

## Распознавание шрифтов по изображениям на основе анализа структуры букв

А.Л. Липкина<sup>1</sup>, Н.А. Ломов<sup>1</sup>, Л.М. Местецкий<sup>1</sup>

lipkina96@mail.ru|nikita-lomov@mail.ru|mestlm@mail.ru

<sup>1</sup>Московский государственный университет имени М.В. Ломоносова, Москва, Россия

*Рассматривается задача определения шрифта и его начертания на цифровых изображениях, таких как сканированные изображения и фотографии. Предложен метод выделения типов начертания шрифта: прямого, наклонного и с засечками. В основе предлагаемого метода лежит контурное и скелетное представление фигуры. Наклон текста на изображении можно определить, основываясь на информации о наклонах рёбер контурного представления текста. Определить наличие засечек на изображённой букве можно по количеству рёбер агрегированного скелета буквы. Этот метод используется для предварительной фильтрации базы шрифтов, в результате которой в качестве кандидатов рассматриваются только шрифты в нужном начертании. Дальнейшим признаковым описанием изображения служит совокупность морфологических моментов – функций, отражающих пространственное распределение ширины отдельных символов. Для идентификации шрифта строится взвешенная композиция метрических классификаторов, связанных с этими функциями. Эксперименты показывают, что предварительное определение начертания, а также учёт геометрических особенностей букв повышают качество распознавания по сравнению с методом на основе морфологической ширины.*

**Ключевые слова:** цифровой шрифт, наклонное начертание, засечки, граничное представление, агрегированный скелет, морфологические моменты.

## Font Recognition from Images Based on Letter Structure Analysis

*We consider the problem of the recognition of the typeface and its style on the digital images, such as scanned images and photographs. A method is proposed for detection of font style: normal, italic and serif. The proposed method is based on the contour and skeleton representations of the figure. The slope of the text in the image can be determined using information about the slopes of the edges of the contour representation of the text. It is possible to determine the presence of serifs on the represented letter by the number of edges of its aggregated skeleton. This method is used to pre-filter the font database, resulting in the fact that only the fonts in the desired style are considered as candidates. A further feature description of the image is the set of morphological moments, i.e. functions, which reflect the spatial distribution of the width of individual symbols. To identify the font, a weighted composition of the metric classifiers associated with these functions is constructed. Experiments show that the preliminary style detection, as well as the consideration of the geometric features of letters, increase the quality of recognition in comparison with the method based on the morphological width.*

**Keywords:** digital font, italic style, serifs, boundary representation, aggregated skeleton, morphological moments.

### 1. Введение

Типографика является неотъемлемой частью графического дизайна. У графических дизайнеров часто возникает желание определить шрифты, с которыми они сталкиваются в реальной жизни, с целью их дальнейшего использования. Хотя они могут сфотографировать образец шрифта, вызвавшего особый интерес, и попросить эксперта определить шрифт, процесс распознавания «вручную» оказывается очень утомительным и подверженным ошибкам. Поэтому существует потребность в методах автоматической идентификации шрифта по изображению, причём нужный шрифт нужно будет определить из тысяч кандидатур, ведь шрифтов, используемых в современной типографике, – великое множество. Тем не менее работ, посвящённых этой задаче и предполагающих взаимодействие с обширной (более 1000 шрифтов) базой, очень немного, последние из них основаны на формировании кодовой книги из векторов локальных признаков, в качестве которых может использоваться HOG [7] или SIFT [8]. Известен также метод, который по аналогии с методом собствен-

ных лиц можно назвать «методом собственных букв» [10]. Самым успешным оказался метод на основе глубинного обучения [11], внедрённый в редактор Adobe Photoshop в виде инструмента Match Font.

Ряд веб-сайтов позволяет пользователям совершать поиск по шрифтовой базе и ранжировать шрифты по степени соответствия запросу. Среди наиболее популярных из таких сервисов – WhatTheFont [2] и Identifont [1]. Хотя алгоритмы, лежащие в основе этих сервисов, не раскрываются, сам характер их взаимодействия с пользователем говорит о том, что принципы их работы кардинально различаются. Сервис WhatTheFont получает на вход картинку с образцом текста, набранного интересующим пользователя шрифтом, и делает его побуквенную разметку: выделяет и распознаёт на изображении отдельные символы. Это позволяет сделать вывод, что сходство шрифтов оценивается на основе сходства их отдельных символов. При работе с сайтом Identifont картинка не требуется – вместо этого пользователю предлагается ответить на ряд вопросов, касающихся как общего облика шрифта (на-

пример, есть ли в нём засечки), так и особенностей отдельных букв (например, какой формы хвостик у буквы Q).

