

DOI: 10.15514/ISPRAS-2026-38(3)-41



Определение строковой структуры рукописных документов на основе модели векторного поля

*Л.М. Местецкий, ORCID: 0000-0001-6387-167X <mestlm@mail.ru>
В.С. Смирнова, ORCID: 0000-0000-0000-0000 <smirnovavs@my.msu.ru>
Московский государственный университет имени М.В. Ломоносова,
Россия, 119991, Москва, Ленинские горы, д. 1.*

Аннотация. В работе предложен оригинальный метод построчной сегментации рукописных документов, состоящий в построении векторного поля локальных направлений текста и формировании осевых линий строк на основе интегральных кривых построенного поля. Разработан и реализован алгоритм, робастно работающий с реальными изображениями архивных документов, включая сложные случаи изгибающихся строк, неравномерных межстрочных интервалов и вариативности почерка. Экспериментальные результаты демонстрируют высокую эффективность ($F_1 = 0.92$).

Ключевые слова: сегментация строк; рукописные документы; векторное поле; векторные линии; триангуляция Делоне.

Для цитирования: Местецкий Л.М., Смирнова В.С. Определение строковой структуры рукописных документов на основе модели векторного поля. Труды ИСП РАН, том 38, вып. 3, часть 3, 2026 г., стр. 135–148. DOI: 10.15514/ISPRAS-2026-38(3)-41.

Благодарности: Работа поддержана грантом РНФ № 22-68-00066 «Культурное наследие России: интеллектуальный анализ и тематическое моделирование корпуса рукописных текстов».

Handwritten Text Line Structure Detection via Vector-Field Modeling

*L.M. Mestetskiy, ORCID: 0000-0001-6387-167X <mestlm@mail.ru>
V.S. Smirnova, ORCID: 0000-0000-0000-0000 <smirnovavs@my.msu.ru>
Lomonosov Moscow State University,
GSP-1, Leninskie Gory, Moscow, 119991, Russia.*

Abstract. We propose an original method for line-by-line segmentation of handwritten documents consisting in constructing a vector field of local text orientations and forming centerlines of lines based on integral curves of the constructed field. We develop and implement an algorithm that robustly handles real images of archival documents, including challenging cases with curved lines, non-uniform interline spacing, and handwriting variability. Experimental results demonstrate high effectiveness ($F_1 = 0.92$).

Keywords: text line segmentation; handwritten documents; vector field; vector lines; Delaunay triangulation.

For citation: Mestetskiy L., Smirnova V. Handwritten Text Line Structure Detection via Vector-Field Modeling. Trudy ISP RAN/Proc. ISP RAS, vol. 38, issue 3, part 3, 2026, pp. 135-148 (in Russian). DOI: 10.15514/ISPRAS-2026-38(3)-41.

Acknowledgements. The work was supported by the Russian Science Foundation (project no. 22-68-00066: «Cultural Heritage of Russia: intellectual analysis and thematic modeling of the corpus handwritten texts»).

1. Введение

Большинство современных OCR-систем выполняют распознавание на уровне отдельных строк, поэтому для автоматизированного распознавания текста в рукописных документах важно обеспечить точную построчную сегментацию.

Процесс выделения строк в рукописных документах осложняется вариативностью почерка, плотным расположением и нерегулярностью наклона строк, частыми перекрытиями выносных элементов букв, а также фоновым шумом в виде пятен и потертостей на бумаге, неоднородным освещением и ракурсом съемки страниц документов.

Существуют четыре основных подхода к задаче сегментации строк. Проекционные методы строят профиль плотности пикселей по горизонтали и определяют границы строк по локальным минимумам, что хорошо работает на страницах с горизонтальными строками, имеющими статичный интервал, но слабо приспособлено к изогнутым строкам и перекрытиям [1-2]. Морфологические алгоритмы, такие как RLSA (Run-Length Smoothing Algorithm), объединяют близко расположенные компоненты текста посредством морфологического замыкания, однако выбор оптимального структурирующего элемента остаётся нетривиальной задачей [3-5]. Графовые модели представляют документ в виде графа, где узлы – это связанные компоненты текста, а рёбра отражают их пространственные отношения; сегментация достигается решением задачи минимального графового разреза, обеспечивая гибкость при разделении тесно переплетённых строк [6]. Нейросетевые подходы (FCN, U-Net, RNN/LSTM) демонстрируют наивысшую точность на больших размеченных датасетах, но их применение ограничено высокими вычислительными затратами и необходимостью разметки больших объёмов данных [7].

Существует потребность в методе, сочетающем интерпретируемость и робастность графовых моделей с возможностью локального анализа без дорогостоящего этапа обучения. В данной работе предлагается алгоритм построчной сегментации, в котором строковая структура описывается векторным полем локальных направлений, а строки представляются осевыми линиями (baseline), пригодными для последующего распознавания.

