

Построение быстрых вычислительных схем настройки алгоритма бинаризации Ниблэка*

М. А. АЛИЕВ, Д. П. НИКОЛАЕВ, А. А. САРАЕВ

Аннотация. В работе предложены вычислительно эффективные методы автоматической настройки параметров двухпараметрической модификации алгоритма бинаризации Ниблэка. Рассмотрены их преимущества и недостатки. Основными инструментами, использованными в данной работе, являются интегрирование изображения и быстрое преобразование Хафа. Рассмотрен ряд особенностей реализации быстрого преобразования Хафа.

Ключевые слова: обработка изображений, бинаризация, оценка качества изображений, OCR.

Введение

Многие системы обработки изображений в своей работе в качестве одного из этапов используют бинаризацию (binarization, thresholding) — процесс преобразования полутонового или цветного изображения в бинарное. Данная техника получила широкое распространение в области распознавания документов: она является важной частью многих популярных OCR систем [1, 2, 3], применяется для выделения графических примитивов [4]. Примерами использования бинаризации также служат анализ медицинских снимков [5], неразрушающий контроль [6].

Качество работы алгоритмов бинаризации зачастую сильно зависит от выбранных параметров. Определение их значений экспертом часто представляется излишне трудоёмким (особенно в случае большого количества данных), поэтому процесс нахождения оптимальных параметров желательно автоматизировать.

В работе [7] был предложен метод настройки алгоритма Ниблэка [8] (точнее, его небольшой модификации [9]). Предложенный способ основывался на идее минимизации невязки для ключевых фрагментов изображений обучающей выборки и представлял собой фактически перебор в пространстве параметров. Подробно был рассмотрен вопрос выбора разумного критерия оценки качества, на основе которого была выбрана оптимизируемая функция ошибки. Однако в упомянутой статье не были рассмотрены методы, позволяющие ускорить процесс

настройки. Данная работа призвана устранить указанный недостаток.

1. Основные определения

Рассмотрим кратко необходимые в дальнейшем понятия. Более подробное объяснение можно найти в [7].

Пусть G — множество полутоновых изображений, B — множество бинарных изображений. Тогда, если p_B это параметры бинаризации, то алгоритм бинаризации можно выразить как отображение:

$$\langle G, P_B \rangle \rightarrow B. \quad (1)$$

Пусть $I \in G$, $A \in B$. Тогда бинаризация сводится к следующему правилу:

$$A(x, y) = \begin{cases} 0, & \text{если } I(x, y) \leq T(x, y) \\ 1, & \text{если } I(x, y) > T(x, y) \end{cases}, \quad (2)$$

где $T(x, y) \in R$ — порог в соответствующей точке изображения.

В данной работе исследуется модификация алгоритма Ниблэка [9]:

$$T(x, y) = \mu(x, y) + k\sigma(x, y) + a, \quad (3)$$

$$\mu(x, y) = \frac{1}{(2w_x + 1)(2w_y + 1)} \sum_{i=-w_x}^{w_x} \sum_{j=-w_y}^{w_y} I(x+i, y+j), \quad (4)$$

$$\sigma(x, y) = \sqrt{D(x, y)} = \sqrt{\mu(I(x, y)^2) - \mu(I(x, y))^2}, \quad (5)$$

где a, k — параметры метода, w_x, w_y — ширина и высота окна бинаризации.

* Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (проекты 13-07-12179 офи_м и 13-07-12172 офи_м).

Настройка алгоритма бинаризации осуществляется на основе размеченной репрезентативной выборки, представляющей собой набор прямоугольных областей сравнимого размера, для каждой из которых задан глобальный порог. Размечаются характерные в смысле яркостных свойств или размера объектов интереса области.

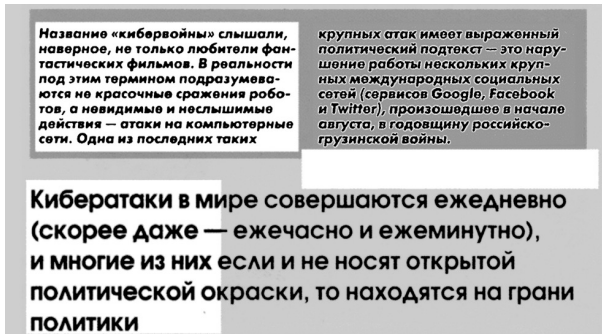


Рис. 1. Пример разметки изображения

В данной работе настройка параметров рассматривается относительно двух расстояний: MSE (Mean Square Error) и CPM (Count Pseudometric Measure, представляет собой разницу количества черных пикселей). Если $A, D \in B$:

$$MSE = \frac{\sum_{x=1}^M \sum_{y=1}^N |A(x, y) - D(x, y)|}{MN} \quad (6)$$

$$c(A) = \sum_{x=1}^w \sum_{y=1}^h \begin{cases} 1, & \text{если } A(x, y) = 0 \\ 0, & \text{если } A(x, y) = 1 \end{cases} \quad (7)$$

$$CPM(A, D) = |c(A) - c(D)|. \quad (8)$$

2. Оптимальные параметры алгоритма Ниблэка

Если θ — функция Хевисайда вида:

$$\theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}, \quad (9)$$

то результат алгоритма Ниблэка для точки изображения (x, y) :

$$A(x, y) = \theta(I(x, y) - \mu(x, y) - k\sigma(x, y) - a). \quad (10)$$

Как показано в [7], значения параметров a и k можно ограничить.

Пусть $k \in [k_{\min}, k_{\max}]$, $a \in [a_{\min}, a_{\max}]$. Если дискретизировать указанные интервалы с шагом k_{step} ,

a_{step} , то область параметров будет представлять собой матрицу или двумерную гистограмму, обозначаемую в дальнейшем как H .

