

Поиск оптимального положения при сравнении оцифрованных образов

Аннотация. В статье описывается задача сравнения оцифрованных (отсканированных) образов. Приводится теоретическое решение задачи поиска параллельного переноса эталонного изображения, при котором его совпадение с тестируемым изображением максимально. Описываются вычислительные эксперименты, связанные с анализом сдвигов бинарных оцифрованных символов. Производится сравнение конструктивно вычисляемой границы возможного сдвига одного графического образа относительно другого и реальных сдвигов при рассмотрении класса бинарных оцифрованных символов.

Ключевые слова: изображение, функция сравнения, оптимальное положение фигур, сдвиг изображений.

Введение

Во многих алгоритмах распознавания производится сравнение двух бинарных образов с помощью некоторой функции близости, чтобы понять, являются ли они образами двух «одинаковых» объектов. Функция близости выбирается так, чтобы удовлетворять требованиям алгоритма распознавания и учитывать особенности образов, получаемых из заданного источника.

Мы рассматриваем пространство образов U , представленных в виде матриц $R = \|r_{ij}\|$ одинакового размера $0 < i \leq m$, $0 < j \leq n$. Размеры реального образа существенно меньше размеров матрицы $m \times n$. Размещение образа с размерами $m_1 \times n_1$ ($m_1 < m$, $n_1 < n$) в матрице с размерами $m \times n$ назовем *центрированием*. Если реальный образ представлен матрицей $m_1 \times n_1$, то можно, например, центрировать этот образ так, чтобы центры матриц $m_1 \times n_1$ и $m \times n$ совпадали. Можно придумать и применять другие способы центрирования, для нас важно, что способ центрирования един для всех образов.

При кластеризации образов и их распознавании $[0, 0]$ используется неотрицательная функция близости d , которая сравнивает два образа $A = \|a_{ij}\|$ и $B = \|b_{ij}\|$, $A \in U$, $B \in U$. Условно мы будем называть образ A *эталонным*, а B - *тес-*

тируемым. При получении образов с помощью цифровых устройств (сканеры, фото- и видеокамеры) возникает ряд *эффектов оцифровки*. Простейший из них состоит в том, что из одного прообраза могут получаться образы, различающиеся как между собой, так и с прообразом. Это объясняется случайностью совмещения прообраза и цифровой сетки (например, сетки сканера) и иллюстрируется Рис.1, в котором показано как из одного исходного прообраза (серые полосы) получают два различных оцифрованных бинарных образа (черные полосы).

Среди других эффектов оцифровки отметим искажения, связанные с наклоном, и искажения, происшедшие от случайного шума.

К функции близости d , используемой в кластеризации и распознавании, предъявляются два требования:

- аксиома тождества - $d(A, B) = 0 \Leftrightarrow A = B$;
- аксиома симметрии - $d(A, B) = d(B, A) \forall A, B$.

Вклад эффектов оцифровки при сравнении бинарных образов устраняются (минимизируется) выбором функции близости и применением процедуры сдвига. В настоящей работе в качестве функции близости рассматривается метрика Хэмминга:

$$\beta(A, B) = \sum_{i=1}^m \sum_{j=1}^n |a_{ij} - b_{ij}|.$$

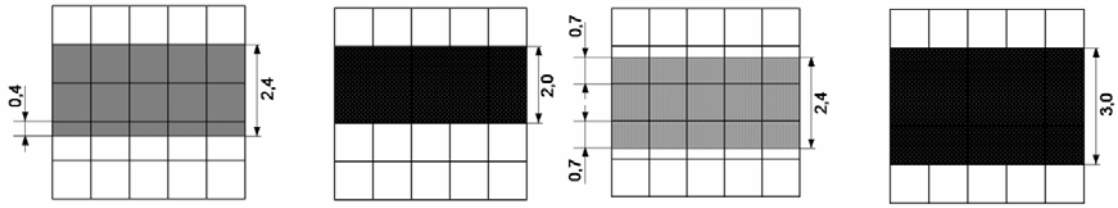


Рис. 1. Примеры оцифровки образа при различном наложении на сетку сканера
серые полосы – прообразы, черные – бинарные образы

При сравнении образов A и B из пространства U будем производить всевозможные сдвиги образа $A = \|a_{ij}\|$ в разных направлениях и рассматривать функцию близости между получившимся образом и B . Определим сдвиг матрицы на величину (h, w) :

$$\bar{A} = A(h, w) = \|a_{(i-h) \bmod m, (j-w) \bmod n}\|.$$

Другими словами, $A(h, w)$ получается из A в результате переноса на вектор (h, w) на торе, являющимся замыканием области прямоугольника с размерами $m \times n$ (Рис. 2).

Теперь мы можем определить "расстояние" d_0 между образами A и B :

$$d_0(A, B) = \min_{h, w} (\beta(A(h, w), B)) \quad (1)$$

Матрицу $A(h, w)$, на которой достигается минимум, мы будем называть оптимальным положением образа A относительно B , а вектор (h, w) – оптимальным сдвигом.

Существуют различные способы определения $d_0(A, B)$ и оптимального сдвига (h, w) . Один из простейших – вычислять функцию близости $d(h, w) = \beta(A(h, w), B)$, ограничив значения h и w : $h \leq n_0$, $w \leq n_0$.

Вообще говоря, наименьшее значение n_0 , гарантирующее достижения реального минимума $d(h, w)$, заранее неизвестно, верхняя гра-

ница n_0 определяется величиной сдвига, при котором образы A и B не пересекаются, т.е. $a_{ij} \wedge b_{ij} = 0 \forall i$ и j . Для таких n_0 способ определения $d_0(A, B)$ является *точным*. При значениях n_0 , которые не гарантируют достижения реального минимума $d(h, w)$, способ определения оптимального положения является *приближенным* и позволяет вычислять $d_0(A, B)$ с некоторой погрешностью.

