

Классификация штрихов рукописного текста на основе расстояния Фреше

Н. М. Пронина¹, Л. М. Местецкий^{1,2}

¹МГУ имени М. В. Ломоносова, ГСП-1, Ленинские горы, д. 1, Москва, 119991, Россия

²НИУ ВШЭ, Покровский бульвар, д. 11, Москва, 109028, Россия

Аннотация

В работе ставится задача по разработке интерпретируемой метрики сходства каллиграфических элементов рукописного текста, штрихов. Потребность в такой метрике часто возникает в задачах навигации по тексту.

Для выделения штрихов текст переводится в бинарное изображение, для всех связанных компонент строится скелетное представление. Оно предоставляет достаточно точную информацию о том, как двигалось перо во время написания текста. Далее производится разрезание скелетного представления в виде геометрического графа. Штрихи формируются в виде подграфов.

В результате работы предложен новый метод, в основе которого лежит использование расстояния Фреше между кривыми на плоскости. Представленная метрика была применена в задачах классификации штрихов и распознавания букв на основе биграмм и триграмм штрихов.

Ключевые слова

Метрика, штрих, расстояние Фреше, скелет.

Strokes classification of handwritten text based on the Fréchet distance

N. M. Pronina¹, L. M. Mestetskiy^{1,2}

¹ M. V. Lomonosov MSU, GSP-1, 1 Leninskiye Gory, Moscow, 119991, Russia

² HSE University, 11 Pokrovsky Bulvar, Moscow, 109028, Russia

Abstract

The paper aims to develop an interpretable similarity metric of the calligraphic handwritten elements, strokes. The need for such a metric often arises in text navigation tasks.

To extract strokes, the text is converted into a binary image and a skeletal representation is constructed for all connected components. It provides accurate information about how the pen moved while writing the text. Next, the skeletal representation is sliced into a geometric graph. The strokes are formed as subgraphs.

As a result, the new method is proposed which is based on the use of Fréchet distance between plane curves. The presented metric has been applied to the tasks of stroke classification and letter recognition based on stroke bigrams and trigrams.

Keywords

Metric, stroke, Fréchet distance, topological skeleton.

ГрафиКон 2023: 33-я Международная конференция по компьютерной графике и машинному зрению, 19-21 сентября 2023 г.,

Институт проблем управления им. В.А. Трапезникова Российской академии наук, г. Москва, Россия

✉ natalka-pronina@mail.ru (Н. М. Пронина); mestlm@mail.ru (Л. М. Местецкий)

🆔 0009-0008-1336-4512 (Н. М. Пронина); 0000-0001-6387-167X (Л. М. Местецкий)

© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Введение

Задача перевода изображений рукописного или печатного текста в текстовые данные (OCR) популярна уже давно. Точность для распознавания печатных документов по чётким изображениям в настоящее время превышает 99% [1, 2]. Точность для распознавания документов с рукописным текстом составляет 80%-90%, и как правило она достигается с применением нейронных сетей.

Возникла гипотеза, что общие представления о каллиграфии могут помочь в навигации по тексту. Так как каждый человек использует ограниченное количество каллиграфических элементов, штрихов, можно попробовать составить «алфавит» штрихов для конкретного почерка и свести задачу распознавания букв к задаче распознавания штрихов. Предполагается, что алфавит штрихов будет сильно меньше алфавита букв, что позволит существенно уменьшить объём данных для обучения.

Подходы с использованием штрихового разложения текста встречаются в литературе крайне редко. Такой подход анонсировала компания ParaScrip [3] в начале 2000-х годов, но он не был реализован в полном виде. В статье [4] приведена и реализована идея построения штрихового представления, которое в дальнейшем предположительно может использоваться при генерации признаков для распознавания текста.

2. Используемые понятия

Далее приведены определения основных понятий.

Штрих — схема движения пишущего инструмента при написании букв, формально — ломаная без самопересечений.

Фигура — многоугольник с многоугольными отверстиями.

Пустой круг фигуры — круг, расположенный внутри фигуры, такой, что никакая граничная точка фигуры не лежит внутри него.

Максимальный пустой круг — пустой круг фигуры, не содержащийся ни в каком другом пустом круге фигуры.

Скелет фигуры — множество центров ее максимальных пустых кругов, имеет вид геометрического графа [5].