В предыдущих работах [3, 4] был предложен метод распознавания шрифтов, основанный на анализе ширины штрихов в символах шрифта. В качестве признакового описания при этом использовалась диаграмма морфологической ширины, описывающая площадь дискового покрытия фигур в зависимости от размера покрывающих дисков. Общая схема метода выглядит следующим образом:

- выделение и распознавание отдельных букв в сканированном тексте на основе сегментации исходного изображения;
- построение диаграмм морфологической ширины для каждой буквы;
- сравнение диаграмм морфологической ширины букв с диаграммами эталонных описаний шрифтов (метрика сходства букв);
- вычисление обобщенной меры сходства шрифта текста с эталонными описаниями шрифтов (метрика сходства шрифтов);
- выбор наиболее похожих эталонных шрифтов.

Метод показал хорошие результаты, ошибки составили 4–9% в зависимости от качества сканированного текста. Вместе с тем, эксперименты показали, что оценивание сходства шрифтов только лишь на основе морфологической ширины оказывается недостаточным. Имеются примеры шрифтов, которые при общем сходстве диаграмм ширины существенно отличаются с точки зрения человеческого (дизайнерского) представления. Например, похожие распределения ширины могут иметь шрифты, отличающиеся наклоном букв или наличием штрихов-засечек.

В настоящей статье рассматривается подход к уменьшению ошибок такого рода за счёт учёта структурных признаков шрифта: наклона и засечек. Кроме того, вводится понятие морфологических моментов фигуры, обобщающее понятие морфологической ширины и позволяющее использовать не только распределение ширины штрихов, но также учесть более детальное распределение «массы» в символах шрифта. Это существенно расширяет возможности для классификации шрифтов с похожими диаграммами морфологической ширины.

Вычислительный метод, обеспечивающий получение соответствующих признаков, основан на модели непрерывного медиального представления бинарных растровых изображений [5]. Метод развивает идею, описанную в работе [3].

Структура оставшейся части статьи такова. В разделе 2 описан метод классификации шрифтового начертания по типу наклона – прямое либо курсивное. В разделе 3 представлен метод классификации типа шрифта

по наличию штрихов-засечек. В разделе 4 описывается метод построения признакового описания распределения «массы» в символах шрифта, так называемые морфологические моменты. Там же представлена идея вычисления моментов на основе непрерывного медиального представления дискретных изображений. В разделе 5 описаны общие решающие правила классификации шрифта по сводному набору признаков: морфологические моменты, наклон, засечки. В разделе 6 представлены результаты вычислительных экспериментов.

## 2. Классификация текста по типу наклона

На вход методу подаётся незашумлённое бинарное изображение некоторого текста. На всех стадиях работы метода мы будем оперировать непрерывным представлением формы. С точки зрения такого представления бинарное изображение рассматривается как непрерывная сцена – набор фигур, каждая из которых является многоугольником с многоугольными дырами. На этапе определения наклона шрифта мы будем работать со сценой как с единым целым, дальнейшие этапы потребуют разметки текста – указания, какие буквы присутствуют в тексте и из каких фигур состоит каждая буква. Приведём необходимые определения.

*Фигура* – замкнутая область на плоскости, ограниченная конечным числом непересекающихся замкнутых жордановых кривых.

*Контурное представление фигуры* – аппроксимация границ фигуры многоугольниками минимального периметра.

*Пустой круг фигуры* – круг, целиком лежащий в фигуре.

*Вписанный круг фигуры* – пустой круг фигуры, не содержащийся ни в каком другом пустом круге фигуры.

*Медиальное представление формы* – совокупность скелета (множества центров всех вписанных пустых кругов фигуры) и радиальной функции (функции, которая каждой точке скелета ставит в соответствие радиус вписанного круга).

В действительности скелет представляет собой планарный граф (рис. 1). Вершинами графа являются центры вписанных пустых кругов, имеющих либо одну, либо три общих точки с границей фигуры, а рёбрами – линии из центров вписанных пустых кругов, касающихся границы ровно в 2 точках.

Методы аппроксимации формы многоугольниками фигурами и построения скелета подробно изложены в [5].



Рис. 1. Фигура (слева) и её скелет (справа).

2.1. Подготовка данных

Предлагаемый алгоритм классификации опирается только на контурное представление фигуры. Идея этого алгоритма состоит в том, что наклон ребер контурного представления фигуры в среднем достаточно хорошо описывает наклон фигуры.

Каждой связной компоненте бинарного изображения соответствует непрерывная фигура. Для каждой фигуры (чаще всего это вся буква целиком) на непрерывной сцене можно получить явное описание всех её контуров. Совокупность внешнего и внутренних контуров связной компоненты будем называть *полной границей* компоненты.

Каждый контур представляется набором вершин, лежащих в порядке обхода многоугольника по или против часовой стрелки. По соседним вершинам можно получить описание ребра: его длину  $l_i$  и угол наклона  $\alpha_i \in [-\pi, \pi)$  относительно оси абсцисс. Угол наклона для удобства приводится к отрезку  $[0, \pi)$  по формуле:

$$\alpha_i = \begin{cases} \alpha_i, & \alpha_i \in [0, \pi); \\ \alpha_i + \pi, & \alpha_i \in [-\pi, 0). \end{cases}$$

Таким образом, после обработки углов наклона ребер каждая полная граница представляется набором пар  $(l_i, \alpha_i)$ .