Мы полагаем, однако, что исходное положение образа A , которое выбирается обычно отнюдь не случайно, не должно быть слишком далеко от оптимального. Чуть точнее, если наугад взятое положение удачно, т.е. если несовпадение эталонного и тестируемого изображений $\beta(A, B)$ составляет малую часть эталонного изображения, то параллельный перенос (h, w) , максимизирующий совпадение изображений, мал (т.е. по длине составляет малую часть диаметра эталонного изображения). Собственно, именно этот эффект и является обоснованием для ограничения рассматриваемых векторов (h, w) .

Ясно, что чем меньше n_0 , тем меньше времени будет занимать вычисление $d = \min_{h, w} d(h, w)$, но хотелось бы иметь какие-

то оценки влияния сдвига при сравнении двух образов и, тем самым, оценки близости положения некоторого образа к оптимальному.

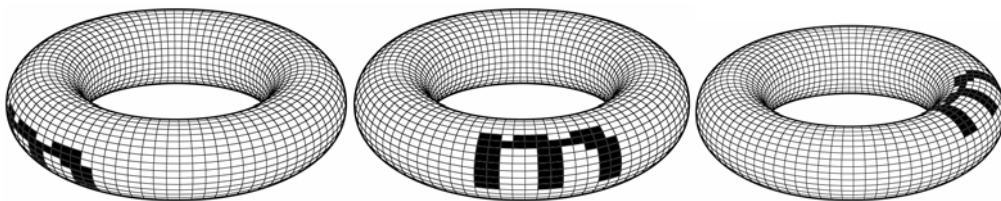


Рис. 2. Примеры переноса образа символа m на различные векторы на торе

Получению таких оценок посвящена настоящая работа. В первом разделе, являющемся развитием работы [3] дается теоретический анализ, во втором – эксперименты над объектами одной из популярных задач распознавания – символами русского алфавита.

1. Исследование влияния сдвигов при сравнении образов

1.1. Описание задачи

Ниже приводится математическая модель, объясняющая приведенный выше эффект на качественном уровне. При этом мы существенно расширим исходные постановки. В модели изображение рассматривается как пара, состоящая из множества I (ненулевые точки изображения), и плотности $\varphi: I \rightarrow \mathbb{R}$ (распределение весов точек изображения). Степень совпадения изображения с плотностями φ_0 и φ_1 моделируется интегралом:

$$\int |\varphi_1 - \varphi_2| d\sigma.$$

Пусть Ω – пространство с мерой $d\sigma$ и X – множество пар $x = (I, \varphi)$, где $I \subset \Omega$ – измеримое подмножество и $\varphi: I \rightarrow \mathbb{R}$ – интегрируемая неотрицательная функция, называемая *плотностью*. На множестве X определена метрика $\rho(x_1, x_2) = \int_{\Omega} |\varphi_1 - \varphi_2| d\sigma$, где

$$x_1 = (I_1, \varphi_1), x_2 = (I_2, \varphi_2).$$

Пусть $\gamma: \Omega \rightarrow \Omega$ взаимнооднозначное отображение, сохраняющее меру, т.е. $\gamma^* d\sigma = d\sigma$. Отображение γ индуцирует изометрию $\bar{\gamma}: X \rightarrow X$, переводящую точку $x = (I, \varphi)$ в точку $\bar{\gamma}(x) = (\gamma(I), (\gamma^{-1})^* \varphi)$. По подмножеству $\Theta \subset \Omega$ и числу α определим подмножество $Y \subset X$, состоящее из точек (I, φ) таких, что:

$$I \subset \Theta, \quad \int_{\Omega} \varphi d\sigma = \alpha.$$

Задача 1. Для функции $\mathcal{E}(\Theta, \alpha, \gamma) = \min_{x \in Y} \rho(x, \bar{\gamma}(x))$ найти (наилучшую) оценку снизу.

Один из возможных вариантов ответа на задачу 1 изложен в следующем пункте.

Остановимся на одной модификации задачи 1.

Расширим множество X и определим множество

$\tilde{X}$ состоящее из пар $\tilde{x} = (I, \varphi)$, в которых

плотность φ может концентрироваться в точках. Точнее, в $\tilde{X}$ в качестве плотностей рассматриваются

обобщенные плотности φ , являющиеся суммами интегрируемых функций и конечных линейных комбинаций δ -функций в точках $q \in \Omega$ (мы будем обозначать их символом δ_q), т.е. $\tilde{\varphi} = \varphi + \sum a_k \delta_{q_k}$. На простран-

ство $\tilde{X}$ автоматически переносятся определения расстояния, отображения $\bar{\gamma}$ и функции

$\mathcal{E}$, которые мы будем обозначать, соответственно, $\tilde{\rho}$, $\tilde{\gamma}$ и $\tilde{\mathcal{E}}$. Пусть даны отображение

$\gamma: \Omega \rightarrow \Omega$, подмножество $\Theta \subset \Omega$ и число

α . Пусть подмножество $\tilde{Y} \subset \tilde{X}$ состоит из точек $(I, \tilde{\varphi})$, где $\tilde{\varphi} = \varphi + \sum a_k \delta_{q_k}$, таких, что:

$$\begin{aligned} I &\subset \Theta, \\ \varphi &\geq 0, a_k \geq 0, \text{ носитель функции } \varphi \\ &\text{лежит в } I \text{ и } q_k \in I, \\ \int_{\Omega} \varphi d\mu + \sum a_k &= \alpha \end{aligned}$$

Задача 1'

Для функции $\tilde{\mathcal{E}}(\Theta, \alpha, \gamma) = \min_{\tilde{x} \in \tilde{Y}} \tilde{\rho}(\tilde{x}, \tilde{\gamma}(\tilde{x}))$

найти (наилучшую) оценку снизу.