Цепь — это конечная последовательность вершин скелетного графа, в которой каждая вершина соединена ребром со следующей в последовательности вершиной.

Цикл — цепь, у которой первая и последняя вершины совпадают.

3. Сегментация штрихов

Алгоритм сегментации основывается на методе, описанном в [5]. Результатом сегментации является представление изображенного текста в виде упорядоченного множества штрихов (рис. 1).



Рис. 1: Результат сегментации штрихов.

Для начала входное изображение переводится в бинарное. Изображение текста аппроксимируется многосвязными многоугольными фигурами — многоугольниками с

многоугольными дырами, строится скелетное представление. Скелетизация многоугольных фигур состоит в построении срединных осей – множества точек центров вписанных в фигуры окружностей.

В полученном скелетном графе выделяются подграфы, описывающие штрихи Кольца и Цепи. Выделение подграфов можно интерпретировать как разрезание геометрического графа по вершинам. Выделение штрихов осуществляется на основе следующих операций:

1. Выделение кольцевых циклических штрихов
2. Разрезание графа по вершинам третьей степени и более высоких степеней
3. Для полученных штрихов определяется направление и последовательность их прохождения в рукописном тексте

4. Расстояние Фреше

В качестве количественной характеристики различия кривых было выбрано расстояние Фреше. Для простоты эту характеристику будем называть метрикой, несмотря на то, что неравенство треугольника для неё не выполняется. Расстояние Фреше хорошо сочетается с нашей задачей, потому что оно рассматривает кривые как упорядоченные множества точек, чем и являются штрихи.

Расстояние Фреше обычно иллюстрируется следующим образом: рассматривается движение собаки вдоль одной кривой, которую держит на поводке хозяин, идущий вдоль другой кривой. Оба проходят от начальной точки до конечной, меняя скорость, но не возвращаясь. Расстояние Фреше между этими двумя кривыми — это длина самого короткого поводка, с которым возможно пройти кривые. Интуитивное определение симметрично относительно двух кривых, можно считать, что собака выгуливает хозяина.

Формализуем это определение [6, 7].

Опр: Кривой называется непрерывное отображение из $[0, 1]$ в \mathbb{R}^d .

Кривая π называется замкнутой, если $\pi(0) = \pi(1)$.

Обозн: $\|\cdot\|$ - норма

Опр: Репараметризацией называется непрерывное взаимно-однозначное отображение $f : [0, 1] \rightarrow [0, 1]$. Репараметризация сохраняет ориентацию, если $f(0) = 0$ и $f(1) = 1$.

Рассмотрим кривые π, σ и их репараметризации - f и g соответственно.

Опр: Шириной двух репараметризаций называется величина

$$width_{f,g}(\pi, \sigma) = \max_{s \in [0,1]} \|\pi(f(s)) - \sigma(g(s))\|$$

В интуитивном определении ширина — это максимальная длина поводка при фиксированном законе движения хозяина и собаки.

Опр: Расстоянием Фреше между двумя кривыми называется величина

$$d_F(\pi, \sigma) = \inf_{f,g:[0,1] \rightarrow [0,1]} width_{f,g}(\pi, \sigma),$$

где f и g - репараметризации, сохраняющие ориентацию кривых π и σ соответственно.

То есть производится поиск двух репараметризаций минимальной ширины.

5. Дискретное расстояние Фреше

На практике для подсчёта расстояния Фреше две кривые аппроксимируются ломаными и вычисляется дискретное расстояние Фреше. Это необходимо, чтобы не иметь дело с оптимизацией по несчётному множеству всех репараметризаций f и g .

В интуитивном определении расстояния Фреше теперь рассматриваются положения поводка только в вершинах двух ломаных, внутри рёбер — не рассматриваются. При таком упрощении нужно будет найти соответствие вершин ломаных, при котором $width$ будет достигать минимума, а таких соответствий конечное число.

Формализуем определение дискретного расстояния Фреше [8].

Опр: Ломаной называется непрерывное отображение из $[0, n]$ в \mathbb{R}^d , такое, что

$$P(i + \lambda) = (1 - \lambda)P(i) + \lambda P(i + 1) \quad \forall i = \overline{0, n-1}, \quad \forall \lambda \in [0, 1]$$

Точки $(P(0), P(1), \dots, P(n))$ называются вершинами ломаной P .

Рассмотрим ломаные P и Q и их вершины - (u_1, \dots, u_p) , (v_1, \dots, v_q) .