2.2. Построение дескриптора наклона контурного представления

Для каждой полной границы на изображении мы построим гистограмму распределения длин ребер по углам наклона из  $n$  равных столбцов. В полученной гистограмме нас будет интересовать значение высоты столбца, отвечающего за угловой интервал, содержащий точку  $\frac{\pi}{2}$ . Так как углы, которые являются равными  $\frac{\pi}{2}$ , на самом деле измеряются с некоторой погрешностью за счёт аппроксимации границы многоугольником, стоит обеспечить, чтобы для малых  $\varepsilon$  углы  $\frac{\pi}{2} - \varepsilon$  и  $\frac{\pi}{2} + \varepsilon$  попали в одну и ту же ячейку гистограммы. Для решения этой проблемы мы будем рассматривать только чётные  $n$  и отобразим углы  $\alpha_i \in [\pi - \frac{\pi}{2n}, \pi) \rightarrow [-\frac{\pi}{2n}, 0)$  за счёт вычитания из них  $\pi$ . Вследствие такого преобразования все углы будут принадлежать отрезку  $[-\frac{\pi}{2n}, \pi - \frac{\pi}{2n})$ , который мы разобьём на  $n$  равных частей при построении гистограммы. Столбец гистограммы, соответствующий отрезку  $[\frac{\pi}{2} - \frac{\pi}{2n}, \frac{\pi}{2} + \frac{\pi}{2n})$ , будем называть показательным. Вклад каждого ребра в гистограмму примем равным его длине. По результатам экспериментов  $n$  полагается равным 32.

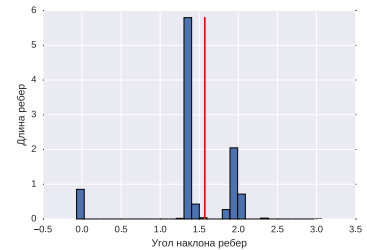
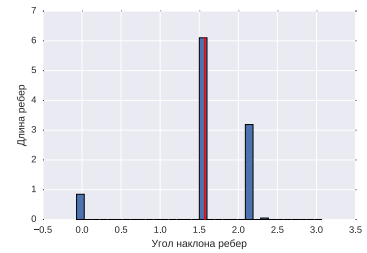


Рис. 2. Гистограмма распределения длине ребер по углам для прямого начертания буквы «N» в шрифте Arial (сверху) и для её наклонного начертания (снизу).

2.3. Определение типа наклона буквы

Найдём показательный столбец и рассмотрим такие столбцы, которые отвечают за углы в окрестности  $U = [\frac{\pi}{4}, \frac{3\pi}{4})$ .

Обозначим за  $U_i^k$  номера таких столбцов гистограммы, которые лежат не более, чем на расстоянии  $k$  столбцов от  $i$ -го столбца, но не считая самого  $i$ -го столбца.

Будем говорить, что столбец является *локальным пиком*, если  $h_i > \gamma \sum_{j \in U_i^k} h_j$  для некоторых фиксированных  $k$  и  $\gamma$ .

В данном исследовании на основании проведённых экспериментов  $\gamma$  берётся равным 0,5.

Определим, есть ли какой-либо пик в окрестности  $U$ . Для этого найдём максимальный столбец из  $U$  (обозначим его номер за  $m$ ) и проверим, является ли он локальным пиком при  $k = 1$ . Если является, и этот столбец совпадает с показательным столбцом, то начертание буквы однозначно классифицируется как прямое. Если он является локальным пиком, и содержит углы, меньшие  $\frac{\pi}{2}$ , то начертание однозначно классифицируется как наклонное. Иначе определить тип начертания невозможно.

Как видно на рис. 2, в верхней части явно выраженный пик гистограммы приходится на значение  $\frac{\pi}{2}$  (красная линия), а в нижней части пик лежит левее красной линии. Это свидетельствует о том, что буква «N» в шрифте Arial является представительной. В то же время, из рис. 3 видно, что у «O» нет каких-либо явно выраженных пиков – это говорит о том, что по этой букве нельзя достоверно определить, является ли она наклонной или нет.

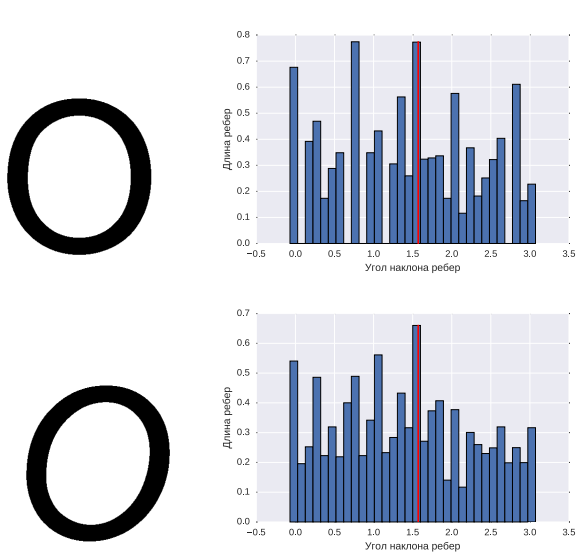


Рис. 3. Гистограмма распределения длин рёбер по углам для прямого начертания буквы «О» в шрифте Arial (сверху) и для её наклонного начертания (снизу).

