# Определение угла наклона текста на сканированных изображениях

Валерий В. ДмитриевУфа, Россия
ufabiz@gmail.com

**Аннотация**

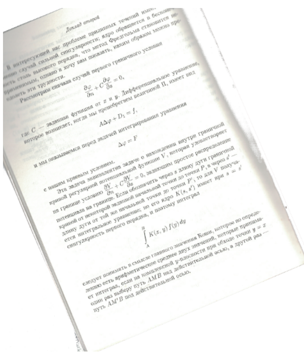
При оптическом распознавании текста на сканированных документах, качество распознавания зависит от того, наклонён ли текст в документе. У выровненных документов качество распознавания заметно лучше. Соответственно, возникает практическая необходимость в средствах автоматического выравнивания угла наклона текста.

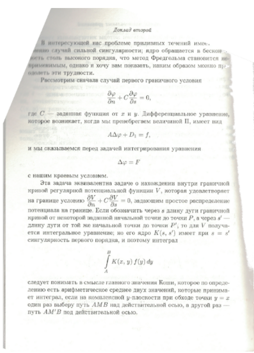
В статье предлагается простой, универсальный и достаточно эффективный алгоритм выравнивания наклона текста, основанный на идее минимизации средней энтропии строк и столбцов растрового изображения.

## 1 Идея

Базовая идея алгоритма состоит в том, что при повороте текста на сканированном изображении, средняя, по строкам и столбцам, энтропия распределения пикселей должна возрасти.

Предположим, нам дан чёрно-белый скан изображения. То есть, каждый пиксель может принимать только два значения: 0 или 1. Как известно, энтропия равномерного распределения максимальна. Если изображение повёрнуто, то в среднем, распределение чёрных и белых пикселей по строкам (и столбцам) будет ближе к равномерному, чем у неповёрнутого изображения. У выровненного изображения распределение пикселей в среднем должно быть менее равномерным.





*Доклад второй*

В интересующей нас проблеме приливных течений име... нению случай сильной сингулярности; ядро обращается в бескон... ость столь высокого порядка, что метод Фредгольма становится не... применимым, однако я хочу вам показать, каким образом можно пр... одолеть эти трудности.

Рассмотрим сначала случай первого граничного условия

$$\frac{\partial\varphi}{\partial n} + C\frac{\partial\varphi}{\partial s} = 0,$$

где $C$ — заданная функция от $x$ и $y$. Дифференциальное уравнение, которое возникает, когда мы пренебрегаем величиной П, имеет вид

$$A\Delta\varphi + D_1 = f,$$

и мы оказываемся перед задачей интегрирования уравнения

$$\Delta\varphi = F$$

с нашим краевым условием.

Эта задача эквивалентна задаче о нахождении внутри граничной кривой регулярной потенциальной функции $V$, которая удовлетворяет на границе условию $\frac{\partial V}{\partial n} + C\frac{\partial V}{\partial s} = 0$, задающим простое распределение потенциала на границе. Если обозначить через $s$ длину дуги граничной кривой от некоторой заданной начальной точки до точки $P$, а через $s'$ — длину дуги от той же начальной точки до точки $P'$, то для $V$ получается интегральное уравнение; но его ядро $K(s, s')$ имеет при $s = s'$ сингулярность первого порядка, и поэтому интеграл

$$\int_A^B K(x, y)\, f(y)\, dy$$

следует понимать в смысле главного значения Коши, которое по определению есть арифметическое среднее двух значений, которые принимает интеграл, если на комплексной $y$-плоскости при обходе точки $y = x$ один раз выберу путь $AMB$ над действительной осью, а другой раз — путь $AM'B$ под действительной осью.

Гипотеза состоит в том, чтобы вычислить среднюю по строкам и столбцам энтропию распределения пикселей для разных углов поворота. И найти такой угол, при котором эта усреднённая энтропия примет минимальное значение.

Для проверки предложенной гипотезы в сети Интернет был собран набор данных различных видов сканированных изображений. После чего предположения были проверены экспериментально.

Предложенный подход работает и позволяет абсолютно точно определить угол поворота в 83 % случаев и с точностью до 1° в 98 % случаев.