Опр: Соответствием L ломаных P и Q называется последовательность пар $(u_{a_1}, v_{b_1}), (u_{a_2}, v_{b_2}), \dots, (u_{a_m}, v_{b_m})$, таких что:

$$a_1 = 1, \quad b_1 = 1, \quad a_m = p, \quad b_m = q$$

$$a_{i+1} = a_i \text{ или } a_{i+1} = a_i + 1, \quad \forall i = \overline{1, p-1}$$

$$b_{i+1} = b_i \text{ или } b_{i+1} = b_i + 1, \quad \forall i = \overline{1, q-1}$$

Опр: Шириной соответствия двух ломаных называется величина

$$width(L) = \max_{i=1, m} \|u_{a_i} - v_{b_i}\|$$

Опр: Дискретным расстоянием Фреше между двумя ломаными называется

$$d_f(P, Q) = \min_L width(L),$$

где L - соответствие ломаных P и Q .

Дискретное расстояние Фреше может быть вычислено с помощью метода динамического программирования за время $O(p \times q)$, где p, q — количество вершин ломаных. С помощью обратного прохода можно также вычислить соответствие ломаных. Алгоритмы вычисления прямого и обратного проходов представлены на рисунках 2, 3.

6. Модификации расстояния Фреше

Так как в определении дискретного расстояния Фреше не учитываются точки внутри рёбер, оно всегда получается не меньше расстояния Фреше. Но можно легко модифицировать алгоритм так, чтобы подсчитывались расстояния не только между вершинами, но и расстояния между рёбрами, между ребром и вершиной. Таким образом для двух ломаных мы сможем посчитать точное расстояние Фреше, но и требуемый объём памяти увеличится в 4 раза.

Стоит отметить, что такая модификация имеет смысл, если выполнена неточная аппроксимация кривой, в этом случае результат будет существенно отличаться от дискретного расстояния Фреше.

```

def frechet(P, Q, dist):
    p = количество вершин P
    q = количество вершин Q
    D = массив размера p * q
    for i in range(p):
        for j in range(q):
            d = dist(P[i], Q[j])
            if i == 0 and j == 0:
                D[i, j] = d
            elif i > 0 and j == 0:
                D[i, j] = max(D[i-1, 0], d)
            elif i == 0 and j > 0:
                D[i, j] = max(D[0, j-1], d)
            elif i > 0 and j > 0:
                D[i, j] = max(d,
                               min(D[i-1, j], D[i-1, j-1], D[i, j-1]))
    return D[p - 1, q - 1]

```

Рис. 2: Прямой проход динамического программирования для подсчёта расстояния Фреше

```

def L(D):
    i, j = p - 1, q - 1
    stack = [(i, j)]
    while i != 0 and j != 0:
        m = argmin([D[i - 1, j], D[i - 1, j - 1], D[i, j - 1]])
        if m == 0:
            i, j = i - 1, j
        elif m == 1:
            i, j = i - 1, j - 1
        elif m == 2:
            i, j = i, j - 1
        stack.append((i, j))
    while j != 0:
        i, j = i, j - 1
        stack.append((i, j))
    while i != 0:
        i, j = i - 1, j
        stack.append((i, j))
    return stack[::-1]

```

Рис. 3: Обратный проход динамического программирования для нахождения соответствия вершин

Можно заметить, что расстояние Фреше неустойчиво к выбросам. Например, если хозяин с собакой почти всегда шли рядом, но в какой-то момент собака убежала и сразу же вернулась, то расстояние между траекториями будет так же велико, как если бы хозяин с собакой весь путь прошли далеко друг от друга.

Из таких рассуждений рождается ещё одна модификация расстояния Фреше: можно вычислять между соответствующими вершинами не максимальное расстояние, а среднее (рис. 4).

С помощью обратного прохода динамического программирования найдем соответствие ломаных $L = \{(u_{a_1}, v_{b_1}), (u_{a_2}, v_{b_2}), \dots, (u_{a_m}, v_{b_m})\}$.