2.4. Решающее правило

Пусть  $M_r = \frac{i_r}{d_r + i_r}$ , где  $d_r$  – число букв, начертание которых классифицировано как прямое,  $i_r$  – как наклонное. Зададим некоторый порог  $\theta \in [0, 1]$ . Если  $M_r > \theta$ , то говорим, что шрифт имеет наклонное начертание, иначе – прямое. Эксперименты показали, что подходящим является значение  $\theta = 0,4$ .

2.5. Классификация текста по наличию засечек

Введём понятие *агрегированного скелета* – это граф, вершинами которого являются вершины скелета фигуры степени 1 и 3, а рёбрами – цепи, соединяющие вершины степени 1 и 3 и проходящие по вершинам степени 2 (рис. 4). Алгоритм выявления засечек опирается только на агрегированный скелет фигуры. Идея алгоритма состоит в том, что количество рёбер агрегированного скелета фигуры в среднем хорошо описывает наличие в ней засечек.

Скелет связанной фигуры и последующая его стрижка (регуляризация) делаются стандартным способом, описанным в [5]. Экспериментальным путём было выяснено, что оптимальным параметром стрижки скелета является 7% общей высоты фигуры. Число рёбер  $E$  агрегированного скелета можно определить по теореме Эйлера:  $E = V + F - 1$ , где  $V$  – число вершин агрегированного скелета (число вершин исходного скелета степени 1 и 3),  $F$  – число дыр в фигуре (число внутренних контуров). Исключением из этого правила является случай, когда скелет состоит из единственного цикла, что возможно, например, для буквы «О». В этом случае мы по определению считаем, что агрегированный скелет содержит одну вершину и одно ребро-петлю.

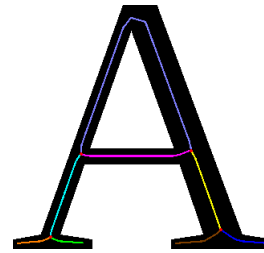


Рис. 4. Выделенные разным цветом рёбра (цепи) агрегированного скелета.

Интуитивно понятно, что у буквы с засечками количество рёбер агрегированного скелета должно быть больше, чем у такой же буквы из той же гарнитуры без засечек. Но порог, по которому можно явно отнести букву к одному из классов, естественно, для каждой буквы свой. В качестве порога для каждой буквы можно взять среднее число рёбер в шрифтах с засечками ( $c_i^{serif}$ ) и без ( $c_i^{sans}$ ):

$$\delta_i = \frac{c_i^{serif} + c_i^{sans}}{2}.$$

Напомним, что на вход этой задаче подается *размеченное* изображение буквы. Это означает, что известно, какая буква представлена на изображении.

Для каждой буквы по обучающей выборке можно определить, какую долю среди шрифтов, у которых число рёбер не больше порогового, составляют шрифты без засечек:

$$p(s = sans | E_i \leq \delta_i) = \frac{\sum_{j: E_i^j \leq \delta_i} [s_j = sans]}{\sum_{j: E_i^j \leq \delta_i} 1},$$

где  $s_j$  – стиль  $j$ -го шрифта,  $E_i^j$  – количество рёбер в агрегированном скелете его  $i$ -й буквы. Аналогичным образом можно рассчитать величины  $p(s = serif | E_i \leq \delta_i)$ ,  $p(s = sans | E_i > \delta_i)$ ,  $p(s = serif | E_i > \delta_i)$ .

Тогда, если

$$\prod p(s = serif | E_k \leq \delta_k)^{[E_k \leq \delta_k]} p(s = serif | E_k > \delta_k)^{[E_k > \delta_k]} > \prod p(s = sans | E_k \leq \delta_k)^{[E_k \leq \delta_k]} p(s = sans | E_k > \delta_k)^{[E_k > \delta_k]},$$

то начертание надписи классифицируется как начертание с засечками (индекс  $k$  пробегает только номера букв, присутствующих на изображении), иначе – как начертание без засечек.

Заметим, что и в этом случае для надёжной классификации необходимо присутствие в слове «представительных» букв, вид которых в шрифтах с засечками и без значительно отличается.

### 3. Идентификация шрифта по морфологическим моментам

#### 3.1. Понятие морфологических моментов

Морфологические моменты развивают концепцию морфологической ширины фигуры, введённую в [3]. Напомним необходимые определения:

*Дисковое  $r$ -покрытие фигуры* – объединение всех пустых кругов радиуса  $r$ .

