррентные нейронные сети. Такая задача часто является более простой в решении, чем задача offline-распознавания, но основной проблемой такого подхода является необходимость использования специального записывающего оборудования при написании текста, что существенно ограничивает применение этого подхода на практике. Во многих задачах, в частности, в оцифровке архивов, запись следа пера отсутствует в принципе, и для распознавания доступен только сам текст.

Цель данной работы – попытаться свести задачу offline-распознавания рукописных символов к задаче online-распознавания посредством реконструкции возможного следа пера по изображению. Такое сведение позволит решать широкий спектр задач распознавания рукописного текста методами online-распознавания, которые часто оказываются более эффективными и не столь требовательны к объёму обучающей выборки.

2. Восстановление следа пера

Основной идеей предложенного подхода является построение модели символа в виде геометрического графа с вершинами в точках, близких к положению центра пера при

начертании символа, и рёбрами, соответствующими отрезкам траектории движения пера. После построения такой модели восстановленный след пера будет являться маршрутом в данном графе.

2.1. Построение скелета

Первым этапом для восстановления следа пера по изображению текста или символа является построение его модели в виде графа. Для этого используется медиальное представление или скелет символа на этом изображении. Для начала введём несколько базовых понятий.

Скелетом называется множество центров всех вписанных пустых кругов фигуры.

Стрижкой (регуляризацией) скелета называется процесс отсечения ветвей, вносящих незначительный вклад в образование формы. Способ стрижки, который будет использоваться далее, описан в [1]. Если не проводить стрижку, в получившемся скелете будет присутствовать множество коротких лишних ветвей, зашумляющих модель и затрудняющих её обработку.

Скелет многоугольной фигуры можно рассматривать как геометрический граф, рёбрами которого являются отрезки прямых линий и парабол. К такому представлению ведёт получение скелета из диаграммы Вороного [1]. В рамках данной задачи мы будем рассматривать скелет в виде $S = (V; E)$, где каждая вершина $v \in V$ представляется в виде пары координат (x, y) , а каждое ребро $e \in E$ является неупорядоченной парой вершин. Ввиду малой длины рёбер скелета символа отличия при приближении параболических рёбер прямыми линиями незначительны.

Алгоритм построения этой модели подробно описан в [1]. Сначала изображение бинаризуется, после этого связные области на бинарном изображении аппроксимируются многоугольниками а затем на этих многоугольниках строится морфологическая модель согласно алгоритму из [1]. Пример построения скелета показан на рисунке 1.



Рисунок 1. Регуляризованный скелет символа (буква “g”).

Основой предлагаемого подхода является предположение, что скелет изображения символа можно рассматривать как модель следа пера при начертании этого символа. В большинстве случаев точки, входящие в скелет, близки к центру штриха символа, и существенные отклонения возможны только в случае нечётких краёв штриха или значительного перекрытия близко расположенных штрихов. Соответственно, задача восстановления возможного следа пера (с некоторой точностью) сводится к задаче построения обхода скелета; под обходом графа понимается маршрут (последовательность инцидентных вершин и рёбер с возможными повторами) в графе, в который должны входить все рёбра данного графа.

В случае, если символ состоит из нескольких связных компонент, маршрут строится для каждой из них. Предложенный алгоритм не обрабатывает случай, когда символ состоит из нескольких пересекающихся штрихов, и при построении маршрута реальная траектория приближается возможной траекторией начертания символа «без отрыва пера».

2.2. Построение метаграфа

Если рёбра регуляризованного скелетного графа считать фрагментами траектории пера, то прохождение этих фрагментов определяется лишь направлением движения пера. Для последовательности смежных вершин степени 2 (т.е. расположенных на штрихе) траектория однозначно определяется выбором направления движения. Неоднозначность построения маршрута возможна только в вершинах степени 3 и выше (т.е. на развилках), и отдельным случаем неоднозначности является построение обхода цикла, т.к. необходимо так выбирать последовательность обхода, чтобы были обойдены все рёбра цикла. Для того, чтобы облегчить задачу построения такого обхода графа, мы строим на его основе новый вспомогательный граф, в дальнейшем называемый метаграфом исходного графа, таким образом, что каждая область исходного графа, где возможны неоднозначности, представляется в виде одной вершины метаграфа, что упрощает построение обхода.

Каждая вершина метаграфа ставится в соответствие одной или нескольким вершинам исходного графа, причём каждой вершине исходного графа соответствует как минимум одна вершина метаграфа. Таким образом, алгоритм сведения задачи offline-распознавания к задаче online-распознавания, предложенный в этой работе, состоит из следующих шагов:

1. Построение скелета символа и представление его в виде графа.
2. Построение метаграфа по данному графу.
3. Построение обхода метаграфа.
4. Построение обхода исходного графа по обходу метаграфа.
5. Генерализация обхода графа посредством агрегирования вершин и рёбер.

