



О СУБПОЛОСНОМ АНАЛИЗЕ ИЗОБРАЖЕНИЙ¹

Е. Г. ЖИЛЯКОВ
Н. О. ЕФИМОВ

*Белгородский
государственный
национальный
исследовательский
университет*

*e-mail:
n.o.efimov@gmail.com*

Исследование существующих методов распознавания изображений текста позволяет говорить о том, что в общем случае задача решена только для изображений печатного текста. В случае рукописного текста каждое решение является индивидуальным и зачастую требует многочасового обучения либо вмешательства оператора.

В статье предложена новая решающая процедура обнаружения идентичных фрагментов текста по заданному образцу, реагирующая на различия нормированных трансформант Фурье в заданных частотных интервалах. Исследовано поведение решающей функции на искусственно созданных изображениях. Определены вероятности принятия ошибочных решений.

Ключевые слова: обработка изображений, распознавание текстовых и рукописных изображений, субполосный метод, решающая функция, мера близости, трансформанты Фурье.

К настоящему времени накопилось множество цифровых изображений текста: сканированные книги, статьи, журналы, документы, как офисные, так и персональные (например изображения паспорта, ИНН и др). На многих интернет ресурсах, во избежание копирования и плагиата различного рода документов (например дипломных и диссертационных работ), интересующая пользователя информация представляется в формате изображений. Однако, обработка информации, представленной в подобной форме затруднена невозможностью осуществить такие операции, как поиск по тексту, обнаружение и выделение интересующих фрагментов.

Разработанный подход позволяет осуществлять поиск интересующих фрагментов на изображениях текста (в том числе и рукописного). Так, возможно выделить одно слово на изображении и осуществить поиск этого слова по всему документу, не изменяя формат файла. В случае цифрового документооборота, возможно найти все документы, подписанные одним и тем же человеком. Также представленный подход позволяет искать на изображениях заранее заданные слова, для установления его содержания (например для обнаружения пропаганды терроризма, наркотиков и т.д.). Предлагаемый метод прецедентного распознавания фрагментов изображений основан на использовании субинтервальных методов в частотной области [1, 2].

В общем случае, системы распознавания имеют следующую структуру. Входные данные, подлежащие распознаванию, подаются на вход системы и подвергаются предобработке с целью их преобразования в необходимый для следующего этапа вид или для выделения из них необходимых характерных признаков. Далее на этапе принятия решения над обработанным массивом данных производится ряд вычислений и на основе их результатов формируется ответ, содержащий ожидаемые от системы сведения о входных данных. Содержание входных и выходных данных определяется назначением системы.

Цифровое изображение можно определить как двумерную функцию f_{ik} , где $i=1,2,\dots,M$; $k=1,2,\dots,N$ – координаты в пространстве (конкретно на плоскости), а значение f в любой точке, задаваемой парой координат, называется интенсивностью изображения в этой точке.

Частотным представлением функции f_{ik} называется следующее выражение [3]:

$$f_{ik} = \frac{1}{4\pi^2} \sum_{u=1}^M \sum_{v=1}^N F(u, v) e^{ju(i-1)} e^{jv(k-1)}, \quad (1)$$

$i=1,2,\dots,M; k=1,2,\dots,N$

¹ Исследования проведены в рамках государственного контракта № 14.581.21.0003 с Министерством образования и науки РФ и при поддержке Государственного задания НИУ «БелГУ» (код проекта № 358)



где j – мнимая единица ($j^2 = -1$), аргументы u и v – пространственные частоты, отражающие периодичность (цикличность) изменений исходной функции f_{ik} с изменением аргументов i и k .

Частотной областью называется координатная система, задающая соответствие между аргументами $F(u,v)$ и частотными переменными u и v [4,5].

Прямоугольную область размера $M \times N$, задаваемую при $u = 0, 1, 2, \dots, M - 1$ и $v = 0, 1, 2, \dots, N - 1$, принято называть частотным прямоугольником. Частотный прямоугольник имеет те же размеры, что и исходное изображение.