*$r$ -площадь фигуры* – площадь её дискового  $r$ -покрытия.

*Морфологическая ширина фигуры* – это  $r$ -площадь фигуры как функция от  $r$ .

Идея морфологических моментов заключается в том, что вычисление площади фигуры можно рассматривать как взятие интеграла от функции  $f(x, y) = 1$  по рассматриваемой фигуре. В случае моментов рассматривают функции более общего вида:  $f(x, y) = x^p y^q$ . В результате моменты позволяют описать не только то, как меняется «масса» покрытия с ростом радиуса, но и то, как меняется её пространственное распределение.

*Морфологическим моментом* порядков  $(p, q)$  фигуры  $X$  называется функция, описывающая зависимость геометрического момента, вычисленного по дисковому  $r$ -покрытию, от размера примитива:

$$\mathfrak{M}_{pq}(X, r) = \iint_{(x,y) \in X_r} x^p y^q dx dy,$$

где  $X_r$  – дисковое  $r$ -покрытие фигуры  $X$ .

Нетрудно заметить, что в таком виде морфологические моменты оказываются чувствительными к выбору системы координат. Чтобы обеспечить их инвариантность к параллельному переносу фигуры, моменты можно центрировать, поместив начало координат в центр масс фигуры:

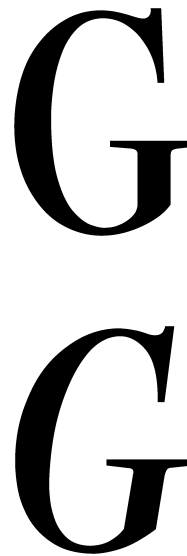
$$\mathfrak{M}_{pq}^*(X, r) = \iint_{(x,y) \in X_r} (x - \bar{x})^p (y - \bar{y})^q dx dy,$$

$(\bar{x}, \bar{y})$  – координаты центра масс, которые находятся по формулам

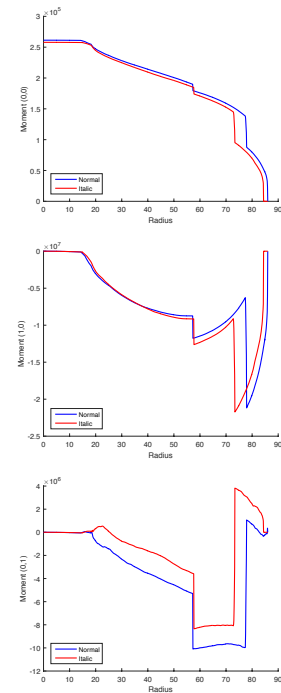
$$\bar{x} = \mathfrak{M}_{10}(X, 0) / \mathfrak{M}_{00}(X, 0), \bar{y} = \mathfrak{M}_{01}(X, 0) / \mathfrak{M}_{00}(X, 0).$$

При этом центральные моменты по-прежнему остаются чувствительными к масштабированию и повороту фигуры. Пользуясь особенностями задачи, эту проблему можно решить следующим образом: установим систему координат, в которой начало совпадает с центром масс буквы, ось  $OX$  параллельна базовой линии шрифта, а масштаб выбран так, что высота заглавной буквы шрифта оказывается равной фиксированному значению.

Как показывает верхний график (рис. 5), ширинные характеристики символа часто не меняются при смене начертания с прямого на курсивное. Однако эти символы могут быть различимы по моментам более высоких порядков (два нижних графика).



(a)



(b)

Рис. 5. Прямое и курсивное начертание символа (a). Графики моментов нулевого и первого порядка (b); синий – прямое, красный – курсивное.

#### 3.2. Процедура вычисления

Как и морфологическую ширину, морфологические моменты можно рассчитать с высокой точностью в результате высокоэффективной вычислительной процедуры, использующей медиальное представление формы. Дисковое  $r$ -покрытие представляет собой геометрическую фигуру (или совокупность фигур), ограниченную отрезками прямых и выпуклыми дугами окружностей. В ходе работы алгоритма [3] дисковое покрытие получает явное представление в виде объединения и исключения простых геометрических фигур (рис. 6), среди которых:

- собственные области рёбер скелета (выделены светло-серым);
- внешние сектора меньших концевых кругов усечённых рёбер (тёмно-серый);
- линзы в пересечении внешних секторов (красный).

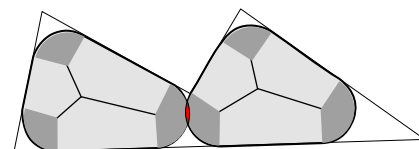


Рис. 6. Составные части дискового покрытия многоугольной фигуры.

Площадь покрытия складывается из суммы площадей собственных областей и секторов за вычетом суммы площадей линз. Но ничто не мешает нам сохранить общий вид алгоритма, заменив на третьем этапе вычисление *площадей* составляющих фигуру областей на вычисление их моментов.