1.2. Частичное решение задачи 1

Справедлива следующая теорема.

Теорема 1. Если для некоторого целого n множество $\gamma^n(\Theta)$ не пересекается с множеством Θ , то $\mathcal{E}(\Theta, \alpha, \gamma) \geq 2\alpha / n$.

Доказательство. С точкой $y \in Y$ свяжем последовательность $y_0, y_1, \dots, y_n$ точек в X , определенных соотношениями:

$$y_0 = y, y_1 = \bar{\gamma}(y), \dots, y_n = \bar{\gamma}^n(y).$$

Пусть $y = (I, \varphi)$.

Поскольку $y \in Y$, то $I \cap \gamma^n(I) = \emptyset$. Поэтому

$$\begin{aligned} \rho(y_0, y_n) &= \int_{\Omega} |\varphi - (\gamma^{-n})^* \varphi| d\sigma = \int_A \varphi d\sigma + \\ &+ \int_{\gamma^n(A)} (\gamma^{-n})^* \varphi d\sigma = 2\alpha \end{aligned}$$

Так как отображение $\bar{\gamma}$ является изометрией, то для любого k имеем $\rho(\bar{\gamma}^{k-1}(y), \bar{\gamma}^k(y)) = \rho(y, \bar{\gamma}(y))$. По аксиоме треугольника

$$\rho(y_0, y_1) + \dots + \rho(y_{n-1}, y_n) \geq \rho(y_0, y_n),$$

т.е. $n\rho(y, \bar{\gamma}(y)) \geq 2\alpha$ или $\rho(y, \bar{\gamma}(y)) \geq 2\alpha/n$. Что и требовалось доказать.

Покажем, что при выполнении некоторых дополнительных условий оценку из теоремы 1 нельзя улучшить. Главное из них - следующее условие.

Условие $\wp_n$: существует точка $q \in \Omega$, такая, что точки $q, \gamma(q), \dots, \gamma^n(q)$ различны и все они, за исключением точки $\gamma^n(q)$, лежат во множестве Θ .

Предположим дополнительно, что множество Ω снабжено топологией, в которой множество Θ открыто и отображение γ непрерывно.

Теорема 2. Пусть выполнены перечисленные выше условия и $\gamma^n(\Theta) \cap \Theta = \emptyset$. Тогда $\varepsilon(\Theta, \alpha, \gamma) = 2\alpha/n$.

Доказательство. Выберем точку q , удовлетворяющую условию $\wp_n$, и возьмем ее окрестность E положительной меры, столь малую, что для каждого $k = 0, 1, \dots, n$ окрестности $\gamma^k(E)$ не пересекаются друг с другом и все они, за исключением окрестности $\gamma^n(q)$, лежат во множестве Θ . Пусть φ_0 – неотрица-

тельная функция на Ω , носитель которой лежит в области E , такая, что $\int_U \varphi_0 d\sigma = \alpha/n$.

Пусть $x = (I, \varphi)$ – точка из Θ , определенная следующими соотношениями:

$$I = \bigcup_{0 < k \leq n} \gamma^k(E), \varphi = \sum_{0 < k \leq n} (\gamma^{-k})^* \varphi_0.$$

Тогда $\bar{\gamma}(x) = (\gamma(I), (\gamma^{-1})^* \varphi)$. Имеем

$$\gamma(I) = \bigcup_{0 < k \leq n} \gamma^k(E), (\gamma^{-1})^* \varphi = \sum_{0 < k \leq n} (\gamma^{-k})^* \varphi_0.$$

Легко видеть, что

$$\rho(x, \bar{\gamma}(x)) = \int_U \varphi_0 d\sigma + \int_{\gamma^n(U)} (\gamma^{-n})^* \varphi_0 d\sigma = 2\alpha/n.$$

Поэтому $\varepsilon \leq 2\alpha/n$. По теореме 1 $\varepsilon \geq 2\alpha/n$. Теорема 2 доказана.

Частичное решение задачи 1': теорема 1 дословно обобщается на случай задачи 1'. Мы не будем останавливаться на формулировке и доказательстве этого обобщения. Теорема 2 для случая задачи 1' допускает усиление: не нужно дополнительно требовать, чтобы множество Ω было снабжено топологией, в которой множество Θ открыто и отображение γ непрерывно.

Теорема 2'. Пусть выполнено условие $\wp_n$ и $\gamma^n(\Theta) \cap \Theta = \emptyset$. Тогда $\varepsilon(\Theta, \alpha, \gamma) = 2\alpha/n$.

Для доказательства достаточно рассмотреть точку $\tilde{x} \in \tilde{X}$, равную набору точек $q, \gamma(q), \dots, \gamma^{n-1}(q)$, наделенных плотностями, равными δ -функции, умноженной на α/n . Дальнейшее доказательство теоремы 2' аналогично доказательству теоремы 2.

1.3. Случай евклидова пространства

Пусть пространство Ω с мерой $d\sigma$ – это пространство $\mathbb{R}^n$ с евклидовой мерой. Отображение γ – это сдвиг на фиксированный вектор $a \in \mathbb{R}^n$ (т.е. $\gamma(x) = x + a$), а множество Θ – это любое n -мерное выпуклое подмножество в $\mathbb{R}^n$. Рассмотрим совокупность неотрицательных интегрируемых функций φ в $\mathbb{R}^n$, тождественно равных нулю вне выпуклого множества Θ и удовлетворяющих соотношению:

$$\int_{R^n} \varphi(x) d\sigma = \alpha.$$

Задача 1 для евклидова пространства. Найти (наилучшую) оценку снизу функции $\varepsilon(\Theta, \alpha, a)$, заданной формулой:

$$\varepsilon = \min_{\varphi} \int_{R^n} |\varphi(x) - \varphi(x-a)| d\sigma$$

Задача 1' в рассматриваемом случае формулируется аналогично.