Так как изображение является функцией не времени, а координаты, используется термин пространственные частоты (волновые числа), которые подразумеваются под словом «частота» далее по тексту.

В качестве весовой функции $F(u,v)$ можно использовать трансформанту Фурье:

$$F(u,v) = \sum_{i=1}^M \sum_{k=1}^N f_{ik} e^{-ju(i-1)} e^{-jv(k-1)}, \quad (2)$$

где $i=1,2,\dots,M; k=1,2,\dots,N$

Величины $F(u,v)$ в уравнениях (1), (2) принято называть коэффициентами разложения Фурье.

На основе равенства Парсеваля [5] энергию изображения можно представить в виде суммы:

$$\sum_{i=1}^M \sum_{k=1}^N f_{ik}^2 = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |F(u,v)|^2 dudv = \sum_{r=1}^{R_1} \sum_{m=1}^{R_2} P_{\Omega_{r,m}} \quad (3)$$

где значение доли энергии $P_{\Omega_{r,m}}$ двумерного сигнала f_{ik} в двумерной частотной области Ω :

$$P_{\Omega_{r,m}} = \frac{1}{4\pi^2} \iint_{(u,v) \in \Omega} |F(u,v)|^2 dudv \quad (4)$$

где интервалы $\Omega_{r,m}$ определяют разбиение частотной области на интервалы.

$$\Omega_{r_1 r_2} : \{ \Omega(u,v) \mid (u \in [\sigma_1, \sigma_r], v \in [\gamma_1, \gamma_r]) \cup (u \in [\sigma_1, \sigma_r], v \in [-\gamma_r, -\gamma_1]) \cup (u \in [-\sigma_r, -\sigma_1], v \in [-\gamma_r, -\gamma_1]) \cup (u \in [-\sigma_r, -\sigma_1], v \in [\gamma_1, \gamma_r]) \},$$

Одна из проблем дискретного преобразования Фурье – приближенное нахождение доли энергии в некотором частотном интервале. В данном случае оно строится, как сумма дискретных значений, попадающих в интервал, что приводит к неточным результатам.

В связи с этим для нахождения точных значений долей энергии изображений в заданных частотных интервалах в работах [1, 6, 7, 8] был разработан и исследован следующий субинтервальный метод.

Если в правую часть представления (4) подставить определение (2), то после преобразований можно получить соотношение [6]:

$$P_{\Omega} = \sum_{i_2=1}^M \left(\sum_{k_1=1}^N \left(\sum_{i_1=1}^M a_{i_1 i_2} f_{i_1 k_1} \right) \left(\sum_{k_2=1}^N b_{k_1 k_2} f_{i_2 k_2} \right) \right), \quad (5)$$

где

$$a_{i_1 i_2} = \begin{cases} \frac{\sin(\alpha_2(i_1 - i_2)) - \sin(\alpha_1(i_1 - i_2))}{\pi(i_1 - i_2)}, & i_1 \neq i_2, \\ \frac{\alpha_2 - \alpha_1}{\pi}, & i_1 = i_2, \end{cases} \quad (6)$$



$$b_{k_1 k_2} = \begin{cases} \frac{\sin(\beta_2(k_1 - k_2)) - \sin(\beta_1(k_1 - k_2))}{\pi(k_1 - k_2)}, & k_1 \neq k_2, \\ \frac{\beta_2 - \beta_1}{\pi}, & k_1 = k_2. \end{cases} \quad (7)$$

Введем матрицы $A=(a_{ii2})$ и $B=(b_{kk2})$, размерности $M \times M$ и $N \times N$, в соответствии с выражениями (3, 4). Матрицы A и B , в соответствии с определением, данным в работах [1, 2, 6, 7], называются субполосными матрицами.

Тогда, выражение (5) позволяет записать формулу вычисления точных значений энергии P_Ω дискретного двумерного сигнала Φ в частотной двумерной области Ω , используя матричные обозначения, в следующем виде:

$$P_\Omega = \text{trac}(A \cdot \Phi \cdot B \cdot \Phi^T), \quad (8)$$

где Φ – исходное изображение, A и B – субполосные матрицы.