Тогда определим значение в точке H :

$$H(k', a') = |\{x, y \mid \exists I(x, y) \wedge P(x, y)\}|. \quad (11)$$

Здесь $|\cdot|$ — мощность множества, $P(x, y)$ — предикат соответствующий выбранному критерию оптимизации:

$$P_{MSE}(x, y) \equiv A(x, y) \wedge GT(x, y), \quad (12)$$

$$P_{CPM}(x, y) \equiv !A(x, y), \quad (13)$$

где $GT \in B$ — «правильная» разметка (ground truth). При этом помним, что 0 — обозначает черный пиксель (false), а 1 — белый (true).

Рассмотренный в [7] способ построения гистограммы представляет собой не что иное, как прямой расчет для каждой ячейки гистограммы $\langle k', a' \rangle$ количества пикселей, соответствующих выбранному предикату при бинаризации исходного изображения I с параметрами $k = k_{\min} + k_{step} k'$ и $a = a_{\min} + a_{step} a'$ по формуле (10).

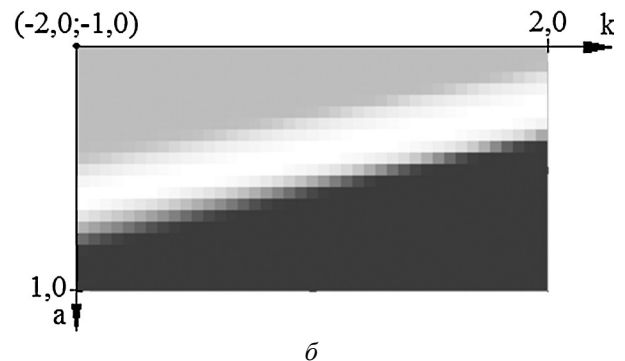
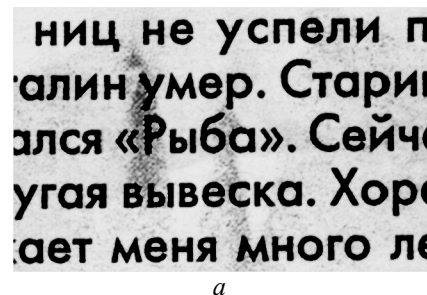


Рис. 2. Пример гистограммы изображения. a — исходное изображение; b — гистограмма для MSE критерия. Чем светлее ячейка, тем больше ее значение, т. е. для указанного критерия оптимальное решение содержится в белой области

После построения гистограммы не составляет труда найти и оптимальные параметры бинаризации:

$$(a'_*, k'_*) = \underset{\substack{0 \leq a' \leq a'_{\max} \\ 0 \leq k' \leq k'_{\max}}}{\text{argmax}} H_{MSE}(a', k'), \quad (14)$$

$$(a'_*, k'_*) = \underset{\substack{0 \leq a' \leq a'_{\max} \\ 0 \leq k' \leq k'_{\max}}}{\text{argmin}} |H_{CPM}(a', k') - c(GT)O|, \quad (15)$$

где O — матрица из одних единиц.

Значения $\mu(x, y)$ и $\delta(x, y)$ можно посчитать и сохранить заранее линейным проходом точек изображения, поэтому трудоемкость нахождения параметров бинаризации определяется сложностью построения гистограммы. Прямое вычисление H_{MSE} и H_{BCM} имеет трудоемкость порядка $O(n_w n_h m_w m_h)$, где n_w — ширина изображения I , n_h — высота I , m_w — ширина H ($(k_{\max} - k_{\min})/k_{step}$), m_h — высота H ($(a_{\max} - a_{\min})/a_{step}$). Если принять $n_w = n_h = n$, $m_w = m_h = m$, трудоемкость составит $O(n^2 m^2)$.

В дальнейшем обозначим изложенный выше способ вычисления гистограммы как ДНС (Direct Histogram Calculation).

3. Интегрирование гистограммы

В соответствии с (10) пиксель будет классифицирован как черный, если выполнено следующее неравенство:

$$a \geq I(x, y) - \mu(x, y) - \sigma(x, y)k. \quad (16)$$

Определим положение параметров бинаризации, при которых пиксель $\langle x, y \rangle$ классифицирован как черный, на гистограмме H . Границей множества будет прямая:

$$p(x, y) : a = I(x, y) - \mu(x, y) - \sigma(x, y)k. \quad (17)$$

Интересующие нас точки гистограммы расположены непосредственно на прямой и «ниже» (поскольку ось ординат направлена вниз). Так как H — дискретная решетка, то границей множества будет «дискретная» прямая:

$$p'_B(x, y) : a' = \text{ceil} \left(\frac{I(x, y) - \mu(x, y) - \sigma(x, y)(k_{\min} + k_{step}k') - a_{\min}}{a_{step}} \right) \quad (18)$$

Тогда можно «суммируя» прямые (18) получить \tilde{H}_B :

$$\tilde{H}_B(k', a') = \left| \{x, y \mid \langle k', a' \rangle \in p'_B(x, y)\} \right|. \quad (19)$$

После этого не составляет труда получить гистограмму «черных» пикселей H_B («часть» H для пикселей I , которые помечены как черные в GT) простым интегрированием \tilde{H}_B сверху вниз. То есть, для каждого столбца a' последовательно от $k' = 1$ до $k' = m_h - 1$:

$$H_B(k', a') = H_B(k' - 1, a') + \tilde{H}_B(k', a'). \quad (20)$$

В случае построения H_{CPM} все тривиально. $H_{CPM} \equiv H_B$. Если нужно построить H_{MSE} , то гистограмму черных пикселей необходимо сложить с гистограммой белых пикселей \tilde{H}_W (построенную для пикселей, помеченных белыми в GT).