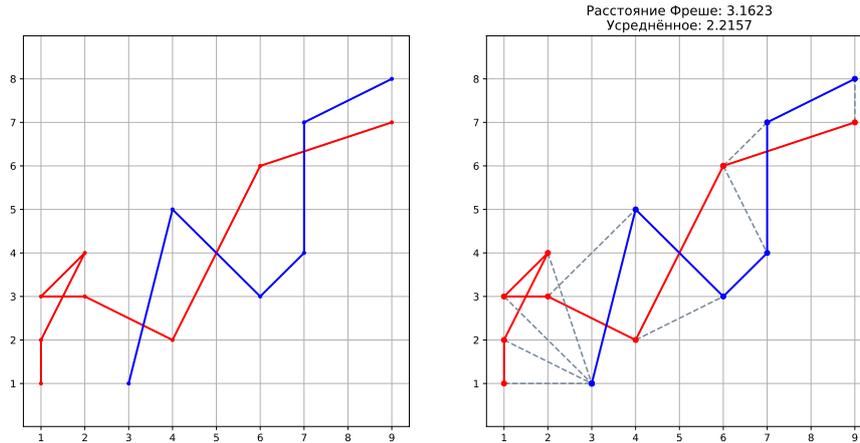


Рис. 4: Соответствие ломаных

Опр: Средним расстоянием Фреше между двумя ломаными называется

$$d_{mid}(P, Q) = \frac{1}{m} \sum_{i=1}^m \|u_{a_i} - v_{b_i}\|$$

7. Нормализация штрихов

Важно отметить, что расстояние Фреше довольно сильно зависит не только от взаимного расположения кривых, но и от направления обхода каждой кривой, поэтому так важно сначала произвести нормализацию.

Опр: Диаметром штриха назовём отрезок, соединяющий две наиболее удалённые точки.

Опр: Центром штриха назовём середину диаметра штриха.

Опр: Центром тяжести штриха назовём точку, координаты которой равны усреднённой сумме соответствующих координат ломаной, в виде которой представлен штрих.

Есть два принципиально разных типа штрихов — замкнутые и незамкнутые, кольца и цепи соответственно. Можно заметить, что любой аккуратный рукописный текст имеет один наклон букв. Поэтому можно предположить, что диаметр штриха — это устойчивая характеристика для колец. В качестве нормализации колец будем совмещать штрихи с помощью параллельного переноса так, чтобы центры штрихов совпали. Для цепей — чтобы совпали центры тяжести штрихов, так как диаметр цепей неустойчив из-за длинных соединительных участков.

Осталось исправить возможное неправильное направление обхода штрихов. Все кольца упорядочим против часовой стрелки, а начальными точками колец назначим пару точек, на которых достигается минимальное евклидово расстояние между множествами точек штрихов.

$$p_0, q_0 : d(p_0, q_0) = \inf_{p \in P, q \in Q} \{d(p, q)\}$$

Для цепей сравним расстояния между началами штрихов, между началом и концом. Если второе расстояние меньше, то меняем направление одного из штрихов на противоположное.

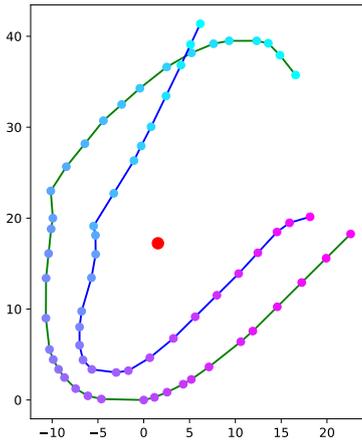


Рис. 5: Нормализация для цепей

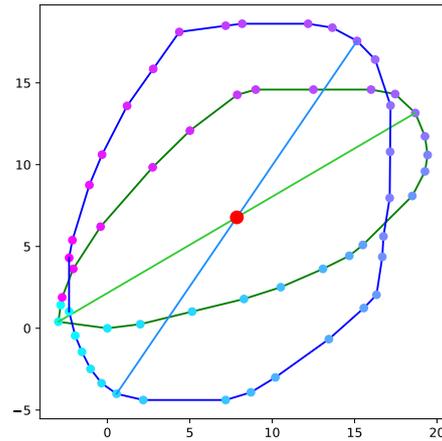


Рис. 6: Нормализация для колец

8. Эксперименты

8.1. Сравнение метрик

Решим, какая метрика лучше подходит для сравнения штрихов: точное расстояние Фреше или дискретное, с усреднением или без. На рисунке 7 приведены результаты для штрихов реального почерка. Ожидается получить выраженную блочно-диагональную матрицу, так как сравнивались 3 группы похожих штрихов. Видно, что все четыре метрики справляются с задачей. Далее будем использовать усредненное дискретное расстояние Фреше, так как для подсчёта точного требуется больше времени, а результат отличается не сильно.