Вектор $a \in R^n$ мы будем записывать в виде $a = |a|v$, где v - вектор единичной длины и $|a|$ - длина вектора a . Для вектора v единичной длины назовем v - *диаметром* $d_v(\Theta)$ выпуклой фигуры Θ точную верхнюю грань множества длин отрезков, являющихся сечениями области Θ прямыми, параллельными вектору v . Для ненулевого вектора $a = |a|v$ назовем a - *диаметром* $d_a(\Theta)$ выпуклой фигуры Θ отношение $d_v(\Theta)/|a|$ ее v -диаметра к длине $|a|$ вектора a .

Приведем решение задач 1 и 1' для вектора $a = |a|v$ в случае, когда a - диаметр фигуры Θ не является натуральным числом.

Теорема 3. Если для некоторого целого $n \geq 0$ справедливы неравенства

$$n|a| < d_v(\Theta) < (n+1)|a|,$$

то $\varepsilon = 2\alpha / (n+1)$.

Доказательство. В условиях теоремы 3 имеем $\gamma^{n+1}(\Theta) \cap \Theta = \emptyset$. Поэтому согласно теореме 1 выполняется неравенство $\varepsilon \geq 2\alpha / (n+1)$. Так как точная верхняя грань длин сечений области Θ прямыми, параллельными вектору v , больше числа $n|a|$, то существует точка y , лежащая строго внутри области Θ , такая, что все точки $y, y+a, \dots, y+na$ лежат строго внутри Θ . По теореме 2 имеем $\varepsilon \leq 2\alpha / (n+1)$, откуда и вытекает теорема 3.

Приведем решение задачи 1' для вектора $a = |a|v$ в случае, когда a - диаметр фигуры Θ является натуральным числом.

Теорема 4. Если для некоторого целого $n \geq 0$ справедливо равенство

$$n|a| = d_v(\Theta),$$

то $\varepsilon = 2\alpha / (n+1)$

Доказательство теоремы 4 аналогично доказательству теоремы 3. Если выполняется равенство $n|a| = d_v(\Theta)$, то минимум в задаче не достигается на обычных функциях. Однако минимум достигается на линейной комбинации δ -функций, и он равен $2\alpha / (n+1)$.

1.4. Следствия найденных оценок

Формула для функции $\varepsilon(\Theta, \alpha, a)$ достаточно громоздка. В следствии 1 приводится неточная, но достаточно хорошая оценка этой функции, которая вполне обозрима. В следствии 2 приводится оценка этой функции, которая для фиксированного вектора v при малых значениях длины $|a|$ линейна по длине $|a|$, т.е. равна $\lambda|a|$ с явно выписанной константой λ .

Следствие 1. Справедлива оценка:

$$\varepsilon(\Theta, \alpha, a) \geq \frac{2\alpha}{d_v(\Theta) + |a|} |a|.$$

Эта оценка точна при $|a| = d_v(\Theta), d_v(\Theta)/2, d_v(\Theta)/3, \dots$ и при $|a| \rightarrow \infty$ (т.е., образно говоря, точна при $|a| = d_v(\Theta)/0$). Если $\varepsilon(\Theta, \alpha, a) < 2\alpha$, то это неравенство можно переписать в виде:

$$|a| \leq \frac{\varepsilon(\Theta, \alpha, a)}{2\alpha - \varepsilon(\Theta, \alpha, a)} d_v(\Theta).$$

Следствие 2. Для любого $0 \leq y_0$ при $0 \leq |a| \leq y_0$ имеем

$$\varepsilon(\Theta, \alpha, a) \geq \frac{2\alpha}{d_v(\Theta) + y_0} |a|,$$

при $|a| > y_0$ имеем

$$\varepsilon(\Theta, \alpha, a) \geq \frac{2\alpha y_0}{d_v(\Theta) + y_0}.$$

В частности, для $y_0 = d_v(\Theta)$ получаем следующее: при $0 \leq |a| \leq d_v(\Theta)$ имеем

$$\varepsilon(\Theta, \alpha, a) \geq \frac{\alpha}{d_v(\Theta)} |a|,$$

при $|a| > d_v(\Theta)$ имеем $\varepsilon(\Theta, \alpha, a) \geq \alpha$ (отметим, что на самом деле, при $|a| > d_v(\Theta)$ справедливо равенство $\varepsilon(\Theta, \alpha, a) = 2\alpha$).

Доказательство. Следствие 1 автоматически вытекает из теорем 3 и 4, следствие 2 - из следствия 1.

1.5. Качественное объяснение рассматриваемого эффекта

Пусть $\varphi(x)$ - неотрицательная интегрируемая функция на R^n , равная нулю вне выпуклого множества Θ . Пусть

$$\alpha = \int_{R^n} \varphi(x) d\sigma.$$

Пусть $g(x)$ другая неотрицательная интегрируемая функция на R^n . Пусть

$$\beta = \int_{R^n} |\varphi(x) - g(x)| d\sigma. \quad (2)$$

Допустим, что интеграл β значительно меньше, чем α .

Задача 2. Меняя параметр $a \in R^n$, добиться, чтобы интеграл

$$\psi(a) = \int_{R^n} |\varphi(x-a) - g(x)| d\sigma$$

принял, возможно меньшее значение.