Гистограмма белых пикселей строится в целом аналогично гистограмме черных пикселей. При этом, во-первых, сама прямая (17) не входит в искомое множество:

$$p'_W(x, y) : a' = \text{floor} \left(\frac{I(x, y) - \mu(x, y) - \sigma(x, y)(k_{\min} + k_{step}k') - a_{\min}}{a_{step}} \right), \quad (21)$$

а во-вторых, интегрирование

$$\tilde{H}_W(k', a') = \left| \{x, y \mid \langle k', a' \rangle \in p'_W(x, y)\} \right|$$

происходит снизу вверх.

Надо учитывать, что указанные прямые могут пересекать область гистограммы по горизонтали — нижнюю или верхнюю границы области поиска. Тогда, чтобы описанный способ интегрирования работал корректно, при построении \tilde{H}_B можно заменить точки $\langle a', k' \rangle \in p'_B(x, y)$, для которых $a' < 0$ на точки вида $(0, k')$. Аналогично, при построении гистограммы белых пикселей следует заменить точки, для которых $a' > m_h - 1$ на $(m_h - 1, k')$.

Построение одной прямой имеет трудоемкость порядка $O(m)$, количество прямых равняется количеству точек на исходном изображении и равно n^2 . Интегрирование линейно относительно гистограммы. Таким образом, общая трудоемкость составляет $O(n^2 m) + O(m^2) = O(n^2 m)$, что лучше ДНТ, имеющего, как было показано, трудоемкость $O(n^2 m^2)$. Обозначим изложенный способ расчета гистограммы как АНС (Accelerated Histogram Calculation).

4. Преобразование Хафа

Рассмотрим произвольную точку (x_i, y_i) . Очевидно, что через данную точку проходит бесконечное множество прямых, удовлетворяющих уравнению $y_i = x_i a + b$ при различных значениях a и b . Если переписать это уравнение в виде $b = -x_i a + y_i$ и рассмотреть плоскость ab , называемую пространством параметров, то для заданной пары (x_i, y_i) получаем уравнение единственной прямой.

Из вышесказанного следует, что «правильно» отмеченной точке в исходном пространстве, соответствует прямая с желаемыми свойствами в пространстве параметров. Трансформировать пространство изображения в пространство параметров позволяет преобразование Хафа (ПХ) [10]. Так как самой трудоемкой частью настройки алгоритма Ниблэка «интегрированием» является как раз построение прямых, то эффективная реализация преобразования Хафа может существенно ускорить настройку, ведь вместо того чтобы отмечать m точек, достаточно будет отметить лишь одну. Все остальное делает преобразование Хафа.

Вернемся к свойствам пространства параметров. Пусть в этот раз есть две точки (x_i, y_i) и (x_j, y_j) , и соответствующие им прямые (в пространстве параметров) пересекаются в точке (a', b') . Тогда a' — угловой коэффициент прямой, проходящей через точки (x_i, y_i) и (x_j, y_j) в исходном пространстве, а b' — координата ее пересечения с осью ординат.

Заметим, что если прямая, проходящая через точки (x_i, y_i) и (x_j, y_j) , близка к вертикали, то ее угловой коэффициент стремится к бесконечности. Поэтому можно перейти к представлению $x \cdot \cos(\phi) + y \cdot \sin(\phi) = \rho$, где ϕ — угол между перпендикуляром к прямой из начала координат и осью абсцисс, а ρ — длина перпендикуляра. В данной интерпретации точкам соответствуют в пространстве параметров не прямые, а синусоидальные кривые; а пересечение нескольких кривых в одной точке свидетельствует о том, что точки исходного пространства лежат на одной прямой.

Пространство параметров можно разбить на так называемые ячейки накопления: дискретные значения, которые могут принимать параметры. Каждой ячейке присваивается значение, равное количеству синусоидальных кривых, пересекающихся в этой точке пространства параметров — что, в свою очередь, есть количество точек исходного множества, лежащих на соответствующей прямой.

Очевидно, что алгоритм, который реализует преобразование Хафа над изображением, принимает на

вход двумерный массив (матрицу) и возвращает также двумерный массив, каждый элемент которого является ячейкой накопления. Далее по тексту под словами «результат преобразования Хафа» будет понимать именно полученный двумерный массив.

5. Быстрое преобразование Хафа

Для дальнейшего изложения и понимания нам понадобятся понятия, описанные в алгоритме, приведенном в [11]. Пусть мы имеем входное изображение размером $w \times h$, где w — его ширина, а h — высота. Те линии, для которых $|tg(\phi_i)| \leq 1$, будем называть «преимущественно вертикальными», те, для которых $|ctg(\phi_i)| \leq 1$, будем называть «преимущественно горизонтальными». Здесь ϕ_i , как уже было сказано, угол между перпендикуляром к прямой из начала координат и осью абсцисс.

В отличие от классической параметризации преобразования Хафа (ρ, ϕ) мы будем использовать $(x_0, shift)$, где x_0 — координата пересечения параметризуемой прямой с верхней строкой, а $shift = h * ctg(\alpha) = h * tg(\phi)$ — горизонтальный сдвиг «преимущественно вертикальной» линии вдоль изображения (входного массива), где α — угол наклона прямой. Для вычисления преобразования Хафа зафиксируем знак сдвига $shift$ (не больше или не меньше нуля). Отрицательный сдвиг в этом случае можно записать как $-shift$.