Отметим, что существующие подходы к сегментации строк без обучения, например, минимально-энергетические «швы» в ограниченной обрезке швов (constrained seam carving) из работы [8] или голосование в пространстве Хафа по блокам связанных компонент [9], имеют

ограничения: высокую чувствительность к предобработке и бинаризации; геометрические допущения, нарушаемые сильно искривленными строками, неравномерным межстрочным интервалом, пересечениями и слияниями штрихов, что приводит к множественным локальным минимумам и фрагментации/слиянию линий; выраженная масштабная зависимость (высота строки, толщина штриха, интервал), требующая ручной калибровки окон, фильтров и порогов и ухудшающая перенос между коллекциями.

Цель исследования – разработать и экспериментально оценить предложенный метод построчной сегментации рукописных документов с математически обоснованной структурой, обеспечивающей интерпретируемость и универсальность при обработке разнообразных архивных материалов.

2. Постановка задачи

Исходными данными задачи являются изображения страниц исторических рукописных документов, таких как письма, дневники и заметки, представленные в виде растровых изображений. Для каждого исходного изображения требуется автоматически выполнить сегментацию текстовых строк путём построения множества осевых линий строк. Каждая осевая линия должна быть представлена упорядоченной ломаной, которая аппроксимирует центральную траекторию расположения компонент текста, отражая как глобальную кривизну строки, так и её локальный наклон. Выходом предложенного алгоритма является набор таких ломаных линий, предназначенный для последующего использования в системах оптического распознавания текста и других приложениях анализа рукописных документов. Важно отметить, что как правило в работах по сегментации строк [10] результат задаётся полигональными областями, охватывающими строку. Однако на достаточно искривленных строках, а также строках с пересечениями выпадающих элементов букв такое разделение становится неоднозначным и слабо поддаётся строгой оценке качества, тогда как представление в виде непересекающихся осевых линий [11] надёжно в этих условиях и более удобно для последующей нормализации, выравнивания по траектории и распознавания.

Основные сложности задачи обусловлены вариативностью почерка авторов, неравномерностью межстрочных интервалов и изменчивостью угла наклона строк. Для формализации задачи приняты следующие предположения:

- Осевая линия строки является кусочно-гладкой.
- Преобладающий наклон соседних строк считается постоянным.
- В пределах одной строки локальные участки текста в малой окрестности имеют схожие углы наклона.
- Строки имеют ограниченный наклон относительно горизонтали.

С учетом этих допущений предлагается метод построения осевых линий строк, основанный на модели векторного поля.

3. Модель векторного поля

Предлагаемый подход к строковой сегментации страницы основан на использовании модели векторного поля. Векторное поле – это *отображение*, которое каждой точке плоскости ставит в соответствие *вектор* с началом в этой точке. Векторной линией поля называется линия, в каждой точке которой направление касательной совпадает с вектором поля. Структуру строк можно интерпретировать как семейство векторных линий.

Представим страницу текста как случайное поле, в котором каждая точка может принадлежать к тексту или к фону, а вероятность принадлежности точки тексту задаётся как функция точки. Обозначим точку на странице $g=(x, y)$, а $f(g)$ – вероятность принадлежности точки p к тексту. Не противоречит интуиции предположение, что функция $f(g)$ является дифференцируемой.

Распределение вероятностей $f(g)$ заранее неизвестно, а картинка с текстом рассматривается как конкретная реализация этого распределения. Функция распределения $f(g)$ имеет волнообразный вид, волны соответствуют строкам, и общая форма строки описывается гребнем волны. Тогда задача сегментации строк на странице состоит в восстановлении функции $f(g)$ и вычислении линий, проходящих по гребням волн. Решение задачи должно быть выполнено на основе единственной наблюдаемой реализации этого распределения, представленной страницей текста.

Эти предположения могут быть положены в основание модели, описывающей строковую структуру рукописной страницы. Во-первых, линия гребня волны состоит из точек максимума по направлению, ортогональному направлениям строк. Это значит, что частная производная $f(g)$ в вертикальном направлении в точках на гребне волны равна нулю. А во-вторых, вдоль линии гребня значение функции меняется медленно по сравнению с поперечным гребню направлением, что позволяет определить касательное направление строки как направление минимального изменения функции.

Обозначим точку на странице $g=(x, y)$, а $f(g)$ – вероятность принадлежности точки p к изображению текста. Линия гребня представляет собой точки локальных максимумов $f(g)$ вдоль вертикального направления, в которых $\frac{df(g)}{dy}=0$. А сама эта линия имеет направление $d(g)=(x_g, y_g)$ такое, что $\frac{df(g)}{dv}=0$. Эти свойства линии гребня волны служат основанием для предлагаемого метода выделения строк на странице рукописного текста. Метод основан на идее реконструкции картины векторного поля по изображению рукописной страницы и построения векторных линий поля, соответствующих строкам текста. Метод включает три основных элемента (рис. 1).