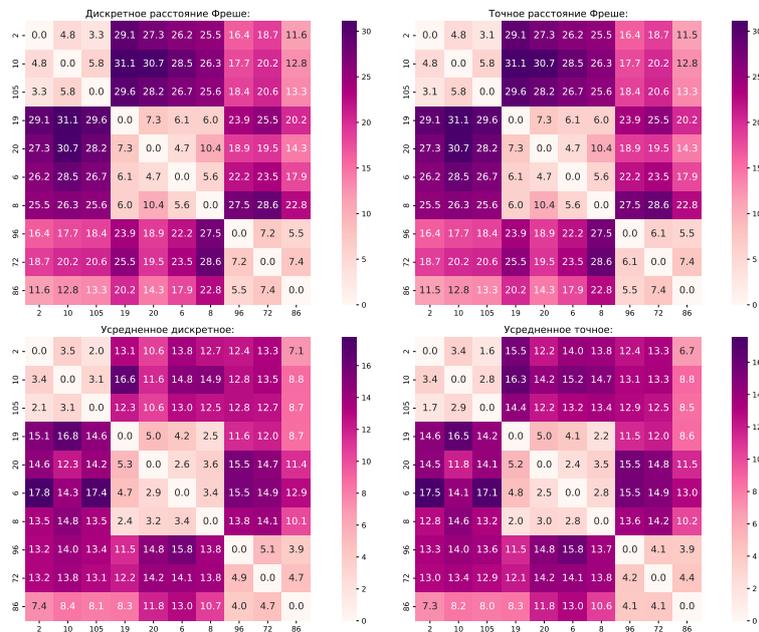


Рис. 7: Сравнение метрик

8.2. Формирование алфавита штрихов

Посчитаем полную матрицу попарных расстояний для эталонного почерка и проведём иерархическую кластеризацию по полученной матрице. При кластеризации укажем допустимое для объединения межкластерное расстояние, так размер алфавита штрихов будет подбираться автоматически. В результате экспериментов получилось 14 кластеров.

На рисунке 8 приведено по 5 представителей самых больших кластеров. В них узнаются основные каллиграфические элементы письма, а значит метрика Фреше хорошо подходит для данной задачи.