Таким образом, из описания видно, что алгоритм вычисления быстрого преобразования Хафа приведенный в [11], вводит понятия квадранта направления прямых для преобразования Хафа, задавая четыре возможных квадранта прямых. Для лучшего понимания данные квадранты можно задать интервалом угла ϕ для каждого типа:

- 1) $[-45^\circ, 0^\circ]$ — «преимущественно вертикальные с положительным сдвигом»;
- 2) $[0^\circ, 45^\circ]$ — «преимущественно вертикальные с отрицательным сдвигом»;
- 3) $[45^\circ, 90^\circ]$ — «преимущественно горизонтальные с положительным сдвигом»;
- 4) $[90^\circ, 135^\circ]$ — «преимущественно горизонтальные с отрицательным сдвигом».

Работать с неотрицательными углами удобнее, поэтому будем рассматривать первый квадрант в пределах интервала $[315^\circ, 360^\circ(0^\circ)]$. Для упоминания этих квадрантов в дальнейшем обозначим их соответственно 315–0, 0–45, 45–90, 90–135.

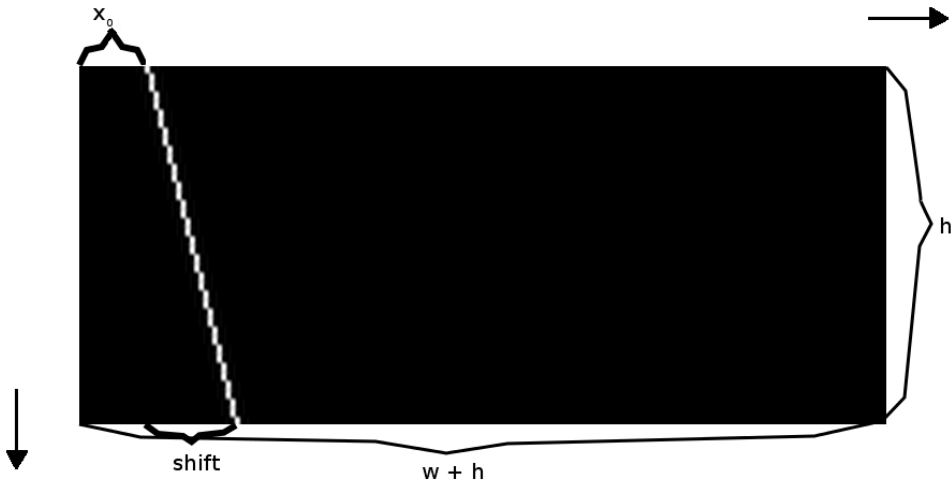


Рис. 3. Результат преобразования Хафа. Точка в исходном пространстве – (15, 20). Квадрант 0–45

Отметим, что при такой параметризации каждой строке результирующего ПХ соответствует свой угол.

Например, в случае 315–0 угол

$$\phi = \arctg(\text{shift} / h).$$

В случае других квадрантов зависимость аналогична. Причем угол монотонно изменяется от первой строки до последней.

При выбранной параметризации точке в исходном пространстве соответствует прямая в пространстве параметров в каждом из квадрантов. Докажем данный факт для квадранта 0–45.

Из рис. 3 видно, что

$$a = -\frac{h}{\text{shift}}, \quad 0 = ax_0 + b \Rightarrow b = \frac{x_0 h}{\text{shift}}.$$

Тогда для заданной точки (x_i, y_i) исходного пространства имеет место равенство

$$y_i = -\frac{h}{\text{shift}} x_i + \frac{x_0 h}{\text{shift}},$$

откуда можно выразить уравнение прямой в пространстве параметров:

$$\text{shift} = -\frac{h}{y_i} x_0 + \frac{x_i h}{y_i} \quad (22)$$

Как видим, (22) действительно является уравнением прямой.

В случае «более горизонтальных» прямых мы, очевидно, будем иметь графики в координатах не (x_0, shift) , а (shift, x_0) . Чтобы привести все графики к одному виду, нам нужно транспонировать результат ПХ для «более горизонтальных» прямых.

Заметим, что вследствие того, что часть прямых, которые могут проходить через точку, пересекают

ось абсцисс вне отрезка $[0, w-1]$, результат применения алгоритма, описанного в [11], может быть перетекающим через границы и быть циклическим (см. рис. 4), что затрудняет интерпретацию результатов ПХ.

Цикличность не возникает, если $w > h$, так как $h * \text{tg}(\phi) \leq h \leq w$. Тогда для решения проблемы цикличности в случае «преимущественно вертикальных» прямых преобразование Хафа следует выполнять для расширенной матрицы размером $h \times (w + h)$, где «лишние» точки заполняются нулями. При этом результат также увеличится до $h \times (w + h)$.

Из-за транспонирования для «преимущественно горизонтальных» прямых, входную матрицу следует увеличить до размеров $(h + w) \times w$, а выходную до $w \times (h + w)$. Таким образом, результат ПХ для любого квадранта имеет длину $(h + w)$. Результата ПХ

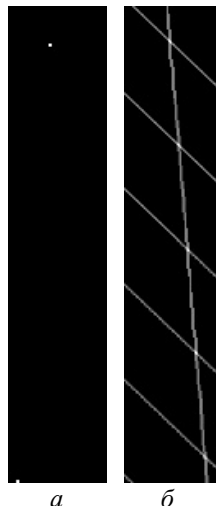


Рис. 4. Пример преобразования Хафа для квадранта 0–45. *a* – в исходном пространстве отмечены две точки; *b* – результат преобразования Хафа. Точка, находившаяся в нижней строке рисунка *a*, дала циклический результат