1. Определение множества точек, лежащих на гребнях волн функции $f(g)$. Это конечное множество точек $G = \{g_1, g_2, \dots, g_N\}$ образующее нерегулярную решётку на изображении, будем называть гребневым множеством.
2. Определение направления гребней волн $V = \{d_1, d_2, \dots, d_N\}$ во всех точках гребневого множества. Множество пар $(g_i, d_i), i = 1, \dots, N$ представляет собой дискретную картину векторного поля.
3. Построение множества всех векторных линий, проходящих через точки гребневого множества.

Задача сегментации строк сводится к автоматическому выделению осевых линий текста на бинаризованном изображении страницы D . Для точной формализации введём следующие понятия.

Определение 1 (Узловая точка). Пусть $D \subset Z^2$ – множество точек на целочисленной решётке, образованное центрами пикселей текста. Точка $g \in D$ называется узловой, если существует окрестность $U_\varepsilon(g)$ такая, что плотность текстовых пикселей в ней максимальна по сравнению с соседними областями.

Определение 2 (Строковая решётка). Множество всех узловых точек $G = g_1, g_2, \dots, g_N \subset D$ называется строковой решёткой.

Определение 3 (Поле направлений строк).

$$d: G \rightarrow S, S = \{v \in R^2: \|v\|=1\},$$

где $d(g_i)$ – единичный вектор локального направления строки в окрестности g_i .

Определение 4 (Осевая линия строки). Непрерывная упорядоченная ломаная

$$l = (v_1, \dots, v_k), v_k \in G,$$

такая что для каждого её сегмента вектор $v_k v_{k+1}$ совпадает по направлению с $d(v_k)$, называется *осевой линией*.

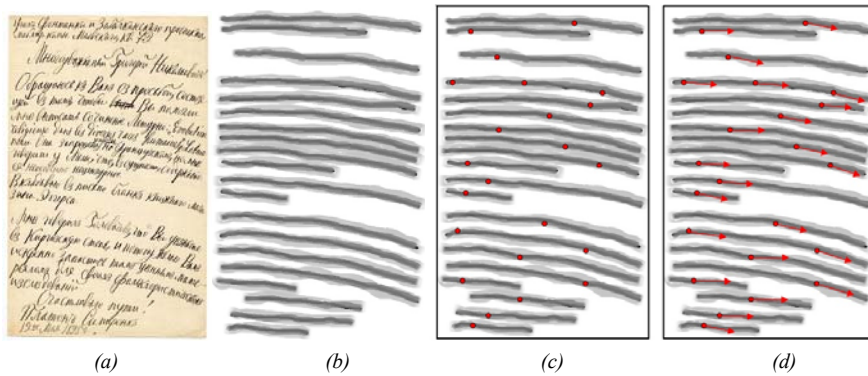


Рис. 1. Векторное поле строк: (a) изображение страницы текста, (b) функция $f(p)$ – вероятность принадлежности точек страницы тексту,

(c) гребневое множество точек, (d) векторы направлений строк в точках гребневого множества.
Fig. 1. Vector field of lines: (a) the image of the page, (b) the function $f(p)$ is the probability of the points belonging to the text, (c) the comb set of points, (d) vectors of the directions of the lines at the points of the comb set.

4. Метод решения

4.1 Предобработка изображения

Входные данные – изображения рукописных страниц в каком-либо растровом графическом формате. Сначала проводится медианное сглаживание для подавления шумов при сохранении структуры текста. Затем выполняется адаптивная бинаризация (рис. 2) с локальным порогом

$$T(x, y) = \frac{1}{\sum_{(i,j) \in N} w(i,j)} \cdot \sum_{(i,j) \in N} I(i,j)w(i,j) - C, N=9 \times 9, C=6,$$

где $w(i, j)$ – веса гауссова ядра. Пиксель (x, y) считается чёрным, если $I(x, y) > T(x, y)$. Наконец, для удаления мелких артефактов применяются последовательно морфологические преобразования: сначала открытие, затем закрытие (композиции эрозии и дилатации) с квадратным структурирующим элементом 3×3 .



Рис. 2. Примеры бинаризации: слева исходное изображение, справа – бинаризованное.
Fig. 2. Binarization examples: the original image on the left and the binarized image on the right.

4.2 Построение строковой решётки

Рукописная строка представляется как область с повышенной плотностью чёрных пикселей по сравнению с межстрочными промежутками. Центры таких скоплений позволяют определить узловые точки, по которым восстанавливается траектория строки. Локальная согласованность направлений текстовых пикселей в окрестности узловых точек позволяет оценить направление векторного поля строки и, тем самым, её наклон.