В качестве первого шага при построении метаграфа в исходном графе выделяется минимальный базис циклов, т.е. множество циклов с минимальной суммарной длиной, образующее базис пространства циклов для данного графа (множество циклов, через сумму которых можно выразить любой цикл графа). Затем метаграф строится по следующим правилам:

I. Построение вершин:

1. Каждая вершина исходного графа, не входящая ни в один из циклов, переходит в метаграф без изменений, т.е. в метаграф добавляется одна новая вершина, соответствующая этой исходной вершине и только ей.
2. Каждый цикл из базиса исходного графа соответствует одной новой вершине метаграфа, т.е. эта вершина соответствует всем вершинам из этого цикла исходного графа и только им. (Вершины исходного графа, входящие в несколько циклов, таким образом, соответствуют нескольким вершинам метаграфа.)

II. Построение рёбер:

1. Каждое ребро исходного графа, соединяющее две вершины, не входящие ни в один из циклов, переходит в метаграф без изменений, т.е. между соответствующими им вершинами проводится ребро той же длины.
2. Если вершина исходного графа не входит ни в один из циклов, но соединена ребром с вершиной, входящей в один или несколько циклов, то на метаграфе проводится рёбра такой же длины, соединяющие эту вершину с каждой из вершин, соответствующих этим циклам.
3. Если две вершины исходного графа, входящие в два разных цикла, соединены ребром, то на метаграфе проводится ребро такой же длины, соединяющее две вершины, соответствующие этим циклам.
4. Если два цикла исходного графа имеют хотя бы одну общую вершину, то на метаграфе проводится ребро нулевой длины, соединяющее две вершины, соответствующие этим циклам.

По этим правилам метаграф строится однозначно при условии однозначности выбора минимального базиса циклов исходного графа, что всегда выполняется в том случае, когда исходный граф является скелетом плоской фигуры, и длина ребра исходного графа определяется как евклидово расстояние между его концами.

Пример построения метаграфа показан на рисунке 2. Метки на рёбрах соответствуют длинам, назначенным данным рёбрам. В этом примере два цикла исходного графа превращаются в две вершины (Ц1 и Ц2) метаграфа, причём вершина Ц1 метаграфа соответствует вершинам 1, 2 и 3 исходного графа, а вершина Ц2 – вершинам 1, 3 и 4. Поскольку из-за того, что циклы могут иметь общие вершины, метаграф не обязательно будет являться деревом, для дальнейшего обхода будет использоваться его минимальное остовное дерево. При этом ни одно ребро, соответствующее ребру исходного графа, не будет потеряно при построении минимального остовного дерева, т.к. циклы в метаграфе возникают только в результате применений правил 2 и 4 построения рёбер метаграфа.

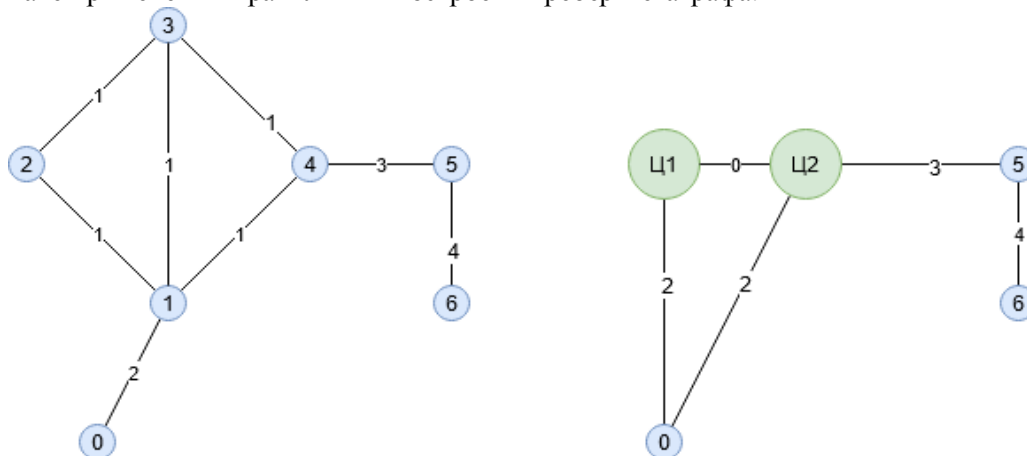


Рисунок 2. Пример графа и его метаграфа.