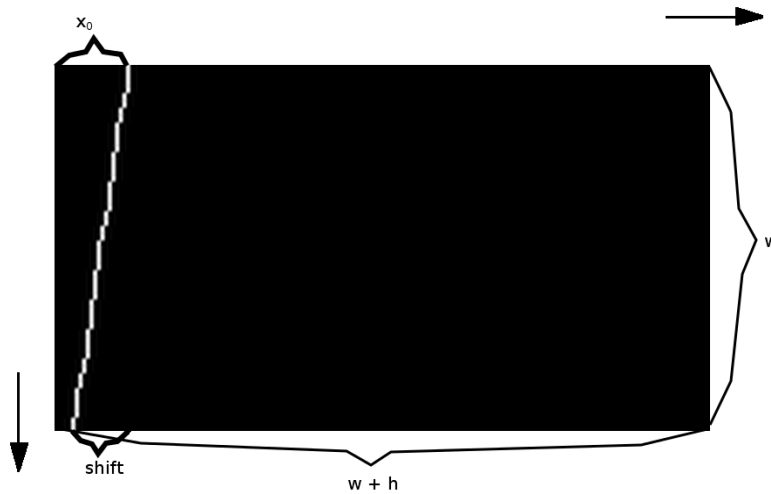


Рис. 5. Результат преобразования Хафа. Точка в исходном пространстве – (15, 20). Квадрант 90–135

для одного квадранта может быть недостаточно, и для полной картины может понадобиться полное ПХ.

Для получения полного ПХ рассмотрим результаты ПХ для отдельных квадрантов. Как уже упоминалось ранее, каждой строке результата ПХ соответствует свой угол. Рассмотрим результат ПХ для 315–0. Первая строка соответствует углу в 0 градусов, в то время как последняя соответствует углу в 315 (–45) градусов. Теперь рассмотрим результат ПХ для 0–45. Первая строка соответствует углу в 0 градусов, в то время как последняя соответствует углу в 45 градусов. Таким образом, получаем, что первые строки ПХ для этих квадрантов совпадают. Аналогично первые строки ПХ для 45–90 и 90–135 соответствуют углу в 90 градусов, а последние углам в 45 и 135 градусов. Таким образом, получаем, что первые строки ПХ для этих квадрантов совпадают, а также что последняя строка 45–90 совпадает с последней строкой 0–45.

Вследствие этого находим, что мы можем получить ПХ для всех направлений прямых, «перевернув вертикально» результаты ПХ для отдельных квадрантов и «склеив» их по общим строкам. Для того чтобы изменение угла происходило от меньшего к большему, выберем следующий порядок:

- 1) 315–0 с переворотом;
- 2) 0–45;
- 3) 45–90 с переворотом;
- 4) 90–135.

6. Построение прямых с помощью преобразования Хафа

В уравнении (17) $\sigma(x, y) \in [0; 0,5]$ [7],

то есть

$$\begin{aligned} \text{tg}(\alpha) \in [-0,5; 0] &\Rightarrow \text{ctg}(\phi) \in [-0,5; 0] \Rightarrow \\ &\Rightarrow \phi \in [90; 135]. \end{aligned}$$

Следовательно, мы можем ограничиться исключительно получением прямых, соответствующих квадранту 90–135 — «преимущественно горизонтальных» с отрицательным сдвигом.

С помощью рис. 5 можно понять, что для указанного квадранта уравнение прямой пространства параметров (22) примет вид:

$$\text{shift} = -\frac{w}{y_i} x_0 + \frac{x_i w}{y_i} \quad (23)$$

В примере на рис. 3 точке исходного пространства (15; 20) соответствовала прямая l_{0-45} в пространстве Хафа. Она пересекала ось абсцисс в точке (15; 0), а ее проекция на ту же ось равнялась 20и (в прежних обозначениях *shift*. Не следует путать с координатами пространства). На рис. 5 той же точке соответствует прямая l_{90-135} , пересекающая ось абсцисс в точке (20; 0), и с проекцией на ось равной 15-ти (следствие транспонирования исходного пространства для «преимущественно горизонтальных» прямых).

Прямая (17) пересекает границы гистограммы H в точках со значением ординат $a_{\text{down}} = a'(k_{\text{max}})$ и $a_{\text{up}} = a'(k_{\text{min}})$, где $a'(k)$ имеет вид:

$$a'(k) = \text{ceil} \left(\frac{I(x, y) - \mu(x, y) - \sigma(x, y)k - a_{\text{min}}}{a_{\text{step}}} \right), \quad (24)$$

«down» и «up» обозначают положение соответствующих точек относительно результата преобразо-

вания Хафа (следствие транспонирования). В то же время, $x_0 = a_{up}$ и $shift = a_{up} - a_{down}$. Следовательно, для того чтобы получить требуемую прямую, необходимо отметить точку $(a_{up} - a_{down}; a_{up})$ в исходной матрице и применить преобразование Хафа. В простейшем случае этим можно ограничиться. Тогда достаточно вырезать из результата преобразования прямоугольную область размерами $h \times w$ (смотрите рис. 6) и транспонировать ее — результатом будет являться \tilde{H}_B . Дальнейшие действия аналогичны методу «интегрирования»

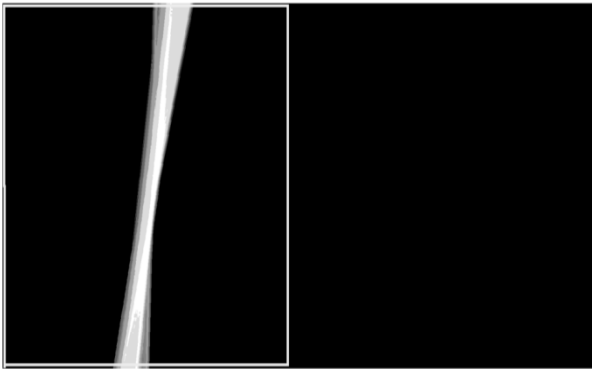


Рис. 6. Результат преобразования Хафа и область интереса, соответствующая гистограмме. Тривиальный случай

К сожалению, если диапазон параметра a недостаточен (или, что равносильно, диапазон параметра k слишком велик), то не все прямые есть возможность построить с помощью преобразования Хафа, так как точку $(a_{up} - a_{down}; a_{up})$ не всегда можно отметить в исходной матрице. Тривиальным решением проблемы является увеличение пространства параметров. Но тогда растет количество используемой памяти и операций.