На основе узлов формируется строковая решётка G , и для каждой точки $g_i \in G$ задаётся вектор локального направления $d(g_i)$, формирующий поле направлений $d: G \rightarrow S$. Далее формируются осевые линии строк как линии, аппроксимирующие линии векторного поля d , проходящие через узловые точки g_i с касательным направлением, совпадающим с вектором $d(g_i)$. Это обеспечивает согласованность траекторий с локальной геометрией и кривизной строк.

В основе метода лежит гипотеза о том, что в пределах короткого фрагмента текста угол наклона текстовой строки статичен, то есть на выбранном фрагменте изображения направление строки практически не меняется. Более того, чем короче фрагмент, тем точнее выполняется это предположение, поскольку локальная структура строки слабо подвержена искажениям и кривизне на малом масштабе. Но при этом длина фрагмента не может быть слишком малой, поскольку необходимо, чтобы пикселей текста было достаточно, для определения локальных характеристик функции $f(p)$.

В соответствии с введенным предположением для перехода к дискретному представлению страницы область изображения D разбивается на N вертикальных полос одинаковой ширины (далее $N = 8$):

$$S_j = \{(x, y) \in D: x \in [j \frac{W}{N}; (j+1) \frac{W}{N}]\}, j=0, \dots, N-1,$$

где W – ширина изображения.

В каждой полосе строится вертикальный переходный профиль (рис. 3) $r_j(y) = \sum_{x: (x, x+1) \in S_j} |D(x+1, y) - D(x, y)|$, где $D(x, y) \in \{0, 1\}$.

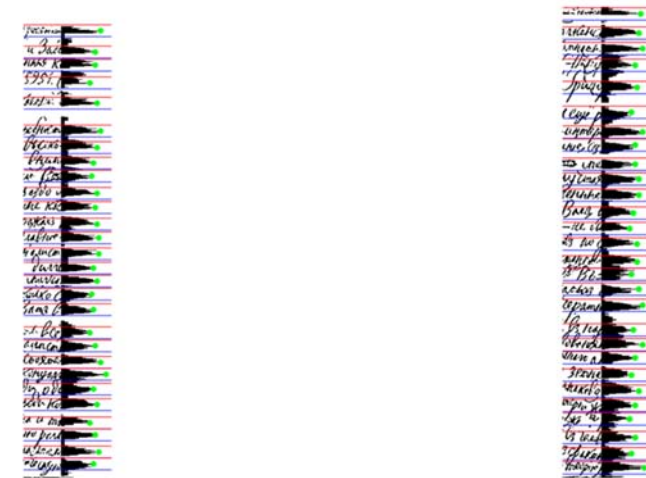


Рис. 3. Фрагменты полученных полос и соответствующих гистограмм строчных профилей. Зеленые точки – локальные максимумы, красные/синие линии – верхняя/нижняя границы описанных фреймов.
Fig.3. Fragments of the extracted strips and the corresponding line-profile histograms. Green points denote local maxima; red/blue lines indicate the upper/lower boundaries of the described frames.

Локальный максимум $r_j(\mathbf{y}_{j,k})$ принимается, если выполняются следующие условия:

$$\begin{aligned} r_j(\mathbf{y}_{j,k-1}) &< r_j(\mathbf{y}_{j,k}) \\ r_j(\mathbf{y}_{j,k}) &< r_j(\mathbf{y}_{j,k+1}) \\ r_j(\mathbf{y}_{j,k}) &\geq \tau_j, \tau_j = 1.5 \times \text{median}(r_j(\mathbf{y})) \end{aligned}$$

Здесь k – индекс дискретной вертикальной позиции в строке j .

Максимумы дают уровни $\mathbf{y}_{j,k}$ с наибольшей плотностью переходов «фон–текст» и служат для определения центров тяжести компонент строки. Порог отсекает шумовые пики и нетекстовые детали.

Для каждого $\mathbf{y}_{j,k}$ вычисляется центр тяжести $x_{j,k} = \text{median}\{x: D(x, \mathbf{y}_{j,k})=1, (x, \mathbf{y}_{j,k}) \in S_j\}$, и формируется прямоугольный фрейм $R_{j,k} = \left[x_{j,k} - \frac{w_j}{2}, x_{j,k} + \frac{w_j}{2} \right] \times \left[y_{j,k} - \frac{h_j}{2}, y_{j,k} + \frac{h_j}{2} \right]$, где $w_j = \frac{W}{N}$, $h_j = \frac{2}{3}w_j$.