Соотношение (8) определяет метод субинтервальной обработки изображений на основе частотных представлений и позволяет для нахождения точных значений энергии двумерного сигнала в любой частотной двумерной области построить вычислительную процедуру, не вычисляя при этом трансформанту Фурье.

Концептуальные основы процедуры идентификации фрагментов изображений текста.

Исходные условия: На изображении текста Φ_0 , размерностью $M \times N$ выделяется прямоугольная область Φ_P , размерностью $MP \times NP$, которая в дальнейшем именуется эталонным фрагментом (индекс p здесь означает прецедент).

Далее из изображения Φ_0 определенным образом формируется множество фрагментов Φ_k , размерностью $MP \times NP$, где k – количество полученных фрагментов, каждый из которых необходимо сопоставить с прецедентом. В случае изображений с текстом процедура поиска фрагментов заключается в перемещении маски по строкам текста, с шагом в 1 пиксель.

На исходном изображении Φ_0 необходимо найти идентичные фрагменты. Для сопоставления фрагментов Φ_k с эталоном Φ_P необходимо иметь решающую функцию, которая задает меру идентичности, значения которой вычисляются по значениям пикселей в сравниваемых фрагментах. Область значений решающей функции, используемой в данной работе – положительные вещественные числа.

Исходная гипотеза H_0 имеет вид: сравниваемые фрагменты идентичны. Для проверки гипотезы используется решающая функция. Гипотеза отвергается, когда значение решающей функции больше порога, величина которого выбирается исходя из требования обеспечения заданного уровня вероятности ошибок первого рода (ложных тревог). Решающая функция должна отвечать принципу максимизации вероятности правильного принятия противоположной гипотезы, когда она верна (минимизации вероятности ошибок второго рода).

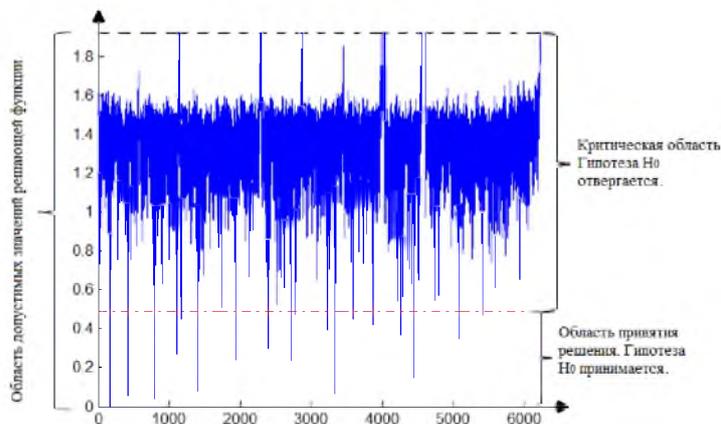


Рис. Множество значений гипотетической решающей функции. Штрихпунктирной линией обозначено пороговое значение



Для определения порога необходимо провести предварительное обучение системы, превышение порога свидетельствует о неидентичности сравниваемых объектов. Обучение проводится на основе единственного образца, который искусственно трансформируется в соответствии с возможными реальными трансформациями.

Решающая функция на основе субполосной меры близости спектров двумерных дискретных сигналов.

В качестве решающей функции предлагается использовать расстояние между нормированными спектрами в заданном частотном интервале:

$$S_{rm} = \frac{1}{2\pi} \iint_{\omega_1 \omega_2 \in \Omega} \left| \frac{F_1(\omega_1 \omega_2)}{\sqrt{P_{rm}^{\Phi_p}}} - \frac{F_2(\omega_1 \omega_2)}{\sqrt{P_{rm}^{\Phi_k}}} \right|^2 d\omega_1 d\omega_2 =$$

$$= \frac{1}{2\pi} \iint_{\omega_1 \omega_2 \in \Omega} \left| \left(\frac{F_1(\omega_1 \omega_2)}{\sqrt{P_{rm}^{\Phi_p}}} \right)^2 - 2 \frac{F_1(\omega_1 \omega_2) F_2(\omega_1 \omega_2)}{\sqrt{P_{rm}^{\Phi_p} P_{rm}^{\Phi_k}}} + \left(\frac{F_2(\omega_1 \omega_2)}{\sqrt{P_{rm}^{\Phi_k}}} \right)^2 \right| d\omega_1 d\omega_2 \quad (9)$$