Наша цель - показать, что если $\alpha \gg \beta$, то для минимизации числа $\psi(a)$ интересны лишь малые по длине вектора $a = |a|v$, которые значительно короче, чем v -диаметр $d_v(\Theta)$ области Θ .

Определим функцию $\varepsilon(a)$ как

$$\varepsilon(a) = \min_{\varphi} \int_{R^n} |\varphi(x) - \varphi(x-a)| d\sigma,$$

где минимум берется по всем неотрицательным функциям φ , удовлетворяющим соотношению (2) и имеющим носитель в области Θ . Мы можем применять найденные выше оценки функции $\varepsilon(a)$.

Теорема 5. Зафиксируем вектор v единичной длины. Пусть $d(v)$ - минимальное число, такое, что для вектора $a = |a|v$, при $|a| > d(v)$ выполнено неравенство $\varepsilon(a) > 2\beta$. Тогда при $|a| > d(v)$ справедливо неравенство:

$$\int_{R^n} |\varphi(x-a) - g(x)| d\sigma > \beta.$$

В частности, для минимизации $\psi(a)$ достаточно рассматривать лишь векторы $a = |a|v$, для которых $|a| < d(v)$.

Доказательство вытекает из аксиомы треугольника. Действительно, имеем

$$\int_{R^n} |\varphi(x) - g(x)| d\sigma = \beta,$$

$$\int_{R^n} |\varphi(x-a) - g(x)| d\sigma \geq \varepsilon(a).$$

Откуда при $|a| > d(v)$

$$\int_{R^n} |\varphi(x-a) - g(x)| d\sigma \geq \varepsilon(a) - \beta > 2\beta - \beta = \beta,$$

что и требовалось доказать.

Следствие 3. Для минимизации интеграла $\psi(a)$ на прямой $a = |a|v$, натянутой на вектор v , достаточно рассматривать лишь векторы a , для которых выполнено неравенство

$$|a| < s_v = \frac{2\beta}{\alpha} d_v(\Theta); \quad (3)$$

При $|a| > 2\beta d(\Theta) / \alpha$ справедливо неравенство $\int_{R^n} |\varphi(x-a) - g(x)| d\sigma > \beta$.

Доказательство вытекает из теоремы 5 и следствия 1.

Следствие 3 доставляет качественное объяснение обсуждаемого эффекта. Действительно, если $\beta \ll \alpha$, то число $\beta / (\alpha - \beta)$ мало. Поэтому минимум $\psi(a)$ достигается при длине вектора a , значительно меньшей, чем диаметр $d_v(\Theta)$ фигуры Θ в направлении вектора v .

Тем самым доказано, что для достижения «оптимального» положения двух фигур достаточно малых сдвигов в том случае, когда мера несовпадения при малых сдвигах незначительна.

2. Эксперименты

2.1. Исходные объекты, параметры и алгоритмы

Доказанные теоремы дают конструктивно вычисляемую границу возможного сдвига одного графического образа относительно другого для получения минимального расстояния. Интересно понять, какие реальные сдвиги будут получаться при рассмотрении конкретного класса объектов? Мы приведем результаты экспериментов, в которых в качестве объектов выступали печатные символы, взятые из отсканированного образа страницы обычной книги.

Итак, исходный материал – тестовое множество M , состоящее из $k=1770$ образов букв русского языка. Рассматриваются различные пары знаков, всего $k(k-1)/2=1565565$ пар.

Мы будем использовать для установки начального (нулевого) положения двух изображений A_1 и A_2 относительно друг друга некоторый алгоритм $\aleph_1$. Значение функции близости, основанной на метрике Хэмминга, в этом положении будем обозначать $\beta(A_1, A_2)$. Кроме того, мы будем рассматривать оптимизационный алгоритм $\aleph_0$. Этот алгоритм перебирает различные параллельные переносы одного из изображений на всевозможные вектора (h, w) относительно начального положения и выбирает тот, при котором метрика Хэмминга принимает минимальное значение. Это (оптимальное) значение мы будем обозначать через $d_0(A_1, A_2)$ а оптимизирующий вектор (h, w) – через $a(A_1, A_2)=|a| \cdot v$, где v – единичный вектор направления. Обычно мы будем рассматривать величину $n(A_1, A_2)=|a(A_1, A_2)|$, характеризующую расстояние между нулевым и оптимальным положением в пикселях.

Количество ненулевых пикселей изображения A будем обозначать α . Как показано в предыдущем разделе, важную роль играет отношение β/α . При этом, поскольку во всех случаях, когда речь идет о паре изображений A_1 и A_2 , величины β/α_1 и β/α_2 выступают симметрично, мы будем рассматривать величину $t(A_1, A_2)=\min(\beta/\alpha_1, \beta/\alpha_2)$.

Это, в частности, «симметризует» формулы п. 1.5, относящиеся к функциям $\varphi(x)$ и $g(x)$.

Это же избавляет нас от необходимости рассматривать большие значения β/α , так как все-

гда $t \leq 2$. Соответствующую величину в оптимальном положении будем обозначать t_0 :

$$t_0(A_1, A_2)=\min(d_0/\alpha_1, d_0/\alpha_2).$$

Отметим, наконец, что в статистических таблицах, а также там, где это не ведет к путанице, мы будем опускать аргументы A_1 и A_2 и писать просто β, d_0, t, t_0, n .

2.2. Общие характеристики

Среди 1770 образов – символов русского алфавита, встретившихся на странице, некоторые являются одноименными (например, буквами "е"). Ясно, что многие характеристики для таких пар могут сильно отличаться от соответствующих характеристик "разноименных" пар. Поэтому иногда мы будем рассматривать статистики по таким парам отдельно.