Такой фрейм охватывает достаточное число пикселей для надёжной локальной оценки направления, минимально пересекается с соседними полосами и, так как высота по построению меньше ширины, повышает точность оценки вдоль строки:

$$P_{j,k} = \{p \in R_{j,k}: D(p)=1\}.$$

4.3 Оценка локального направления методом главных компонент

Для каждого облака точек $P_{j,k}$ направление строки и её центр тяжести оцениваются методом главных компонент (рис. 4), который робастен к шуму, имеет низкую вычислительную сложность и согласуется с допущением о локальной стационарности наклона, а экспериментально подобранное соотношение сторон фрейма $h_j: w_j=2:3$ обеспечивает надёжность оценки локального направления при сильной кривизне строки. Сначала вычисляется центр масс $\bar{p} = \frac{1}{|P_{j,k}|} \sum_{p \in P_{j,k}} p$, затем ковариационная матрица $S_{j,k} = \frac{1}{|P_{j,k}|} \sum_{p \in P_{j,k}} (p - \bar{p})(p - \bar{p})^T$ и её старший собственный вектор v_1 даёт оценку нормированного локального направления $d(g_{j,k}) = v_1, \|d(g_{j,k})\|=1$.

Пара $(g_{j,k}, d(g_{j,k}))$, где $g_{j,k} = \bar{p}$, формирует узел строковой решётки и локальное направление.

Выделенное множество пар $(G, d) = \{(g_{j,k}, d(g_{j,k}))\}$ является робастным к артефактам письма и шумам и служит основой для дальнейшей реконструкции непрерывных осевых линий.

4.4 Построение осевых линий строк

Для получения линий строк строится граф потенциальных связей $N=(G, E_0)$, где рёбра E_0 соединяют узловые точки. Отбирается подмножество рёбер E_s , согласованных с локальными направлениями поля d , что даёт подграф $N_s=(G_{ext}, E_s)$. Его связные компоненты соответствуют фрагментам осевых линий, а также на нём определены две основные операции:

1. Склейка рёбер, имеющих общие вершины и близкие направления – отбор фрагментов ломаной линии при их соединении концами.
2. Склейка фрагментов – объединение связных цепочек рёбер в строки.

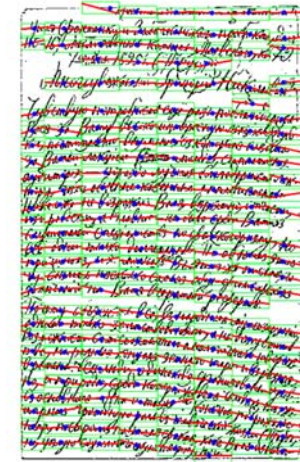


Рис. 4. Пример страницы с выделенными узловыми точками (синие), векторами главных компонент (красные), зеленым обозначены описанные фреймы.

Fig. 4. Example page with detected node points (blue), principal component vectors (red), the corresponding frames highlighted in green.

Чтобы восстановить траекторию каждой строки, к узлам строковой решётки G добавляются четыре вершины для корректного построения модели страницы, совпадающие с углами изображения $C=\{(0,0), (W-1,0), (W-1,H-1), (0,H-1)\}$, что даёт расширенное множество $G_{ext}=G \cup C$. По этому множеству строится триангуляция Делоне (рис. 5), то есть такое разбиение плоскости на треугольники, в описанную окружность каждого из которых не попадает ни одна другая точка G_{ext} . Поскольку узлы G лежат в центрах текстовых кластеров, их ближайшие соседи в этой триангуляции, как правило, принадлежат той же строке, а свойство Делоне минимизирует узкие и длинные треугольники, благодаря чему большинство рёбер ориентировано почти горизонтально, а построение имеет вычислительную сложность $O(|G_{ext}| \log |G_{ext}|)$. Обозначим полный граф триангуляции $N_0=(G_{ext}, E_0)$.

Далее каждое ребро $e=(g, g') \in E_0$ проверяется на согласованность с локальным направлением, определённым в узлах методом главных компонент. Пусть $w=g' - g$ и пусть $d(g), d(g')$ - нормированные векторы, показывающие ориентацию строки в точках g и g' . Тогда $\theta(g) = \arccos \frac{w \cdot d(g)}{\|w\|}$, $\theta(g') = \arccos \frac{w \cdot d(g')}{\|w\|}$.

Мера согласованности рёбер триангуляции и векторного поля оценивается как

$$\delta(e) = \max\{\min(\theta(g), 180^\circ - \theta(g)), \min(\theta(g'), 180^\circ - \theta(g'))\}.$$

Критерий согласованности ребра $\delta(e) \leq \phi_{max}$, где $\phi_{max}=7^\circ$ допускает умеренную кривизну строк. Рёбра, удовлетворяющие этому условию, образуют подграф $N_s=(G_{ext}, E_s)$, $E_s=\{e \in E_0: \delta(e) \leq \phi_{max}\}$. Каждая его связная компонента, как правило, соответствует одной строке. Наконец, рёбра с близкими направлениями, исходящие из общей вершины, склеиваются в ломаные, а соседние ломаные, совпадающие по ориентации и расположенные, соединяются в непрерывные полилинии, которые аппроксимируют интегральные кривые поля направлений d и, таким образом, задают искомые осевые линии строк (рис. 6).