2.3. Обход графа

Помимо необходимости прохода через все рёбра графа, маршрут, восстанавливающий возможную траекторию движения пера, должен правдоподобно приближать реальную траекторию. Одним из эвристических критериев такого приближения можно считать минимизацию повторных проходов через рёбра графа. В случае, если граф имеет форму дерева (с вершиной дерева в той вершине графа, которая выбрана в качестве начальной), это достигается посредством рекурсивного обхода поддеревьев в порядке увеличения их высоты, с возвратом на предыдущий уровень после полного обхода всех поддеревьев обрабатываемой вершины. В данной работе предлагается при построении обхода графа основываться на таком обходе минимального остовного дерева его метаграфа: хоть такой обход и не всегда приводит к оптимальному результату, он позволяет конструировать пригодные для распознавания маршруты, что подтверждается экспериментальным исследованием.

Прежде всего в графе выбирается начальная вершина, представляющая собой вершину степени 1 (конец штриха) с минимальной суммой координат (т.е. находящаяся как можно ближе к левому верхнему углу). После этого производится первый обход метаграфа (как дерева с корнем в начальной вершине), во время которого для каждой вершины определяется высота соответствующего ей поддерева. Дерево обходится в глубину, его листьям присваивается значение высоты 1, а остальным вершинам – значение, на единицу большее максимального из значений высоты её дочерних вершин.

Затем производится второй обход метаграфа и построение трассы в соответствии со следующим алгоритмом:

1. Если текущая вершина не соответствует циклу исходного графа, то занести её в трассу и спуститься в ещё не обойдённую дочернюю вершину с минимальной высотой.
2. Если текущая вершина соответствует циклу исходного графа, то выбрать финальную вершину выхода, соответствующую дочерней вершине с максимальной высотой, а также направление обхода цикла (по часовой стрелке). Двигаться по соответствующему циклу исходного графа от точки входа (из предыдущей вершины) до ближайшей точки выхода, соответствующей ещё не обойдённой вершине метаграфа, не совпадающей с финальной

вершиной выхода. Если не обойдена только финальная вершина выхода, то двигаться по циклу до этой вершины. При этом заносить в трассу все пройденные вершины цикла исходного графа.

3. Если дочерних вершин нет, или они все обойдены, то вернуться на уровень вверх, занеся текущую вершину в трассу. Для вершины, соответствующей циклу исходного графа, это означает движение по циклу до точки первоначального входа в цикл с занесением в трассу всех пройденных вершин.

При таком методе обхода не будет потеряно ни одно ребро, принадлежащее исходному графу, т.к. обходятся все вершины и рёбра метаграфа, причём каждому ребру, не входящему ни в один из циклов исходного графа, соответствует ребро метаграфа (при построении остовного дерева исключаются только рёбра, соединяющие соответствующие циклам вершины метаграфа и не имеющие аналогов на исходном графе), а входящие в один и более циклов рёбра обходятся при обработке вершины метаграфа, соответствующей данному циклу.

Такой обход позволяет по данному метаграфу получить трассу, т.е. последовательность прохода по вершинам исходного графа. Сортировка дочерних вершин по высоте, помимо минимизации количества возвратов, вытекает из специфики задачи: так как короткие ответвления будут обходиться первыми, метод можно применять для построения обхода не только отдельных символов, но и более длинной последовательности рукописного текста (слова, написанного без отрыва пера от бумаги).

2.4. Генерализация маршрута

Скелет фигуры обычно имеет большое количество близко расположенных вершин, возникающих в местах высокой кривизны штрихов. При построении обхода этого скелета в получившемся маршруте будет много последовательно идущих близких вершин, соединённых короткими рёбрами. Агрегирование таких вершин позволяет существенно (в некоторых случаях – на порядок) сократить объём входных данных без существенных потерь информации о форме символа, что облегчает обучение классификатора при решении задачи online-распознавания. Для этого производится прореживание трассы (т.е. последовательности вершин из маршрута обхода): последовательное удаление из неё всех вершин степени 2 (т.е. не терминальных и не узловых), соседи которых в трассе тоже имеют степень 2 и находятся на расстоянии от рассматриваемой вершины, не превышающем некоторого порогового значения.

В результате этих операций из графа получается последовательность его вершин, которая считается реконструированным возможным следом пера при написании рассматриваемого символа или текста.

3. Распознавание символов

Основным критерием пригодности восстановленного возможного следа пера для распознавания символа алгоритмами online-распознавания является сравнение качества работы алгоритма распознавания на реальной траектории пера и на восстановленной по записи символа. Критерием же возможности распознавания посредством сведения задачи offline-распознавания к online-распознаванию является сравнение качества распознавания с использованием предложенного алгоритма и распознавания традиционными методами offline-распознавания.

Для распознавания символов с использованием получившейся трассы использовался алгоритм, реализованный в работе [2]. Этот алгоритм основан на использовании рекуррентной нейронной сети, состоящей из 24 слоёв управляемых рекуррентных блоков с сигмоидной функцией активации.