где F_1, F_2 – трансформанты Фурье эталонного и сравниваемого объектов соответственно,

$$P_{rm}^{\Phi_p} = tr(A_r \Phi_p B_m \Phi_p^T) \quad (10)$$

$$P_{rm}^{\Phi_k} = tr(A_r \Phi_k B_m \Phi_k^T) \quad (11)$$

Где A_r, B_m – субполосные матрицы, элементы которых вычисляются согласно выражениям (6)(7)

В основе этой меры используются нормированные субполосные коэффициенты корреляции в частотном интервале:

$$p_{rm}^{pk} = \frac{1}{2\pi} \iint_{\omega_1 \omega_2 \in \Omega} \frac{F_1(\omega_1 \omega_2) F_2(\omega_1 \omega_2)}{\sqrt{P_{rm}^{\Phi_p} P_{rm}^{\Phi_k}}} d\omega_1 d\omega_2 \quad (12)$$

Для сопоставления фрагментов цифровых изображений в данной работе используется мера близости следующего вида:

$$s_{rm}^{pk} = 2 \left(\frac{1}{R_1 R_2} \sum_{j=1}^{R_2} \sum_{i=1}^{R_1} (1 - p_{rm}^{pk}) \right) \quad (13)$$

где

$$p_{rm}^{pk} = \frac{tr(A_r \Phi_p B_m \Phi_k^T)}{\sqrt{P_{rm}^{\Phi_p} P_{rm}^{\Phi_k}}} \quad (14)$$

Выражение 8 является используемой в данной работе мерой идентичности фрагментов.

Также, для проведения сравнительного анализа, предлагается использовать относительную евклидову норму δ с коэффициентом a :

$$\delta = \sqrt{\frac{\sum_{i=1}^M \sum_{k=1}^N (\tilde{\Phi}_{kik} - a \Phi_{pik})^2}{\sum_i \sum_k \Phi_{pik}}} \quad (15)$$

где Φ_{pik} – эталонный фрагмент;
 $\tilde{\Phi}_{kik}$ – фрагмент, сравниваемый с эталоном;
 a – коэффициент, который вычисляется:



$$a = \frac{\sum_i \sum_k \bar{\Phi}_{k_i} \Phi_{P_k}}{\sum_i \sum_k \Phi_{P_k}^2} \quad (16)$$

Основная гипотеза (H_0 – сравниваемые фрагменты идентичны) отвергается при выполнении неравенства:

$$s_{tm}^{pk} > h_p \quad (17)$$

где h_p – граница критической области, которая определяется на этапе обучения.

Для изучения предлагаемой решающей функции был проведен ряд экспериментов, наиболее значимые из результатов представлены ниже.

Эксперимент 1.

Целью вычислительного эксперимента является исследование эффективности разработанного алгоритма для изображений разного размера.

Таблица 1

Значения вероятности ошибок первого и второго рода при распознавании объектов с использованием субполосной меры близости на текстовых изображениях различных размеров

	Размер изображения						
	1000x714	900x638	800x567	700x500	600x428	500x354	400x283
Вероятность ошибки первого рода	0	0	0	0	0	0.02	0
Вероятность ошибки второго рода	0	0	0	0	0	0	0

Таблица 2

Значения вероятности ошибок первого и второго рода при распознавании объектов с использованием евклидовой нормы с коэффициентом на текстовых изображениях различных размеров

	Размер изображения						
	1000x714	900x638	800x567	700x500	600x428	500x354	400x283
Вероятность ошибки первого рода	0	0	0.02	0.02	0.02	0.02	0.02
Вероятность ошибки второго рода	0	0.0003	0.0003	0.001	0.0012	0.002	0.03

Из таблицы 1 видно, что предлагаемый субполосный метод инвариантен к изменению размеров изображения текстового фрагмента, и используемого шрифта. При использовании евклидовой нормы вероятность ошибок второго рода выше, так как данная мера идентичности зачастую относит фрагменты, соседние с эталонным к заданному классу.