Обратим внимание, что при использовании квадрата 90–135 выполнено условие $a_{up} - a_{down} \geq 0$, так как $a_{up} \geq a_{down}$ (смотрите рис. 5). То есть, точку нельзя отметить на исходном изображении, если $a_{up} < 0$ и $a_{up} > h$. Здесь следует выделить три случая.

Первый случай — $a_{up} < 0$. Это означает, что после транспонирования соответствующая прямая будет лежать целиком выше гистограммы. При построении H_B она голосует за все ячейки гистограммы, а в случае H_W ни за одну. Такие прямые можно исключить из подсчета гистограммы. Если их количество равно c' , то формула (14) не изменится, а формула (15) примет вид:

$$(a'_s, k'_s) = \underset{\substack{0 \leq a'_s \leq a'_{\max} \\ 0 \leq k'_s \leq k'_{\max}}}{\operatorname{argmin}} |H_{CPM}(a', k') - (c(GT) - c')O|, \quad (25)$$

случай — $a_{up} > h$ и $a_{down} \geq h$. Тогда соответствующая прямая после транспонирования лежит «ниже» гистограммы, то есть голосует за все ячейки H_W и ни за одну ячейку H_B . Исключение подобных прямых из подсчета гистограммы не влияет на формулы (14) и (15).

Третий случай — $a_{up} > h$ и $a_{down} < h$. Такая прямая пересекает гистограмму, и ее следует построить. Вспомним, что для решения проблемы цикличности мы расширяем входную и выходную матрицы. Рассмотрим, что произойдет, если ставить точки на уже расширенной матрице. Высота области равна w , поэтому сдвиг прямой не будет превышать w . Если $a_{down} < 0$, то будет иметь место перетекание (смотрите рис. 8), но «перетекающая» часть прямой не попадет в вырезаемую область, поскольку $w + h - shift > w + h - w = h$. Более того, область, содержащая «перетекающие» части прямых, будет ограничена треугольником, делящим «добавленный» к входной матрице квадрат размерами $w \times w$ пополам (рис. 7), так как наклон прямых ограничен. Если же $a_{down} \geq 0$, то $a_{up} < a_{down} + w < h + w$ и перетекания не будет.

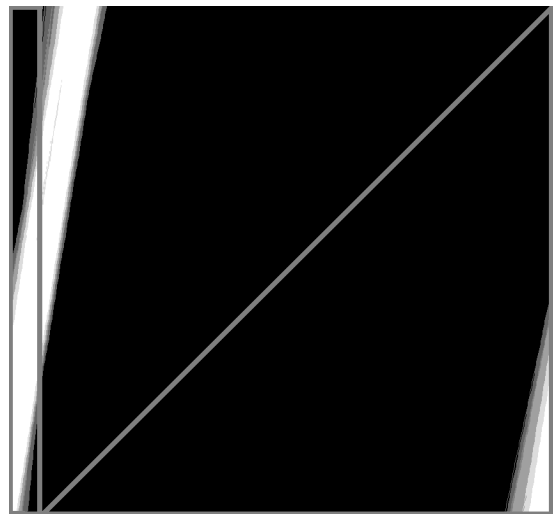


Рис. 7. Перетекание прямых при $a_{down} < 0$.
Рамкой слева выделена область интереса, треугольник ниже диагонали — область перетекания

При использовании указанной выше техники несколько усложняется процедура интегрирования, ведь нужно правильно учесть влияние фрагментов прямых, не попадающих в гистограмму (на рис. 7 прямоугольник в рамке слева). Для гистограммы

черных пикселей важны отрезки прямых, находящиеся выше гистограммы (перетекание, на рис. 7 область ниже диагонали. Напомним, что транспонирование еще не было проведено). В случае гистограммы белых пикселей следует учесть фрагменты прямых, находящихся под гистограммой (выше диагонали, но вне прямоугольника). Таким образом, для гистограммы белых пикселей интегрирование матрицы, показанной на рис. 7, правильно производить справа налево, начиная с диагонали квадрата на который расширено исходное изображение. Для гистограммы черных пикселей интегрирование приводится слева направо от диагонали до правой границы вырезаемой области (прямоугольник в рамке слева). При этом считается, что первый и последний столбцы «склеены». Результат интегрирования при вычислении гистограммы черных пикселей показан на рис. 8.



Рис. 8. Результат интегрирования

После этого область интереса вырезается и транспонируется. Результат — искомая гистограмма.

Трудоёмкость предложенного метода построения гистограммы равняется трудоёмкости преобразования Хафа и составляет $O(m^2 \log(m))$, что, как правило, значительно лучше трудоёмкости для «интегрирования» гистограммы — $O(n^2 m)$. Обозначим описанный метод получения гистограммы как FHE (Fast Histogram Estimation).

В заключение следует отметить причины, по которым мы предлагаем вынести слово estimation (приближение) в название данной схемы. В ДНС и АНС используются прямые, разграничивающие полуплоскости: аппроксимация точек прямых выполняется с помощью функций ceil или floor в зависи-

мости от типа гистограммы, как если бы прямая проходила по границам пикселей. В то же время преобразование Хафа строит прямые, соединяющие центры двух пикселей, к тому же дискретная прямая БПХ приближает эту прямую субоптимально [11]. Также нужно отметить то, что для заданного диапазона параметра k отклонение построенной прямой от идеальной не зависит от диапазона параметра a , так как в рассмотренном алгоритме данный интервал используется только для определения размеров гистограммы, но не для её построения.