Хотя, на первый взгляд, энтропия Шеннона хорошо подходит для этой задачи, было бы разумно не ограничиваться только ей, а рассмотреть весь спектр энтропий Реньи. И с учётом полученных результатов, а также вычислительной сложности, выбрать оптимальное значение параметра энтропии Реньи.

Энтропия Реньи вычисляется с помощью следующего выражения [1]:

$$R_\alpha = \frac{1}{1-\alpha} \log \left( \sum_{i=1}^{n} p_i^\alpha \right), \tag{1}$$

где $p_i$ – вероятности, соответствующие распределению. В нашем случае это частоты чёрных и белых пикселей.

В случае $\alpha = 1$, выражение (1) превращается в энтропию Шеннона:

$$R_1 = H = -\sum_{i=1}^{n} p_i \log(p_i). \tag{2}$$

## 2 Эксперимент

Для проведения эксперимента был собран набор различных документов в сети Интернет. Каждое изображение из набора было повёрнуто на случайный угол в интервале от -45° до 45° и, затем, был вычислен угол поворота с помощью предложенного алгоритма.

В таблице ниже представлены результаты для различных значений параметра энтропии Реньи $\alpha$:

| Параметр энтропии Реньи $\alpha$ | 1/8 | 1/4, | 1/2 | 3/4 | 1 | 2 | 5 |
|---|---|---|---|---|---|---|---|
| Среднее абсолютное отклонение | 0.498 | 0.299 | **0.211** | 0.283 | 0.240 | 5.827 | 41.099 |
| Доля полных совпадений (точность) | 0.822 | **0.834** | 0.828 | 0.815 | 0.805 | 0.641 | 0.009 |
| Доля совпадений с точностью до 1° | 0.942 | 0.969 | **0.980** | **0.982** | **0.983** | 0.822 | 0.015 |
| Доля совпадений с точностью до 2° | 0.967 | 0.983 | **0.991** | **0.993** | **0.994** | 0.841 | 0.015 |

Всего было обработано 1665 документов.

Из таблицы можно сделать вывод, что наименьшее среднее абсолютное отклонение достигается при $\alpha = \frac{1}{2}$.

Наилучшая точность (доля полных совпадений) достигается при $\alpha = \frac{1}{4}$.

Наилучшая приемлемая точность, как доля совпадений с точностью до 1°, достигается при $\alpha \in \{1, \frac{3}{4}, \frac{1}{2}\}$. И наилучшая доля совпадений с точностью до 2° достигается также при $\alpha \in \{1, \frac{3}{4}, \frac{1}{2}\}$.

Если рассматривать предложенную методику, как часть комплекса оптического распознавания документов, то наилучшим значением оказывается $\alpha = \frac{1}{2}$.

При $\alpha = \frac{1}{2}$, среднее абсолютное отклонение составит всего 0.211°. При этом достигается оптимальная приемлемая доля совпадений с точностью до 1°.

Для выбора $\alpha = \frac{1}{2}$ есть ещё одна причина: при этом значении достигается оптимальная вычислительная сложность.

Ниже представлены результаты бенчмарка многократного вычисления энтропий (1) для различных значений параметра $\alpha$ [3].

Бэнчмарк вычислялся для 1000000000 примеров [1].

| $\alpha$ | nanoseconds | miliseconds | % of Shannon |
|---|---|---|---|
| 1 | 10249895206 | 10249 | 100 |
| 1/2 | 8677368472 | 8677 | 84.66 |
| 1/4 | 10421639934 | 10421 | 120.1 |
| 1/8 | 13235709810 | 13235 | 127 |
| 3/4 | 11403406522 | 11403 | 86.16 |
| 2 | 7245386547 | 7245 | 63.54 |
| 5 | 7771674801 | 7771 | 107.26 |
| 10 | 10809162384 | 10809 | 139.08 |

Из таблицы мы видим, что из подходящих нам значений $\alpha$, наилучшая производительность достигается при $\alpha = 1/2$, что вполне ожидаемо.

## 3 Алгоритм

**Замечание.** Предлагаемый ниже алгоритм [2] предполагает, что для определения угла поворота мы используем бинарное чёрно-белое растровое изображение, в котором каждый пиксель может принимать два значения: 0 или 1.