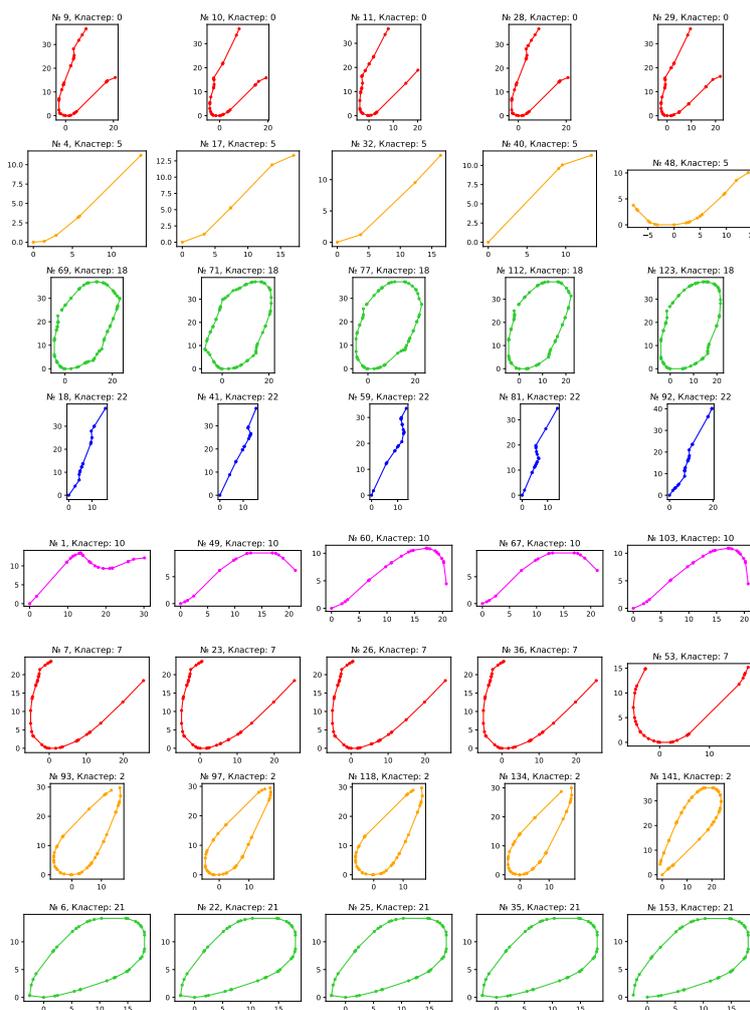


Рис. 8: Кластеризация штрихов

8.3. Классификация штрихов

Теперь распознаем штрихи реального почерка, используя найденный алфавит. Для каждого штриха из реального почерка найдём ближайшего соседа из алфавита.

На рисунке 9 указаны номер штриха из распознаваемого почерка, номер ближайшего штриха из алфавита и его кластер, расстояние между ними. Визуально качество хорошее, но есть и ошибки. Штрих 75 был неправильно распознан, так как расстояние Фреше неустойчиво к выбросам в виде длинных соединительных участков между

буквами, поэтому как ближайший был найден штрих 7 с таким же длинным участком. Предложенная метрика может быть улучшена путём уменьшения влияния длинных соединительных участков на близость штрихов [9]. Несмотря на это для штриха 0 нашёлся верный штрих, но расстояние Фреше получилось слишком большим.

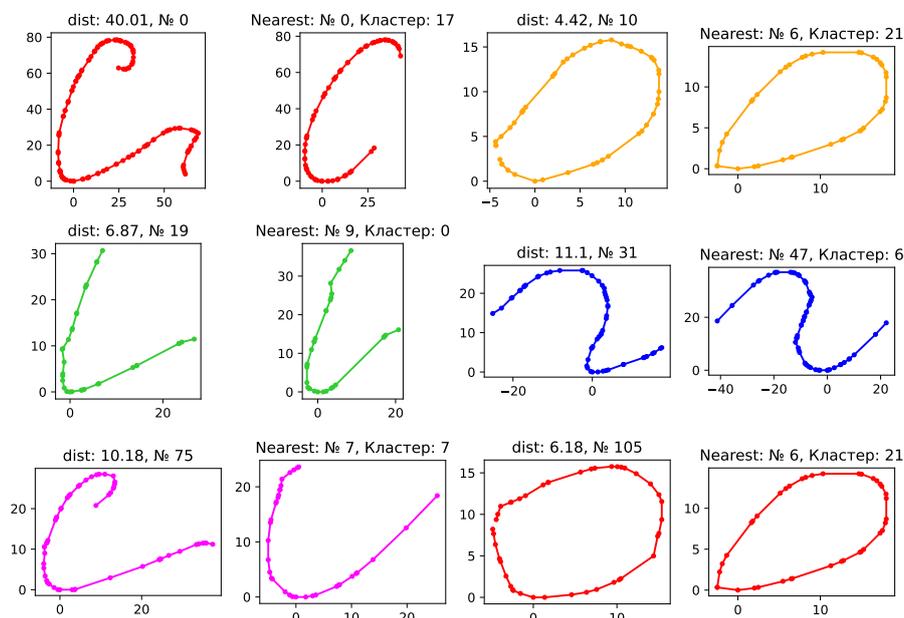


Рис. 9: Классификация штрихов

8.4. Распознавание букв

Поставим задачу поиска букв по тексту на основе полученных классифицированных штрихов. Требуется найти все связанные биграммы/триграммы штрихов, подходящие под запрос. Например, для поиска буквы «а» запрос составим следующим образом: первый штрих принадлежит множеству небольших кругов из алфавита, второй — короткая изогнутая линия из алфавита.

На рисунках 10, 11, 12 приведены примеры распознавания некоторых букв. Красным выделены ошибочные штрихи/неверно распознанные буквы. Довольно много ошибок связано с неправильной сегментацией — несколько штрихов сливаются в один.