В тех случаях, когда в графике по одной из осей откладывается количество пар, обладающих каким-то свойством, максимальные значения могут быть велики, но для нас важно видеть и малые значения. В таких случаях мы будем пользоваться логарифмическим масштабом: $n \rightarrow \log_2(n+1)$.

Обратимся к статистике. Прежде всего, покажем общее распределение количества пар по различным значениям t_0 и t . На Рис. 3 дана гистограмма зависимости $n(t_0)$ и $n(t)$ для всех пар символов.

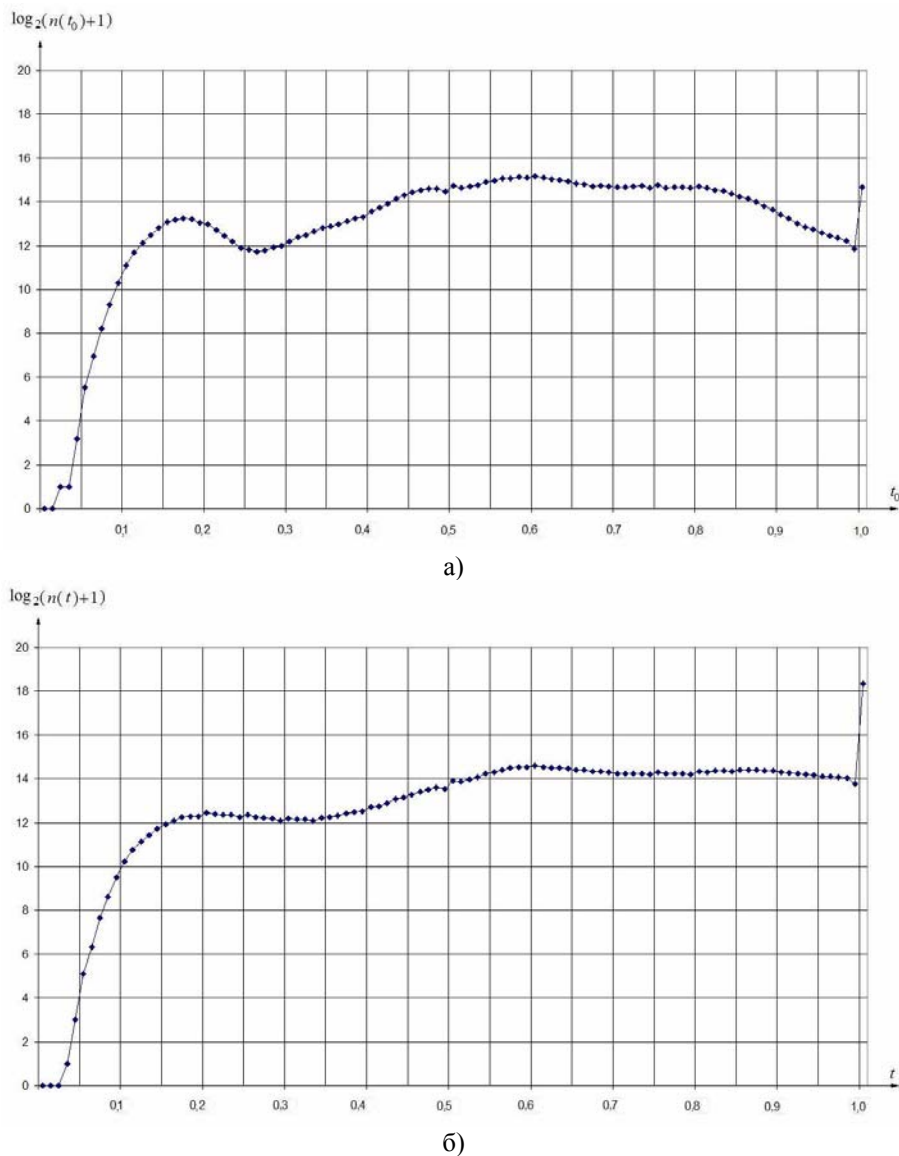
Добавим, что среднее значение t_0 для всех пар равно 0,6, среднее значение t – 0,66.

Рассмотрим на Рис. 4 аналогичную гистограмму для одноименных пар.

Здесь среднее значение t_0 для всех пар равно 0,17, среднее значение t – 0,23.

В среднем алгоритм $\aleph_1$ дает значения не очень далекие от оптимальных. Для одноименных пар, в «правой части» значения $n(t)$ гораздо менее круто спадают к нулю, чем $n(t_0)$. Это означает, что большие значения t отнюдь не всегда указывают на то, что соответствующее оптимальное значение t_0 велико. Однако для небольших значений t_0 и t оба графика ведут себя примерно одинаково.

Заметим, что, как с точки зрения применения теоремы 5, так и в контексте распознавания пары с большим t_0 (например, $t_0 > 0,5$) мало интересуют нас, поэтому для них качество алгоритма $\aleph_1$ и поиск оптимального сдвига не слишком актуальны.

Рис. 3. Общее количество пар в зависимости от t_0 (а) и t (б)

Дополнительно рассмотрим на Рис. 5 гистограмму количества пар в зависимости от величины t_0/t .