Таким образом, для применения алгоритма необходимо, в первую очередь, получить бинаризованную копию изображения.

Я реализовал алгоритм с применением библиотеки *libleptonica* для манипуляции растровыми изображениями, т.к. эта библиотека используется в TesseractOCR [2].

Для этого я использовал последовательное преобразование pixContrastTRC с $contrast\_factor = 1.0$ и затем pixConvertTo1 с $threshold = 170$.

Пусть $h$ высота и $w$ ширина исходного изображения.
И пусть $d = \sqrt{w^2 + h^2}$ – длина диагонали.

Будем поворачивать изображение на угол $\phi$ относительно центра изображения и считать среднюю энтропию по строкам и столбцам.

Чтобы в результате поворота не выйти за границы изображения, мысленно расширим полотно до размеров с высотой и шириной равной $d$.

Определим целые числа $x_{from}$, $x_{to}$, $y_{from}$ и $y_{to}$ как:

$$\begin{aligned}
x_{from} &= \frac{d}{2} - \frac{h|\sin(\phi)| + w|\cos(\phi)|}{2}, \\
x_{to} &= d - x_{from}, \\
y_{from} &= \frac{d}{2} - \frac{h|\cos(\phi)| + w|\sin(\phi)|}{2}, \\
y_{to} &= d - y_{from},
\end{aligned}$$

где $x_{from}$ и $x_{to}$, соответственно, определяют левую и правую границы повёрнутого изображения.

А $y_{from}$ и $y_{to}$, соответственно, определяют верхнюю и нижнюю границы повёрнутого изображения.

Нужно посчитать среднюю энтропию по строкам и по столбцам. Для примера покажем, как считать по строкам. Расчёт по столбцам аналогичен.

Пусть $V(x, y)$ – цвет точки в пикселе с координатами $(x, y)$ (принимает значения $\{0, 1\}$).

[1] Исходный код бенчмарка можно найти на странице
`https://gist.github.com/valmat/6a737cc3783449c4f7a829e77c77393e`

[2] Исходный код алгоритма: `https://github.com/valmat/rotate_detection`

$R(\{p,q\})$ – энтропия распределения $\{p,q\}$.

Приведённый ниже алгоритм рассчитывает среднюю энтропию $S_\phi$ растрового бинарного изображения, повёрнутого на угол $\phi$.

**Algorithm 1** Рассчёт средней энтропии строк

```
S_φ = 0                                              ▷ Средняя энтропия по строкам.
y = y_from
while y < y_to do
    b = 0                                            ▷ Количество чёрных пикселей в строке.
    x = x_from
    while x < x_to do
        x̃ = x − d/2                                  ▷ Устанавливаем начало координат
        ỹ = y − d/2                                  ▷ в центр полотна.
        x′ = x̃ · cos(φ) − ỹ · sin(φ) + w/2           ▷ Поворачиваем полотно на угол φ.
        y′ = x̃ · sin(φ) + ỹ · cos(φ) + h/2
        if x′ ≥ 0 AND x′ < w AND y′ ≥ 0 AND y′ < h then
            b = b + V(x′, y′);                       ▷ Считаем чёрные пиксели.
        end if
        x = x + 1
    end while
    p = b/d, q = 1 − p
    S_φ = S_φ + R({p,q})/d
    y = y + 1
end while
```

Полученное в результате значение $S_\phi$ и будет искомой средней энтропией для угла поворота $\phi$.

Для нахождения искомого угла поворота $\phi_0$ нужно найти минимум $S_\phi$:

$$\phi_0 = -\arg\min_\phi S_\phi$$

Знак “$-$” берётся, потому-что для выравнивания изображения, нужно повернуть его в обратную сторону.

## Список литературы


[1] WikiPedia “Энтропия Реньи” https://ru.wikipedia.org/wiki/Энтропия_Реньи

[2] Исходный код алгоритма на C++ https://github.com/valmat/rotate_detection

[3] Исходный код бенчмарка производительности энтропий на C++ https://gist.github.com/valmat/6a737cc3783449c4f7a829e77c77393e