7. Экспериментальные результаты

Методы тестировались на репрезентативной выборке количеством 15 изображений, с заданными 36-ю областями. Размер окна при обучении был принят равным 121 пикселю.

Тестирование можно разбить на три части. В первой было произведено сравнения предложенных в работе методов настройки (прямое построение гистограммы, «интегрирование», использование преобразования Хафа). В этом случае область параметров имела следующие ограничения: $-4 \leq k \leq 4$, $-3 \leq a \leq 0$, $k_{step} = a_{step} = step$.

Во второй части было изучено влияние размера гистограммы на качество результата. При этом было принято $-3 \leq a \leq 0$, $k_{step} = a_{step} = 0,01$. Варьировался диапазон параметра k .

В третьей части рассматривалось влияние шага дискретизации. Для уменьшения используемого количества памяти и, соответственно, для уменьшения минимального шага область поиска была уменьшена. Размеры области интереса: $-1 \leq a \leq 0$, $-2 \leq k \leq 2$.

Тестирование проводилось на процессоре Intel Core i5-2450M с частотой 2.50 GHz под управлением операционной системы Debian GNU/Linux. Результаты представлены в таблицах 1–6. Значения критериев качества представлены относительно общей размеченной площади. Данные для различных метрик приведены отдельно.

Таблица 1. Результаты настройки бинаризации Ниблека различными алгоритмами. Оптимизация по CPM.

Размер гистограммы: $-4 \leq k \leq 4$, $-3 \leq a \leq 0$

№	Метод	Step	Время, с	(a; k)	CPM $\times 10^{-2}$	MSE $\times 10^{-2}$
1	ДНС	0,01	22120	-0,32; 1,03	6,63308	6,82442
2	АНС	0,01	159	-0,32; 1,03	6,63308	6,82442
3	АНС	0,001	1752	-0,295; 0,899	6,63133	6,82337
4	FHE	0,01	4,53	-0,30; 0,93	6,63338	6,82338
5	FHE	0,001	102	-0,316; 1,007	6,63226	6,82342

Таблица 2. Результаты настройки бинаризации Ниблэка различными алгоритмами. Оптимизация по MSE.
Размер гистограммы: $-4 \leq k \leq 4$, $-3 \leq a \leq 0$

№	Метод	Step	Время, с	(a; k)	MSE $\times 10^{-2}$	CPM $\times 10^{-2}$
1	ДНС	0,01	33064	-0,31; 0,97	6,82300	6,63335
2	АНС	0,01	194	-0,31; 0,97	6,82300	6,63335
3	АНС	0,001	2060	-0,307; 0,949	6,82170	6,63324
4	FHE	0,01	4,98	-0,30; 0,93	6,82338	6,63338
5	FHE	0,001	178	-0,307; 0,957	6,82326	6,63375

В таблицах 1–2 представлена первая часть опытов. На основании полученных данных следует резюмировать, что, во-первых, АНС значительно быстрее ДНС, но медленнее чем FHE, во-вторых, значения параметров, вычисленные с помощью FHE, отличаются от результатов, полученных с использованием ДНС и АНС при одном и том же шаге. Это подтверждает, что FHE не является ускорением ДНС или АНС. В FHE оптимизируемый функционал несколько отличается, так как при использовании быстрого преобразования Хафа используется другая дискретизация прямой. Поэтому видно, что при одинаковых параметрах дискретизации оптимальными получаются другие точки, в которых и другие значения ошибки. При этом с увеличением разрешения наблюдается сходимость результата.

Таблица 3. Исследование зависимости результатов настройки бинаризации Ниблэка с использованием FHE от шага дискретизации. Оптимизация по CPM.
Размер гистограммы: $0 \leq k \leq 2$, $-1 \leq a \leq 0$

№	Step	Время, с	(a; k)	CPM $\times 10^{-2}$	MSE $\times 10^{-2}$
1	0,01	4,04	-0,30; 0,92	6,63375	6,82421
2	0,005	4,24	-0,29; 0,87	6,63405	6,82559
3	0,001	9,37	-0,304; 0,940	6,63298	6,82284
4	0,0005	28,53	-0,296; 0,902	6,63240	6,82419
5	0,0002	179,99	-0,2956; 0,9016	6,63137	6,82351

Таблица 4. Исследование зависимости результатов настройки бинаризации Ниблэка с использованием FHE от шага дискретизации. Оптимизация по MSE.
Размер гистограммы: $0 \leq k \leq 2$, $-1 \leq a \leq 0$

№	Step	Время, с	(a; k)	MSE $\times 10^{-2}$	CPM $\times 10^{-2}$
1	0,01	4,1	-0,30; 0,92	6,82421	6,63375
2	0,005	4,27	-0,29; 0,87	6,82559	6,63405
3	0,001	13,12	-0,306; 0,940	6,82430	6,63682
4	0,0005	45,55	-0,309; 0,959	6,82270	6,63431
5	0,0002	324,64	-0,3026; 0,9308	6,82242	6,63329

Следует заметить, что даже при этом различия в оптимизируемом критерии ничтожны.

Бросаются в глаза результаты опытов 1 и 2 таблицы 3. При уменьшении шага в методе FHE найденное оптимальное значение ухудшилось. Однако, принимая в расчет последующие опыты, можно сформулировать следующее утверждение: ухудшение результатов при уменьшении шага возможно, но лишь при небольших изменениях шага. По-видимому, эффект связан с упомянутым выше отклонением «сдвинутой» прямой относительно хаффовской. При существенном уменьшении шага можно рассчитывать и на значительное улучшение результата, что подтверждается остальными экспериментами. Данный недостаток, очевидно, не характерен для прямого вычисления гистограммы и «интегрирования», так как они всегда дают точный ответ.