Рис. 5. Триангуляция Делоне узловых точек.
Fig. 5. Delaunay triangulation of nodal points..

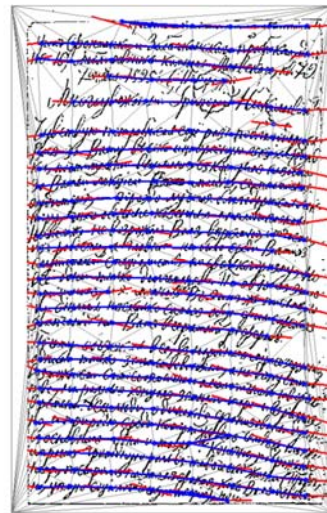
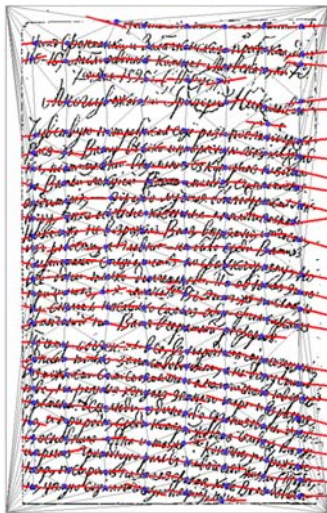


Рис. 6. Пример страницы с построенной триангуляцией (серый) и векторами направлений (красный).
Справа синим выделены ребра, отобранные по критерию согласованности.

Fig. 6. Example page with the constructed Delaunay triangulation (gray) and direction vectors (red).
On the right, the triangulation edges selected by the consistency criterion are highlighted in blue.

4.5 Постобработка осевых линий

После отбора рёбер получен граф, где каждая компонента уже аппроксимирует строку, но могут оставаться два типа локальных артефактов:

- **Разрывы цепочек:** между некоторыми последовательными вершинами отсутствует ребро.

- **Шумовые развилки:** вершина графа имеет степень более двух, образуя неоднозначные ветвления.

Устранение разрывов цепочек. Пусть u и w – правый и левый концы двух ломаных, для которых разница $\|u - w\|$ мала. Ищем путь P^{opt} внутри узкой вертикальной полосы и выбираем

$$P^{opt} = \underset{(p,q) \in P}{\operatorname{argmin}} \sum \left(\lambda \|q - p\| + (1 - \lambda) \left[1 - \frac{(q - p) \cdot d(p)}{\|q - p\|} \right] \right), 0 < \lambda < 1$$

что одновременно минимизирует длину вставки и её отклонение от поля направлений d , тем самым восстанавливая непрерывность линии.

Фильтр шумовых развилок. Для вершины n с $\deg(n) > 2$ оставляем два ребра с максимальным $\frac{(q-n) \cdot d(n)}{\|q-n\|}$, остальные удаляем. Эта фильтрация убирает ложные ответвления, оставляя ребра, наиболее согласованные с локальным направлением.

После обеих процедур фрагменты цепочек превращаются в единую ломаную; упорядочив её вершины по x , получаем правильную осевую линию строки (рис. 7).

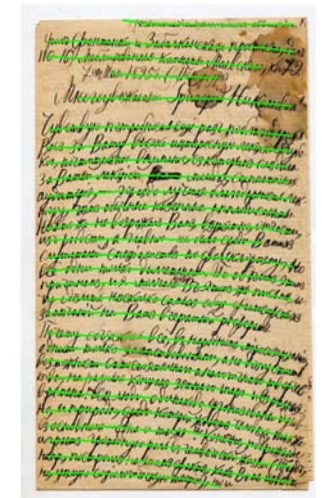
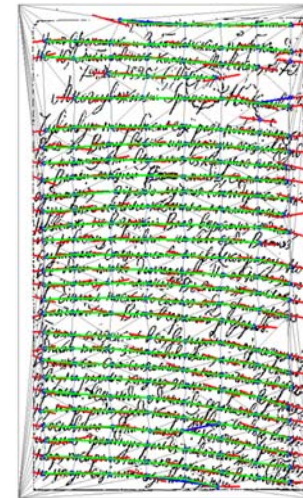


Рис. 7. Постобработка: зелёным – итоговые рёбра, синим – рёбра до фильтрации, красным – локальные направления.

Fig. 7. Post-processing: green – final edges, blue – edges before filtering, red – local directions.

5. Результаты экспериментов

Для тестирования алгоритма была использована коллекция фотографий архивных писем XIX века. Письма написаны разными авторами, имеют разную структуру бумаги, интенсивность чернил, размер написания, на страницах отсутствуют линейки. Всё это помогает оценить разработанный алгоритм на робастность к различным видам данных.