Предобучение модели не производилось; разбиение на обучающую и тестовую выборки проводилось в соотношении 4:1. Обучающая выборка подверглась искусственному увеличению посредством случайных сдвигов точек исходных последовательностей.

В качестве альтернативного метода offline-распознавания использовалась нейросеть VGG16[4], предобученная на базе ImageNet. Разбиение на обучающую и тестовую выборки проводилось также в соотношении 4:1.

4. Эксперименты

Эксперименты проводились на наборе данных, основанном на том, который был предоставлен в работе [2]. Набор данных из [2] представляет собой последовательности точек, соответствующих положениям пера, записанные с помощью инструмента Low Power Projected Capacitive Touchpad Development Kit от Microchip [4], сенсорного контроллера, записывающего последовательности координат при движении пера. Набор данных состоит из шести классов, соответствующих различным символам: A, B, E, P, S, W, и седьмого класса None, содержащего различные символы, не соответствующие ни одной из букв других классов. Для каждого класса имеется по 100 примеров.

Для получения изображений текста для offline-распознавания эти символы были растеризованы. Впоследствии алгоритм восстановления возможного следа пера работал с этими растеризованными символами как с изображениями, без использования исходных данных о положении пера. Примеры получившихся изображений показаны на рисунке 3.

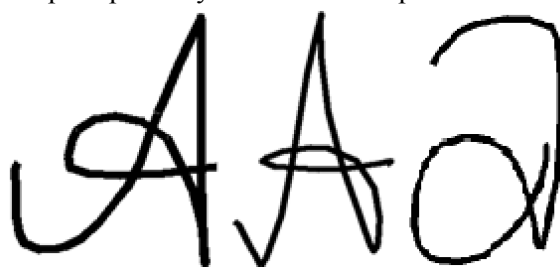


Рисунок 3. Примеры изображений, класс «А».

Результаты экспериментального исследования методов распознавания приведены в таблице 1. Методы обозначены следующим образом:

RNN-True – рекуррентная нейронная сеть, обученная на истинных данных о положении пера

RNN-Offline – рекуррентная нейронная сеть, обученная на восстановленной траектории пера.

VGG – свёрточная сеть VGG16, предобученная на ImageNet и дообученная на изображениях из набора.

Таблица 1. Результаты экспериментального исследования.

Метод	Точность
RNN-True	0.93
RNN-Offline	0.87
VGG	0.60

Видно, что предложенный метод предсказуемо отстаёт от online-распознавания по истинным данным, но при этом он значительно опережает по своим характеристикам метод offline-распознавания, несмотря на то, что сеть VGG16 предобучалась на наборе данных ImageNet, а модели online-распознавания были обучены с нуля на достаточно малой обучающей выборке.

5. Заключение

Был предложен метод сведения задачи offline-распознавания рукописных символов к задаче online-распознавания с последующим распознаванием посредством рекуррентной нейронной сети. Экспериментальная оценка показала, что использование рекуррентной нейронной сети совместно с предложенным алгоритмом для online-распознавания существенно превосходит по точности свёрточную нейронную сеть и обеспечивает распознавание даже в условиях малой обучающей выборки.

Дальнейшие направления исследования могут включать в себя совершенствование метода восстановления траектории пера, эксперименты с распознаванием непрерывных последовательностей символов, а также эксперименты за зашумлённых данных.

6. Благодарности

Работа выполнена при поддержке РФФИ, грант № 20-01-00664.

7. Литература

- [1] Местецкий, Л.М. Непрерывная морфология бинарных изображений: фигуры, скелеты, циркуляры / Л. М. Местецкий – Москва: Физматлит, 2009. – 288 с.
- [2] Tobias Fischer. Проект «Online Handwritten Character Recognition with capacitive sensors» [Электронный ресурс]. – Режим доступа: <https://github.com/tobiasfshr/online-handwritten-character-recognition-capacitive-sensors>.
- [3] Simonyan, K. Very deep convolutional networks for large-scale image recognition / K. Simonyan, A. Zisserman // arXiv preprint arXiv: 1409.1556, 2014.
- [4] Инструмент Low Power Projected Capacitive Touchpad Development Kit [Электронный ресурс]. – Режим доступа: <https://www.microchip.com/DevelopmentTools/ProductDetails/dm160219>.

Handwritten Text Recognition Using Reconstructed Pen Trace with Medial Representation

S.P. Arseev¹, L.M. Mestetsky²

¹Lomonosov Moscow State University, Leninskiye Gory 1, Moscow, Russia, 119991

Abstract. This article is dedicated to the reduction of the offline handwritten character recognition task (using a handwriting image as input) to the online recognition task (using a pen trace as input). A method of the pen trace reconstruction from handwritten text image using its medial representation (a skeleton) is proposed. Experimental comparison of symbols recognition quality using the image, the true pen trace and the reconstructed pen trace is carried out.