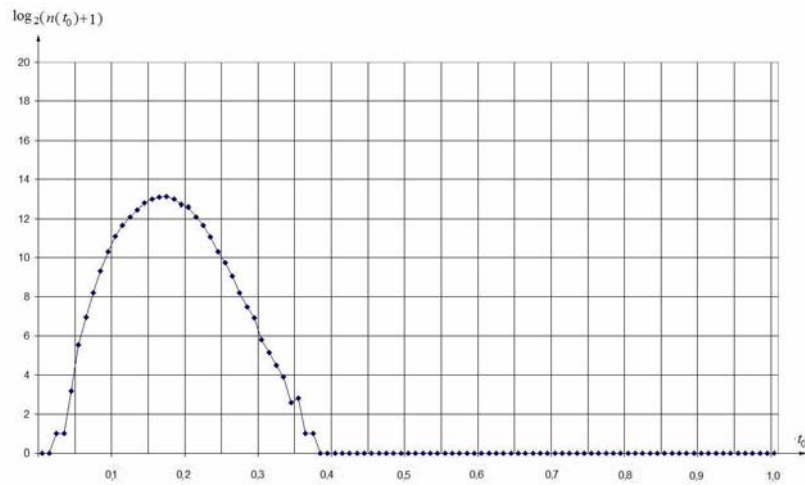
Как мы видим, более 15% всех рассмотренных пар получают при использовании алгоритма $\aleph_1$ значение, отличающееся от наилучшего из возможных (идеального) не более чем на 1%, а затем их число монотонно уменьшается вместе с t_0/t .

Таким образом, можно считать, что статистики сдвигов, полученные при сравнении положения, определенного алгоритмом $\aleph_1$ с оптимальным положением (алгоритм $\aleph_0$), будут достаточно представительными.

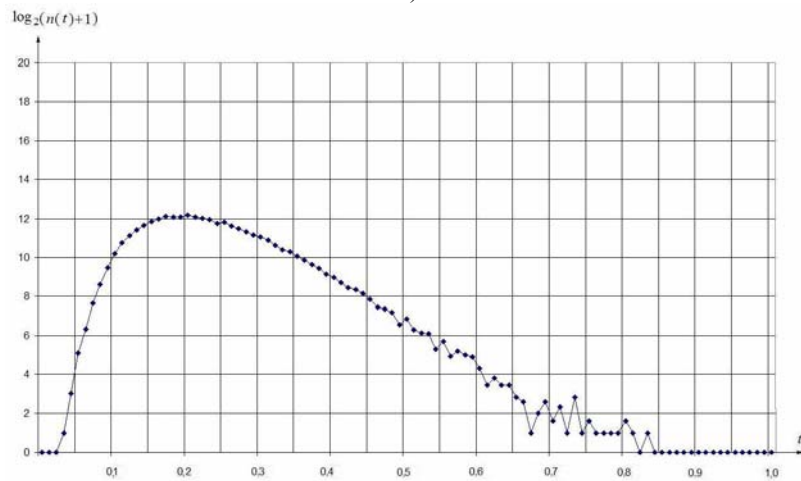
2.3. Анализ отклонений начального положения от идеального

Рассмотрим отклонение от идеального положения в зависимости от t более детально. На Рис. 6 приведено количество пар n_i , отклоняющихся при нулевом положении от идеала на $i=0, 1, 2, 3, 4$ и $i>4$ как функцию от t .

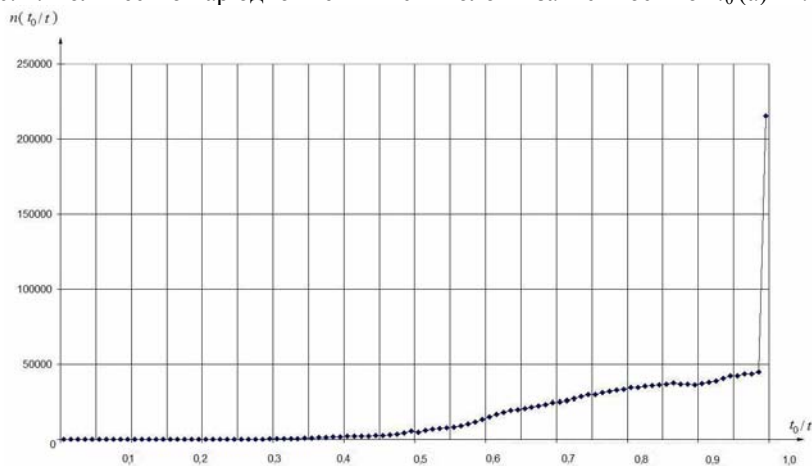
Из графика видно, что для $t<0,13$ все пары символов расположены идеально, но уже при $t>0,22$ количество пар, требующих сдвига на 1 пиксель, превышает количество расположенных идеально. И только при $t>0,3$ появляются пары, требующие сдвига на 2 пикселя.



а)



б)

Рис. 4. Количество пар одноименных символов в зависимости от t_0 (а) и t (б)Рис. 5. Зависимость $n(t_0/t)$

При этом заметим, что хотя алгоритм определения начального положения, как уже отмечалось, не плох, зависимость от t здесь существенна, как показывает Табл. 1.

В ней дано количество пар, в которых отклонение начального положения от идеального равно $i=0, 1, 2$ и $i>2$ для одноименных символов. Мы видим, что даже здесь имеется не так

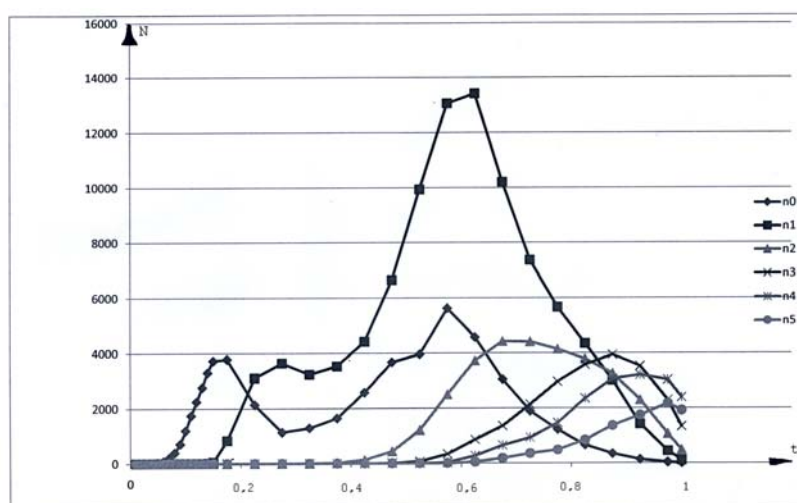
Рис. 6. Количество пар n_i отклоняющихся от идеального положения на i пикселей, как функция t