Из полного архива, включающего более 2600 изображений, вручную были размечены и отобраны 120 наиболее сложных для сегментации экземпляров. Эти изображения содержат типичные проблемные случаи: сильно искривлённые строки, малый межстрочный интервал, фоновый шум и нерегулярность написания. Такой выбор обеспечивает тщательную проверку надёжности работы алгоритма на реальных задачах анализа исторических рукописей. Оценка

качества сегментации проводилась по результатам сравнения сегментированных изображений с эталонной разметкой (табл. 1). Введены следующие обозначения:

- **True Positives (TP)** – число эталонных строк, для которых алгоритм выдал хотя бы один фрагмент. Это отражает, что строка найдена, даже если она разбита на части.
- **False Negatives (FN)** – число эталонных строк, для которых алгоритм не выдал ни одного фрагмента, а именно:
 - Полностью пропущенные строки,
 - Строки, «потерянные» в результате слияния нескольких эталонных строк в один фрагмент.
- **False Positives (FP)** – число предсказанных фрагментов, не соответствующих ни одной эталонной строке, а именно:
 - Ложно выделенные фрагменты,
 - Лишние части при дроблении строки на фрагменты,
 - Объединённые (слипшиеся) несколько строк в один фрагмент, который уже не может быть отнесён к отдельной строке.

На основе этих значений определяются два ключевых показателя качества:

$$Precision = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN}$$

Первая оценка (точность) показывает то, насколько предложенный метод вообще способен детектировать строки, а вторая (полнота) – насколько точно он это делает. Также рассмотрена величина F_1 -метрика, которая объединяет точность (Precision) и полноту (Recall):

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Табл. 1. Оценки результатов экспериментов.
Table 1. Results of experiments.

Всего строк	TP	FP	FN	Precision	Recall	F_1
1546	1482	181	64	0.89	0.96	0.92

Предложенный метод обладает высокой полнотой (recall) и точностью (precision). Значение метрики $F_1 = 0,92$ свидетельствует о высоком качестве сегментации, поскольку гармоническое среднее не позволяет одной из компонент доминировать над другой и обеспечивает сбалансированную оценку результатов при обработке случаев, имеющих шумовые артефакты, различную толщину строк и междустрочных интервалов (рис. 8).

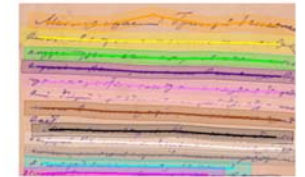
Большинство ошибок, входящих в число **FP**, связано с тем, что метод корректно локализует строки, но внутри одной осевой линии компоненты оказываются разорванными, что создаёт избыточные фрагменты (рис. 9). Ошибки **FN** зачастую обусловлены объединением нескольких истинных строк в один сегмент в тех случаях, где междустрочный интервал сравнительно мал (рис. 10). При этом алгоритм допускает минимальное число полностью ложных выделений и пропусков.

Кроме того, тестовый набор включал наиболее сложные для сегментации страницы (с сильными изгибами строк, нерегулярным междустрочным интервалом и фоновыми шумами), тогда как на документах со стандартной структурой метод работает ещё более робастно.

Большинство ошибок появляются там, где почерк имеет крупный размер, и могут быть связаны с фиксированными размерами фреймов, относительно которых вычисляются локальные направления строк.



Пример корректной сегментации (вариант 1)

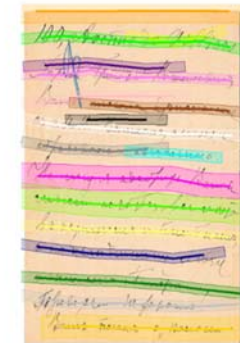


Пример корректной сегментации (вариант 2)

Рис. 8. Все строки сегментированы корректно.
Fig. 8. All lines are segmented correctly

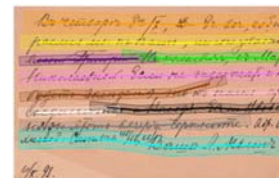


Слипшиеся фрагменты строк: серая линия объединяет две строки, вызывая дополнительные подразделения

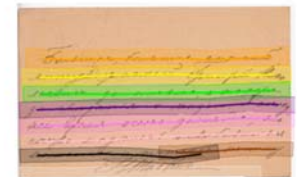


Ложные фрагменты: оранжевая, фиолетовая (2-я строка) и чёрная (4-я) линии

Рис. 9. Ошибки False Positives
Fig. 9. False Positives



Не выделена последняя строка, у предпоследней возможно слияние



Не выделена последняя строка

Рис. 10. Ошибки False Negatives
Fig. 10. False Negatives.