Обратимся теперь к влиянию диапазона параметра k . Результаты экспериментов приведены в таблицах 5–6.

Таблица 5. Исследование зависимости результатов настройки бинаризации Ниблэка с использованием FHE от диапазона параметра k . Оптимизация по CPM.
Параметры гистограммы: $-3 \leq a \leq 0$, $step = 0,01$

№	$\left[\begin{matrix} k_{min} \\ k_{max} \end{matrix} \right]$	Время, с	(a; k)	CPM $\times 10^{-2}$	MSE $\times 10^{-2}$
1	[-2;2]	4,13	-0,32; 1,01	6,64060	6,82671
2	[-3;3]	4,43	-0,34; 1,13	6,63727	6,83351
3	[-5;5]	4,84	-0,28; 0,83	6,63606	6,82691
4	[-10;10]	7,55	-0,3; 0,95	6,64237	6,82750

Таблица 6. Исследование зависимости результатов настройки бинаризации Ниблэка с использованием FHE от диапазона параметра k . Оптимизация по MSE.
Параметры гистограммы: $-3 \leq a \leq 0$, $step = 0,01$

№	$\left[\begin{matrix} k_{min} \\ k_{max} \end{matrix} \right]$	Время, с	(a; k)	MSE $\times 10^{-2}$	CPM $\times 10^{-2}$
1	[-2;2]	4,33	-0,32; 1,01	6,82671	6,64060
2	[-3;3]	4,59	-0,34; 1,13	6,83351	6,63727
3	[-5;5]	5,52	-0,28; 0,83	6,82691	6,63606
4	[-10;10]	10,32	-0,33; 1,09	6,83029	6,63878

Как видно из таблиц 5 и 6, качество результата обучения напрямую не коррелирует с размером диапазона параметра k (сравните опыты 1 и 4 в таблицах 5 и 6).

8. Заключение

В работе исследовалась возможность построения эффективных вычислительных схем настройки алгоритма Ниблэка. Предложена техника интегрирования гистограммы. Исследована возможность использования преобразования Хафа. Показаны преимущества и недостатки рассмотренных схем.

В целом можно сказать, что FHE дает значительный выигрыш в скорости нахождения параметров, но в то же время вносит в итоговый результат некоторую погрешность. В приведенных в статье опытах уменьшение шага на десятичный порядок позволило превзойти результаты других методов (ДНС, АНС) с меньшим шагом в случае оптимизации по СРМ-критерию. При настройке FHE по MSE-критерию данная зависимость не соблюдается. Таким образом, поведение погрешности FHE метода требует дополнительного изучения.

Следует заметить, что все методы показали достаточно близкие результаты относительно выбранных в работе расстояний. Отчасти это говорит о простоте используемой выборки, отчасти о малой информативности выбранных расстояний, что указывает на направление дальнейших исследований.

Литература

1. Арлазаров В. Л., Логинов А. С., Славин О. А. Характеристика программ оптического распознавания текста // Программирование. 2002. № 7.

Алиев Михаил Александрович. М.н.с. ИСА РАН. Окончил в 2008 г. МИСИС. Область научных интересов: обработка изображений. E-mail: aliev.michael@gmail.com

Николаев Дмитрий Петрович. Зав. сектором ИППИ РАН. К. ф.-м. н. Окончил в 2000 г. МГУ. Количество печатных работ: 103. Область научных интересов: быстрые алгоритмы обработки изображений. E-mail: dimonstr@iitp.ru

Сараев Антон Александрович. Программист ООО «БИФИТ Сервис». Окончил в 2013 г. НИТУ МИСиС. Количество печатных работ: 2. Область научных интересов: обработка изображений. E-mail: antonysaraev@gmail.com

2. [Электронный ресурс]
<http://www.3dnews.ru/software/632560/print>
3. Tesseract OCR Engine, What it is, where it came from, where it is going. Ray Smith, Google Inc OSCON 2007.
4. Безматерных П. В., Николаев Д. П., Постников В. В. Быстрый алгоритм детектирования рукописных пометок // Труды ИСА РАН. 2008. Т. 38, НАПРАВЛЕНИЕ II
5. Somasundaram K., Kalavathi P. Medical Image Binarization using Square Wave Representation // Image Processing Lab. — Gandhigram, 2011. P. 152–158.
6. Sezgin M., Sankur B. Selection of thresholding methods for nondestructive testing applications // International Conference on Image Processing. 2001. V. 3. P. 764–767.
7. Николаев Д. П., Сараев А. А. Критерии оценки качества в задаче автоматизированной настройки алгоритмов бинаризации // Труды института системного анализа Российской академии наук (ИСА РАН). 2013. Т. 63. Вып. 3. С. 85–94.
8. Niblack W. An Introduction to Digital Image Processing // Englewood Cliffs, PrenticeHall, 1986.
9. Миронов А. С., Николаев Д. П. Сравнение вариантов реализации алгоритма Ниблэка бинаризации полутоновых изображений // Информационные технологии и системы (ИТиС'10). М., 2010. С. 138–144.
10. patent 1962, Method and means for recognizing complex patterns. US 3069654 A.
11. Nikolaev D. P., Karpenko S. M., Nikolaev I. P., Nikolaev P. P. Hough Transform: Underestimated Tool in the Computer Vision Field // 22st European Conference on Modelling and Simulation, ECMS 2008. Nicosia, Cyprus. P. 238–243.