Любая область первого типа представляет собой простой многоугольник, второго – круговой сектор, третьего – объединение двух областей, каждая из которых – это дополнение треугольника до кругового сектора. Таким образом, нам достаточно уметь интегрировать функцию  $f(x, y) = x^p y^q$  по многоугольнику и круговому сектору. Это можно сделать аналитически, точные формулы выведены в работе [9].

### 3.3. Задание метрики

Для сравнения шрифтов мы вычислим диаграммы моментов – значения функций для аргументов  $r_i$ , взятых с определённым шагом. Обозначим код  $j$ -го символа на изображении  $c_j$ , его диаграммы моментов –  $t_{pq}^j$ , эталонные моменты символа  $i$ -го шрифта с кодом  $j$  –  $f_{pq}^{ji}$ . Расстоянием от тестового образца  $\mathbf{t} = \{t_1, \dots, t_n\}$  до шрифта  $\mathbf{f}_k$  будем считать взвешенную сумму расстояний до эталонов по моментам отдельных символов:

$$d(\mathbf{f}_k, \mathbf{t}) = \sum_{pq} \alpha_{pq} \sum_{j=1}^n \beta_j \|f_{pq}^{c_j k}, t_{pq}^j\|_{L_1}. \quad (1)$$

Здесь  $\alpha_{pq}$  – вес момента порядка  $(p, q)$ , а  $\beta_j$  – вес  $j$ -й буквы. В результате распознавания имеющиеся в базе шрифты ранжируются по мере сходства с представленным образцом, т.е. в порядке убывания расстояния, что позволяет получить как количественную (расстояния), так и качественную (номер в упорядоченном списке) оценку сходства.

## 4. Построение классификатора

Пусть  $\mathbf{t} = (t_1, t_2, \dots, t_n)$  – множество символов, выделенных в исходном тексте на этапе сегментации. В это множество входят все выделенные символы, в том числе это могут быть изображения одних и тех же букв. Их растровые представления могут отличаться из-за особенностей сканирования, бинаризации и т.п. Обозначим

$C(\mathbf{t})$  – результат классификации типа наклона текста  $\mathbf{t}$ : для прямого шрифта  $C(\mathbf{t}) = 0$ , а для наклонного  $C(\mathbf{t}) = 1$ .

$S(\mathbf{t})$  – результат классификации текста по наличию штрихов-засечек: если  $S(\mathbf{t}) = 0$ , то засечек нет, а если  $S(\mathbf{t}) = 1$ , то есть.

$F$  – библиотека эталонных описаний шрифтов, содержащая  $K$  описаний. В ней для каждого шрифта  $\mathbf{f}_i$ ,  $i = 1, \dots, K$  заданы априори известные признаки наклона и засечек  $C_i$  и  $S_i$  соответственно.

Для всех шрифтов выделен набор символов, которые могут быть использованы для распознавания шрифтов. Этот набор из  $M$  символов включает прописные и строчные буквы латиницы и кириллицы, а также

цифры. В каждом шрифте этот набор символов представлен своим набором глифов  $Z_i = z_j^{(i)}$ ,  $j = 1, \dots, M$ .

Эталонное описание шрифта включает в себя набор из  $\frac{(R+1)(R+2)}{2}$  функций моментов  $\{f_{pq}^{ji}\}$ ,  $p, q \geq 0$ ,  $p+q \leq R$  для каждого символа  $j = 1, \dots, M$  каждого шрифта  $i = 1, \dots, K$ . Число  $R$  может быть небольшим, достаточно рассмотреть  $R = 1$  – это означает, что мы принимаем во внимание только площадь и центр масс дискового покрытия символа.

Общая схема работы классификатора для текста  $\mathbf{t}$  выглядит так.

**Шаг 1.** Вычисление признаков наклона  $C(\mathbf{t})$  и засечек  $S(\mathbf{t})$  для текста  $\mathbf{t}$ .

**Шаг 2.** Отбор подмножества эталонных шрифтов  $F_{\mathbf{t}}$  из библиотеки  $F$ , у которых  $C_i = C(\mathbf{t})$  и  $S_i = S(\mathbf{t})$ .

**Шаг 3.** Вычисление сходства фрагмента текста  $\mathbf{t}$  с эталонными шрифтами из  $F_{\mathbf{t}}$ .

**Шаг 4.** Отбор  $top(\mathbf{t})$  – множества эталонных шрифтов, наиболее близких к шрифту текста  $\mathbf{t}$ . Размер списка  $top(\mathbf{t})$  определяется, исходя из потребностей пользователя. По экспертной оценке шрифтовых дизайнеров этот список может включать 5–10 наиболее похожих шрифтов.