Эксперимент 2. Целью вычислительного эксперимента является исследование эффективности предлагаемого метода для эталонных фрагментов разной длины.

Таблица 3

Значения вероятности ошибок первого и второго рода при распознавании объектов различной длины с использованием субполосной меры близости на текстовых изображениях

	Эталонный фрагмент					
	ф	фр	фра	фраг	фрагм	фрагмент
Вероятность ошибки первого рода	0.0067	0	0	0	0	0
Вероятность ошибки второго рода	0	0	0	0	0	0

Для субполосной меры близости в результате прецедентного распознавания буквы ф было обнаружено 150 объектов из 151. 100 объектов в составе слова «фрагмент», и 50 в сос-



таве других слов. В результате прецедентного распознавания эталона «фр» было обнаружено 106 объектов из 106. 100 объектов в составе слова «фрагмент», и 6 в составе других слов.

Таблица 4

Значения вероятности ошибок первого и второго рода при распознавании объектов различной длины с использованием евклидовой нормы с коэффициентом

	Эталонный фрагмент					
	ф	фр	фра	фраг	фрагм	фрагмент
Вероятность ошибки первого рода	0.14	0.09	0.09	0.07	0.01	0
Вероятность ошибки второго рода	0.17	0.32	0.45	0.5	0.17	0.32

Мера близости, основанная на относительной евклидовой норме с коэффициентом менее эффективна, если эталонный фрагмент имеет малый размер (например одна буква), что подтверждается высокой вероятностью ошибок первого и второго рода (таблица 4).

Эксперимент 3. Целью вычислительного эксперимента является исследование решающей функции при обработке изображений рукописного текста.

Таблица 5

Вероятности ошибок первого и второго рода для разных значений порога. Используется субполосная мера близости. Прецедентом является первый фрагмент. max-наибольшее значение эталонной выборки для данного фрагмента. Общее число фрагментов равно 1 308 150

	Численное значение порога						
	max	0.984*max	0.872*max	0.96*max	0.936*max	0.904*max	0.88*max
Количество отнесенных к заданному классу объектов	6888	4335	2919	2020	952	346	165
Количество верно отнесенных объектов (всего 82 объекта)	82	81	78	77	68	48	39
Количество неверно отнесенных объектов	6788	4254	2841	1943	884	298	126
Вероятности ошибки первого рода	0	0.012	0.048	0.06	0.17	0.41	0.524
Вероятность ошибки второго рода	0.0053	0.00325	0.00217	0.00149	0.00068	0.00023	0.0001

Таблица 6

Вероятности ошибок первого и второго рода для разных значений порога. Используется евклидова норма с коэффициентом. Прецедентом является первый фрагмент. max-наибольшее значение эталонной выборки для данного фрагмента. Общее число фрагментов равно 1 303 400

	Численное значение порога						
	max	0.984*max	0.872*max	0.96*max	0.936*max	0.904*max	0.88*max
Количество отнесенных к заданному классу объектов	407882	386689	371550	342469	329655	296956	275574
Количество верно отнесенных объектов (всего 82 объекта)	82	80	79	78	74	70	61
Количество неверно отнесенных объектов	407800	386609	37147	342391	329581	296886	275513
Вероятности ошибки первого рода	0	0.02	0.04	0.05	0.1	0.15	0.26
Вероятность ошибки второго рода	0.31	0.295	0.284	0.26	0.251	0.226	0.21



На основании представленных в таблицах 5 и 6 данных, можно сделать следующие выводы: задача распознавания фрагментов рукописного текста является более трудоемкой, нежели задача распознавания печатного текста. Разработанный алгоритм, использующий субполосную меру близости показывает меньшие ошибки первого и второго рода, чем относительная евклидова норма. Относительная евклидова норма с коэффициентом подвержена ошибкам второго рода.