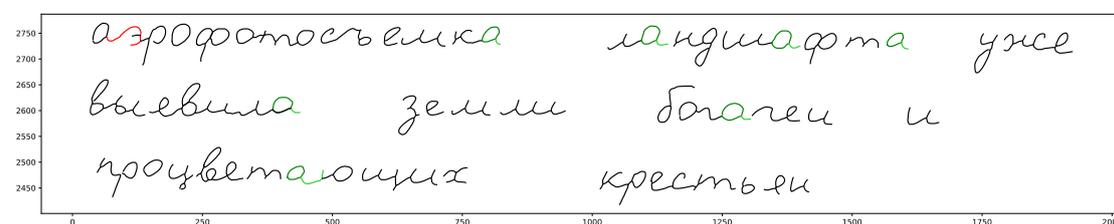


Рис. 10: Поиск буквы «а»

В таблице 1 приведены меры успешности распознавания для некоторых букв.

Обнаружение буквы «и» на месте буквы «ш» не считалось ошибкой, так как без смыслового анализа слова разделить эти случаи невозможно.

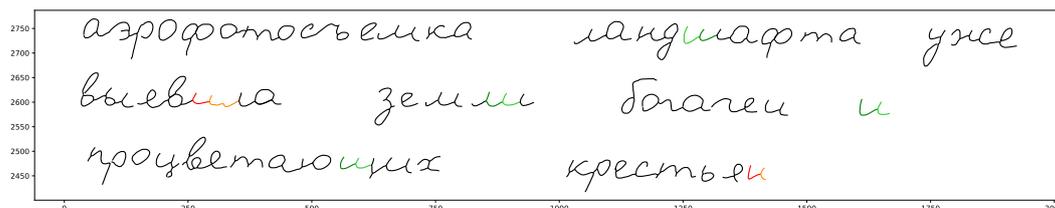


Рис. 11: Поиск «и»



Рис. 12: Поиск «ф»

Таблица 1

Precision и Recall распознавания букв в текстовом фрагменте

Буква	Precision	Recall
«а»	0.9677	0.9677
«и»	0.9069	0.8125
«ф»	0.8333	0.7142

9. Заключение

В данной работе была получена интерпретируемая метрика сходства штрихов, которая соотносится с интуитивным представлением о сходстве линий. Были поставлены эксперименты, подтверждающие целесообразность выбора представленной метрики. Также был предложен новый подход к распознаванию каллиграфических элементов.

В дальнейшем планируется применить предложенную метрику в задачах распознавания стиля почерка, навигации в рукописном контексте по ключевым словам.

10. Благодарности

Работа выполнена при поддержке Российского научного фонда, грант № 22-68-00066, <https://rscf.ru/project/22-68-00066/>.

Список литературы

- [1] K. Wang, B. Babenko, S. Belongie, End-to-end scene text recognition (2011) 1457–1464. doi:10.1109/ICCV.2011.6126402.
- [2] Y. Nagata, J. Otao, D. Haraguchi, S. Uchida, Truetype transformer: Character and font style recognition in outline format (2022). arXiv:2203.05338.
- [3] A. Filatov, V. Nikitin, A. Volgunin, P. Zelinsky, The addressscript™ recognition system for handwritten envelopes (1999) 157–171.
- [4] L. Mestetskii, I. Reyer, T. Sederberg, Continuous approach to segmentation of handwritten text (2002) 440–445. doi:10.1109/IWFHR.2002.1030950.

- [5] L. Mestetskiy, Непрерывная морфология бинарных изображений: фигуры, скелеты, циркуляры, М.: ФИЗМАТЛИТ, 2009. URL: https://www.researchgate.net/publication/298385622_Nepreryvnaa_morfologia_binarnyh_izobrazenij_figury_skelety_cirkulyary.
- [6] H. Alt, M. Godau, Computing the fréchet distance between two polygonal curves., International Journal of Computational Geometry & Applications 05 (1995) 75–91. URL: <https://doi.org/10.1142/S0218195995000064>. doi:10.1142/S0218195995000064.
- [7] A. Driemel, S. Har-Peled, C. Wenk, Approximating the fréchet distance for realistic curves in near linear time, CoRR abs/1003.0460 (2010). URL: <http://arxiv.org/abs/1003.0460>. arXiv:1003.0460.
- [8] T. Eiter, H. Mannila, Computing discrete frechet distance (1994). URL: https://www.researchgate.net/publication/228723178_Computing_Discrete_Frechet_Distance.
- [9] E. Sriraghavendra, K. Karthik, C. Bhattacharyya, Fréchet distance based approach for searching online handwritten documents 1 (2007) 461–465. doi:10.1109/ICDAR.2007.4378752.