## 5. Эксперименты

Для проведения экспериментов была расширена база данных, уже использованная в работе [4]. Новая база содержит изображения 52 символов латиницы (26 строчных и 26 прописных букв) из 2555 шрифтовых начертаний коллекции цифровых шрифтов компании Паратайп [6]. Ранее коллекция содержала 1884 шрифта. Для экспериментов по идентификации начертания использовались только такие шрифты, начертание которых можно установить однозначно, – в частности, декоративные шрифты были исключены из коллекции.

Для каждого шрифта из букв в низком разрешении были получены изображения 1000 наиболее популярных английских слов (средняя длина слова составила 5,33 символа, в наборе были и очень короткие слова, например,  $\mathbb{1}$ ,  $\mathbb{b}\epsilon$ ), случайные 40% из которых были переведены в верхний регистр. Эти изображения и их сегментация использовались как запросы для идентификации шрифта.

### 5.1. Классификация по типу наклона

Для проведения эксперимента из коллекции было отобрано 1760 шрифтов: 967 в прямом начертании и 793 в наклонном. Размер шрифта был подобран так, чтобы высота заглавной буквы  $\mathbb{H}$  составила 50 пикселей.

В результате применения предложенного метода точность классификации по слову составила 97,14% для прямого начертания и 92,71% – для наклонного.

Матрица ошибок представлена в таблице 1:

	Прямое	Наклонное	Не определено
Прямое	939 338	26 242	1 420
Наклонное	57 380	735 202	418

**Таблица 1.** Матрица ошибок при распознавании наклонного начертания по буквам из слова.

Далее для этих же шрифтов в прямом и наклонном начертании были сгенерированы картинки с текстом «the quick brown fox jumps over the lazy dog». Эта фраза примечательна тем, что в ней содержатся все буквы латинского алфавита. Высота заглавной буквы Н снова составила 50 пикселей.

В результате применения предложенного метода точность классификации прямого начертания составила 97,72%, а наклонного – 98,49%. Примечательно, что, так как в тексте заведомо присутствуют буквы, у которых можно определить тип наклона, метод больше не выдаёт отказы от классификации.

Более подробные результаты представлены в нижеприведённой матрице ошибок:

	Прямое	Наклонное
Прямое	945	22
Наклонное	12	781

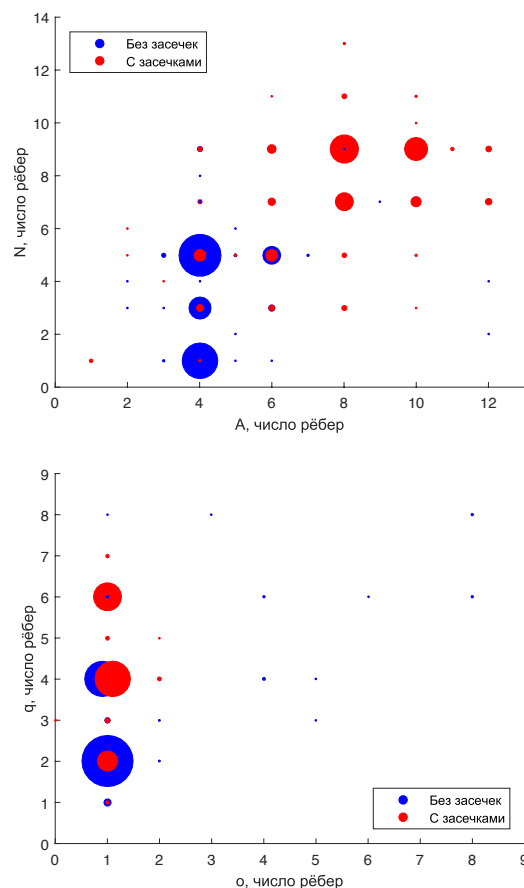
**Таблица 2.** Матрица ошибок при распознавании наклонного начертания по всем буквам латинского алфавита.

## 5.2. Классификация по наличию засечек

Для проведения эксперимента из коллекции был отобран 2101 шрифт: 1285 с засечками и 816 без засечек, по которым был определён порог  $\delta_i$  для каждой буквы. Высота заглавной буквы при растеризации символов составила 50 пикселей.

Результаты показали, что у некоторых букв (например, «А», «Н») показатели среднего количества рёбер для начертания с засечками и без засечек сильно различаются, а у некоторых (например, «о», «d») – практически не отличаются (рис. 7). Понятно, почему так получается: начертания каких-то букв с засечками и без засечек могут отличаться незначительно (например, у буквы «о» вообще нет засечек в обоих случаях).

В ходе проведённого эксперимента для тех букв, где отличие между средним количеством рёбер для начертаний с засечками и без достаточно большое, точность классификации составила примерно 95%. Некоторые буквы оказались совершенно не информативными для алгоритма. Средняя точность классификации по одной букве превысила 80%, а по одному слову составила 94,8%. Можно сделать вывод: достаточно наличия в тексте одной характерной буквы, чтобы определить, имеет ли шрифт засечки.



**Рис. 7.** Распределение числа рёбер агрегированного скелета в шрифтах с засечками и без для букв «А» и «N» (сверху), «о» и «q» (снизу). Площадь каждой отметки прямо пропорциональна количеству шрифтов с данной комбинацией числа рёбер.