Проведенные исследования позволяют говорить о высокой эффективности предлагаемого метода прецедентного распознавания изображений сканированного печатного текста, что подтверждается данными из раздела. Предлагаемый алгоритм инвариантен к изменению размеров изображения текстового фрагмента, и используемого шрифта; осуществляет распознавание эталонных фрагментов с близкой к нулю вероятностью ошибок первого и второго рода. В условиях интенсивного зашумления исходного изображения и неидеального эталона алгоритм превосходит человеческие возможности.

В том случае, если на вход системы распознавания подается изображение сканированного рукописного текста возникают ошибки первого и второго рода. Это связано с особенностями написания слов. В процессе письма изменяется угол написания, нажим, ширина слов, этим и объясняются ошибки алгоритма. Возможно достижение лучшего результата, на изображениях рукописного текста, при проведении дальнейших исследований.

Литература

1. Жилияков, Е. Г., Черноморец А.А. Вариационные алгоритмы анализа и обработки изображений на основе частотных представлений: [Моногр] / – М.: ГиК, 2009. – 146 с.
2. Жилияков, Е.Г. Метод определения точных значений долей энергии изображений в заданных частотных интервалах / Е.Г. Жилияков, А.А. Черноморец, И.В. Лысенко – М.: Вопросы радиоэлектроники, 2007. – Вып. 4. – С.115-123.
3. Семенов, Ю.А. Алгоритмы телекоммуникационных сетей. В 3 частях. Часть 1. Алгоритмы и протоколы каналов и сетей передачи данных [Текст] / Ю.А. Семенов. – М.: Бином. Лаборатория знаний, 2007. – 640 с.
4. Рабинер, Л. Теория и применение цифровой обработки сигналов [Текст] / Л. Рабинер, Б. Гоулд. – М.: Мир, 1978.
5. Сергиенко, А.Б. Цифровая обработка сигналов [Текст] / А.Б. Сергиенко. – СПб.: Питер, 2002. – 603с.
6. Жилияков, Е.Г. Частотный анализ речевых сигналов / Е.Г. Жилияков, Е.И. Прохоренко / М.: Научные ведомости Белгородского государственного университета, 2006. – №2 (3.118), выпуск 3. – С. 201-208. – (Серия: информатика и прикладная математика).
7. Черноморец А.А., Иванов О.Н. Метод анализа распределения энергий изображений по заданным частотным интервалам // Научные ведомости БелГУ. Серия: История. Политология. Экономика. Информатика. 2010. 8. Черноморец А.А., Жилияков Е.Г., Голощапова В.А., Болгова Е.В. Оценка эффективности субполосного внедрения данных в изображение // Научные ведомости БелГУ. Серия: История. Политология. Экономика. Информатика. 2014. №8-1 (179).
8. Черноморец А.А., Жилияков Е.Г., Голощапова В.А., Болгова Е.В. Оценка эффективности субполосного внедрения данных в изображение // Научные ведомости БелГУ. Серия: История. Политология. Экономика. Информатика. 2014. №8-1 (179).

ABOUT SUBBAND IMAGE ANALYSIS

E. G. ZHILYAKOV
E. N. EFIMOV

*Belgorod state
national-research
university*

*e-mail:
n.o.efimov@gmail.com*

The research of existing methods of identification scanning text images allow to draw a conclusion, what in general case problem solve only for printed text images. In case of handwritten text each solution is individual, and frequently need a long time education or operator inspection.

In the article introduce new decision procedure of precedential identical fragment detection, which react on normalized Fourier transform differences in defined frequency slots. The behavior of decision function was researched on the artificial image. Probability of erroneous decision-making was evaluated. In comparison with Euclidean norm, substrip measure have a greater reactivity by similar variability.

Key words: image processing, identification images of printed and handwritten text, substrip method, decision procedure, measure of proximity, Fourier transform.