Табл. 1. Распределение количеств одноименных пар с различными отклонениями начального положения от идеального

t	Отклонение от идеального положения (пикс.)			
	0	1	2	3
0,0 – 0,12	4344	0	0	0
0,12 – 0,31	32158	34271	0	0
0,31 – 0,38	42	8809	38	0
0,38–0,49	0	3328	522	0
0,49–0,66	0	241	463	0
0,66–0,85	0	0	31	7

уж мало случаев, когда начальное положение далеко не идеально, но все они относятся к достаточно большим значениям t .

До сих пор мы использовали для сравнения с идеалом конкретный алгоритм $\aleph_1$. Теперь рассмотрим, как работает в нашем эксперименте теория «в чистом виде». Именно, каждую пару символов сдвинем относительно идеального положения по всем направлениям сначала на 1 пиксель, затем на 2, 3 и 4 и в каждом случае рассмотрим минимальное отклонение по всем направлениям. Таким образом, мы увидим, каково минимальное значение β/α при отклонении от идеала на величину $|a|$ ($|a| = 1, 2, 3$ и более четырех пикселей) по всем 1565565 имеющимся парам символов. Кроме того, для каждого $|a|$ вычислим минимальное значение s_v по формуле (3).

При $|a|=1$ для $t_0>0,12$ всегда находится пара, для которой сдвиг на один пиксель в каком-либо направлении не меняет значения t . Как ни

Табл. 2. Минимальные значения β/α и s_v при отклонении от оптимального положения на 1, 2, 3, 4 и более пикселей

$ a $	$\min \beta/\alpha$	$\min s_v$
1	0,12	2,32
2	0,28	5,81
3	0,44	8,94
4	0,51	11,18

странно, для других значений $|a|$ такие примеры также находятся, но лишь при $t_0>0,5$, выходящих за пределы теоретических оценок (напомним предположение следствия 3: $\beta \ll \alpha$). Результаты эксперимента приведены в Табл. 2.

Данные Табл. 2 хорошо подтверждают основной теоретический вывод. Получив значение $t<0,28$, можно уверенно утверждать, что оптимальное положение находится не более чем в двух пикселях. И даже при $t=0,43$ (расстояние приблизительно равно половине площади символа) взаимное положение отличается от идеального не более чем на три пикселя. В то же время видно, что теоретическая оценка $|a|<s_v$ является слабой. Величина s_v всюду в несколько раз превышает реальный сдвиг от оптимального положения.

Заключение

Мы рассматривали сравнение бинарных растровых изображений, приняв в качестве меры близости β – Хэммингово расстояние между ними. Было доказано, что если β достаточно

мало, то параллельный перенос $a=(h,v)$ одного из объектов, минимизирующий значение β , также не может быть большим. Получена конструктивная оценка s_v для $|a|$.

В рассмотренном эксперименте с образами символов теоретические результаты полностью подтверждаются. Однако теоретическое значение границ возможного сдвига s_v получается несколько завышенным по отношению к реальным, что, впрочем, не удивительно при очень общих теоретических посылках и грубости оценок.

Литература

1. Журавлев Ю. И., Рязанов В. В., Сенько О. В. «Распознавание». Математические методы. Программная система. Практические применения. — М.: Фазис, 2005. — С. 176.
2. Арлазаров В.Л., Котович Н.В., Славин О.А. Адаптивное распознавание // Информационные технологии и вычислительные системы. — 2002, № 4. — С. 11-22.
3. Арлазаров В.Л., Славин О.А., Хованский А.Г. Оценка расстояния между изображениями при параллельном переносе // Доклады Академии Наук. — 2011, Т. 437, № 3. — С. 313-315.

Арлазаров Владимир Львович. Заведующий лабораторией ИСА РАН. В 1961 г. окончил МГУ им. М.В. Ломоносова. Член корреспондент РАН, профессор. Автор более 100 статей и монографий. Область научных интересов: теория графов, распознавания образов, программирование. E-mail: Arl@mail.ru.

Славин Олег Анатольевич. Заведующий лабораторией ИСА РАН. В 1988 г. окончил Московский институт радиотехники, электроники и автоматики. Доктор технических наук. Автор 69 статей и одной монографии. Область научных интересов: распознавания образов, информационные системы. E-mail: OlegSlavin@mail.ru.

Фарсобина Вера Викторовна. Научный сотрудник ИСА РАН. В 1979 г. окончила Московский экономико-статистический институт. Автор 23 статей. Область научных интересов: распознавания образов, вычислительные эксперименты. E-mail: vera@cs.isa.ru.

Хованский Аскольд Георгиевич. Главный научный сотрудник ИСА РАН, В 1970 г. окончил МГУ им. М.В. Ломоносова. Доктор физико-математических наук. Автор 80 статей и двух монографий. Область научных интересов: теория особенностей, комплексный и вещественный анализ, алгебраическая геометрия, дифференциальные уравнения, комбинаторика, геометрия многогранников. E-mail: vera@cs.isa.ru.