## 5.3. Определение шрифта по морфологическим моментам

Для эталонных изображений были получены диаграммы морфологических моментов методом, описанным в статье [9]. Для этого каждый символ был отрисован в виде бинарного растрового изображения в таком масштабе, чтобы высота прописной буквы Н составила 1000 пикселей. По этим изображениям были построены непрерывные скелеты методом, описанным в [5], и на их основе вычислены центральные морфологические моменты нулевого и первого порядков с шагом радиуса в 0,5 пикселя.

Для этих же шрифтов были получены изображения символов в более низком разрешении, так чтобы высота буквы Н составила 100 и 50 пикселей. Для этих символов также были построены диаграммы моментов с тем же шагом радиуса. При этом все координаты в скелетно-граничном представлении предварительно были увеличены в 10 раз для размера 100 и в 20 раз для размера 50, так что, хотя качество аппроксимации формы было значительно хуже эталонного, масштаб сим-

волов был тем же, и диаграммы моментов можно было сравнивать.

При сравнении шрифтов вес момента нулевого порядка был принят равным 1, моментов первого порядка – 0,01. Такая разница объясняется тем, что с ростом порядка момента значения в диаграммах резко возрастают. Вес каждой буквы был взят пропорциональным качеству классификации, которое даёт распознавание по одной этой букве.

Результаты эксперимента показали, что средний уровень правильного распознавания шрифта по слову при использовании только момента нулевого порядка (т.е. морфологической ширины) составил 89,3% при размере 100 и 63,3% при размере 50, добавление центральных моментов первого порядка позволило повысить точность распознавания до 94,1% и 84,6% соответственно.

Таким образом, проведенный эксперимент подтвердил работоспособность предложенного метода и показал его эффективность на практической задаче сравнения большого количества изображений ( $2555 \times 2555 \times 52$ ) с весьма высоким качеством распознавания.

## 6. Заключение

В данной работе были предложены методы классификации начертаний изображений текста: по прямому начертанию/наклонному начертанию/невозможности определить начертание, а также по наличию засечек у буквы. Для идентификации шрифта предложен метод на основе морфологических моментов, обобщающий понятие морфологической ширины.

Алгоритмы для распознавания начертания показывают хорошие результаты на изображениях отдельных слов, особенно, если на них присутствуют буквы, сильно меняющие облик в зависимости от начертания. Применяя эти алгоритмы, можно успешно отфильтровать исходную выборку и исключить из рассмотрения шрифты, начертание которых отличается от установленного.

Экспериментально показано, что добавление моментов первого порядка к диаграммам морфологической ширины повышает качество распознавания. При этом требования к качеству входных данных достаточно невысоки, что позволяет работать с изображениями, полученными в результате обычного сканирования.

## 7. Благодарности

Работа выполнена при поддержке РФФИ (грант № 17-01-00917).

## 8. Литература

- [1] URL: [www.identifont.com](http://www.identifont.com). – Identifont – 2000–2017.  
 [2] URL: [www.myfonts.com/WhatTheFont](http://www.myfonts.com/WhatTheFont). – WhatTheFont! – 1999–2017.

- [3] Ломов Н.А., Местецкий Л.М. Площадь дискового покрытия – дескриптор формы изображения // Компьютерная оптика, 40(4): 516–525, 2016.  
 [4] Ломов Н.А., Местецкий Л.М. Идентификация цифровых шрифтов на основе морфологической ширины // GraphiCon-2016, Нижний Новгород, с. 89–94.  
 [5] Местецкий Л.М. Непрерывная морфология бинарных изображений: фигуры, скелеты, циркуляры // М.: ФИЗМАТЛИТ, 2009.  
 [6] Якупов Э., Петрова И., Фридман Г., Королькова А., Левин Б. 2008–2014. ПАРАТАЙП. Цифровые шрифты. Каталог // Москва: ParaType, 2014.  
 [7] Bui T., Collomosse J. Font finder: Visual recognition of typeface in printed documents // Conference on Image Processing (ICIP), 2015 – pp. 3926-3930.  
 [8] Chen G., Yang J., Jin H., Brandt J., Shechtman E., Agarwala A., Han T. Large-Scale Visual Font Recognition // Conference on Computer Vision and Pattern Recognition, 2014. – pp. 3598–3605.  
 [9] Lomov N., Sidyakin S. Morphological Moments of Binary Images // ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W4, 2017. – pp. 19–25.  
 [10] Solli M., Lenz R. FyFont: Find-your-Font in Large Font Databases // Proceedings of the 15th Scandinavian conference on Image analysis, 2007. – pp. 432–441.  
 [11] Wang, Z., Yang, J., Jin H., Shechtman E., Agarwala A., Brandt J. Huang, T. S. DeepFont: Identify Your Font from An Image // Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 2015. – pp. 451–459.