6. Заключение

В работе предложен метод построчной сегментации рукописных документов, основывающийся на анализе локального распределения текстовых пикселей и построении дискретного векторного поля направлений. Для реализации метода был разработан программный комплекс, а также было произведено тестирование алгоритма на обширном архиве изображений исторических писем. Эксперименты подтвердили робастность метода к фоновому шуму, вариативности почерка и искривлённым строкам: достигнуты высокие значения полноты (recall = 0,96) и точности (precision = 0,89), а $F_1 = 0,92$ свидетельствует о сбалансированном сочетании этих показателей – метод последовательно находит все релевантные строки и минимизирует число ложных срабатываний. Дополнительно показано, что алгоритм эффективно аппроксимирует и корректно сегментирует даже сильно изогнутые или наклонённые текстовые строки. Таким образом, основные полученные результаты таковы:

1. Метод построчной сегментации изображений рукописных документов на основе модели линий векторного поля.
2. Метод определения связей узловых точек, согласованных с векторным полем локальных направлений текста.
3. Метод восстановления линий векторного поля как подграфов в триангуляции Делоне.

Основными направлениями дальнейшей работы являются преобразования всех имеющихся фиксированных порогов в адаптивные на основе структуры каждой страницы для более точного её восстановления.

Список литературы / References

- [1]. Ptak R., Żygadło B., Unold O. Projection-based text line segmentation with a variable threshold. *International Journal of Applied Mathematics and Computer Science*, 2017, vol. 27, pp. 195-206. DOI: 10.1515/amcs-2017-0014.
- [2]. Arivazhagan S., Hemant D.J. Statistical projection methods for text line segmentation. In: *Proc. 9th International Conference on Document Analysis and Recognition (ICDAR)*, 2007, pp. 1029-1033.
- [3]. Wong K.Y., Casey R.G., Wahl F.M. Document analysis system. *IBM Journal of Research and Development*, 1982, vol. 26, no. 6, pp. 647-656. DOI: 10.1147/rd.266.0647.
- [4]. Nikolaou N., Makridis M., Gatos B., Stamatopoulos N., Papamarkos N. Segmentation of historical machine-printed documents using adaptive run length smoothing and skeleton segmentation paths. *Image and Vision Computing*, 2010, vol. 28, pp. 590-604.
- [5]. Papavassiliou V., Stafylakis T., Katsouros V., Carayannis G. Handwritten document image segmentation into text lines and words. *Pattern Recognition*, 2010, vol. 43, pp. 369-377.
- [6]. Kennard D.J., Barrett W.A. Separating lines of text in free-form handwritten historical documents. In: *Proc. 2nd International Conference on Document Image Analysis for Libraries (DIAL'06)*, 2006, pp. 12-23.
- [7]. Renton G., Chatelain C., Adam S., Kermorvant C., Paquet T. Handwritten text line segmentation using fully convolutional network. In: *Proc. 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, pp. 321-326.
- [8]. Zhang X., Tan C.L. Text line segmentation for handwritten documents using constrained seam carving. In: *Proc. 14th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2014, pp. 98-103. DOI: 10.1109/ICFHR.2014.24.
- [9]. Louloudis G., Gatos B., Halatsis C. Text line detection in unconstrained handwritten documents using a block-based Hough transform approach. In: *Proc. 9th International Conference on Document Analysis and Recognition (ICDAR)*, 2007, pp. 599-603.
- [10]. Kurar Barakat B., Cohen R., Drobny A., El-Sana J. Learning-Free Text Line Segmentation for Historical Handwritten Documents. *Applied Sciences*, 2020, vol. 10, art. 8276. DOI: 10.3390/app10228276.

- [11]. Leow C.S., Yajima H., Kitagawa T., Nishizaki H. Single-Line Text Detection in Multi-Line Text with Narrow Spacing for Line-Based Character Recognition. *IEICE Transactions on Information and Systems*, 2023, vol. E106-D, no. 12, pp. 2097-2106. DOI: 10.1587/transinf.2023EDP7070.

Информация об авторах / Information about authors

Леонид Моисеевич МЕСТЕЦКИЙ – доктор технических наук, академик РАН, профессор кафедры математических методов прогнозирования МГУ. Научные интересы: вычислительная геометрия, анализ и распознавание изображений.

Leonid Moiseevich MESTETSKIY – Dr. Sci. (Tech.), Russian Academy of Natural Sciences Member, Professor at the Department of Mathematical Forecasting Methods at Moscow State University. His research interests include computational geometry and image analysis and recognition.

Виктория Станиславовна СМІРНОВА является выпускницей бакалавриата кафедры Математических методов прогнозирования факультета ВМК МГУ имени М.В. Ломоносова и магистрантом 1-го года обучения факультета ВМК МГУ. Её научные интересы включают распознавание образов и компьютерное зрение.

Victoria Stanislavovna SMIRNOVA is a graduate of the B.Sc. program at the Department of Mathematical Methods of Forecasting, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, and a first-year M.Sc. student at the same faculty. Her research interests include pattern recognition and computer vision.