

УДК 004.932.72'1; 004.93'14

## МЕТОД ИДЕНТИФИКАЦИИ ПРОТЯЖЕННЫХ МИКРООБЪЕКТОВ НА СЛАБОКОНТРАСТНЫХ ИЗОБРАЖЕНИЯХ

Д.А. БУШЕНКО, Р.Х. САДЫХОВ

*Белорусский государственный университет информатики и радиоэлектроники  
П. Бровка, 6, Минск, 220013, Беларусь*

*Поступила в редакцию 1 апреля 2010*

Представлен новый метод извлечения и идентификации протяженных микрообъектов, использующий модифицированный алгоритм "Adaptive Fuzzy Select Tool" для извлечения объектов. Показано, что алгоритм позволяет не только качественно извлечь достаточное количество пикселей протяженного объекта, но также и отделить его от прочих объектов, его пересекающих. Для сравнения двух протяженных объектов здесь предлагается использовать корреляцию гистограмм тона.

*Ключевые слова:* цифровая обработка изображений, протяженные объекты, идентификация.

### Введение

Идентификация — одна из важнейших задач интеллектуальной обработки данных. Поскольку не существует единого подхода к идентификации, для каждой конкретной задачи разрабатываются специальные модификации классических алгоритмов. Очевидно, что для идентификации протяженных микрообъектов на слабоконтрастных изображениях также требуется разработать специальную технологию, способную решить эту задачу с максимальной точностью.

Задача идентификации протяженных микрообъектов на слабоконтрастных изображениях возникает чаще всего в криминалистике в виде поиска идентичных волокон на образцах, снятых с одежды жертвы и подозреваемых. Эксперту необходимо просмотреть десятки пленок с образцами при помощи микроскопа, а это часто занимает очень значительное время — дни и даже недели, а иногда и месяцы, что обуславливает необходимость автоматизации процесса идентификации.

Для идентификации текстильных волокон была разработана цифровая система [1], использующая экономичный сканер бизнес-класса Epson Perfection 3200 для ввода образцов, настольный компьютер для обработки данных и принтер для вывода отчетов. Указанная система работает по следующему алгоритму.

1. Пользователь выбирает точку на изображении, которая, по его мнению, принадлежит текстильному волокну.

2. Система предпринимает попытку извлечь текстильное волокно по указанной точке изображения [2, 3].

3. В случае, если извлечение прошло успешно (т.е. данная точка действительно принадлежала текстильному волокну), система предлагает ввести параметры поиска.

4. Пользователь вводит максимально допустимое значение в процентах, на которое могут отличаться искомые волокна от указанного пользователем образца.

5. Система ищет все точки всех изображений, которые совпадают с точками образца.

6. Для каждой найденной точки система предпринимает попытку извлечь волокно.

7. Система строит отчет по всем успешно извлеченным волокнам, идентичным указанному образцу.

Здесь нужно отметить, что найденные точки могут находиться как на текстильных волокнах, так и на ореолах вокруг этих волокон, возникающих при сканировании изображений, а также на фоне изображения, на посторонних объектах и даже на объектах, пересекающих искомое текстильное волокно.

В задаче идентификации протяженных микрообъектов на слабоконтрастных изображениях присутствуют две сложные проблемы. Первая – это отделение извлекаемых протяженных объектов от прочих объектов, их пересекающих. Вторая – способ сравнения двух протяженных объектов, который был бы инвариантен к любым геометрическим искажениям тела объекта. Для решения указанных проблем были проведены специальные исследования и разработаны модифицированные алгоритмы.

### Оценка качества извлеченных объектов

Следующие параметры позволяют разработать технологию извлечения протяженных объектов с отделением их от прочих объектов, их пересекающих.

1. Ширина гистограммы тона.
2. Минимальное среднее отклонение гистограммы тона.
3. Количество пересечений скелета объекта.
4. Длина скелета объекта.

Одиночное монохромное текстильное волокно имеет очень узкую гистограмму тона. Теоретически она должна представлять собой единственный пик на одном значении тона. Однако из-за различных искажений, возникающих в процессе сканирования изображения, гистограмма тона объекта становится несколько шире. Типичная гистограмма тона одиночного монохромного текстильного волокна показана на рис. 1. Здесь и далее значения тона гистограммы принадлежат отрезку [0; 359].

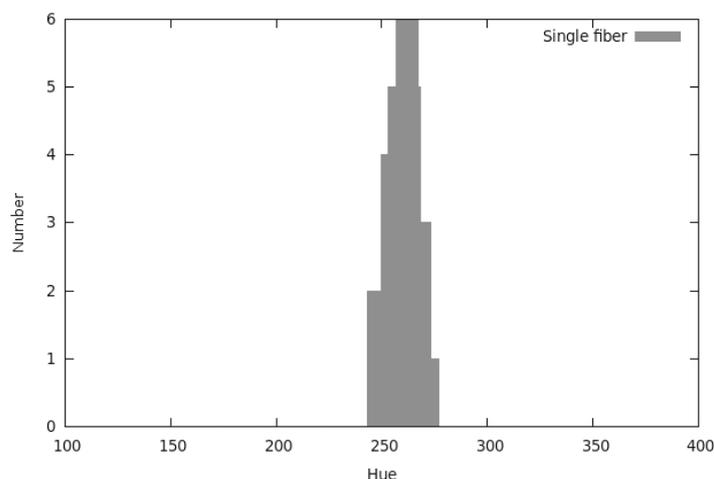


Рис. 1. Гистограмма тона одиночного текстильного волокна

Одиночное текстильное волокно, извлеченное некачественно, имеет, как правило, ширину около десяти пикселей и включает в себя точки, как самого волокна, так и точки ореола и даже фона. Пример гистограммы текстильного волокна, извлеченного некачественно, представлен на рис. 2.

Левая область гистограммы между значениями 240 и 290 принадлежит пикселям фона. Наиболее высокий пик со значением тона 270 принадлежит ореолу вокруг волокна. Остальные значения гистограммы принадлежат точкам волокна.

Исследуя свойства гистограммы тона объекта далее, можно утверждать, что чем шире эта гистограмма, тем хуже извлечен данный протяженный объект. Вот почему указанный параметр можно использовать как один из формальных критериев качества извлеченного объекта.

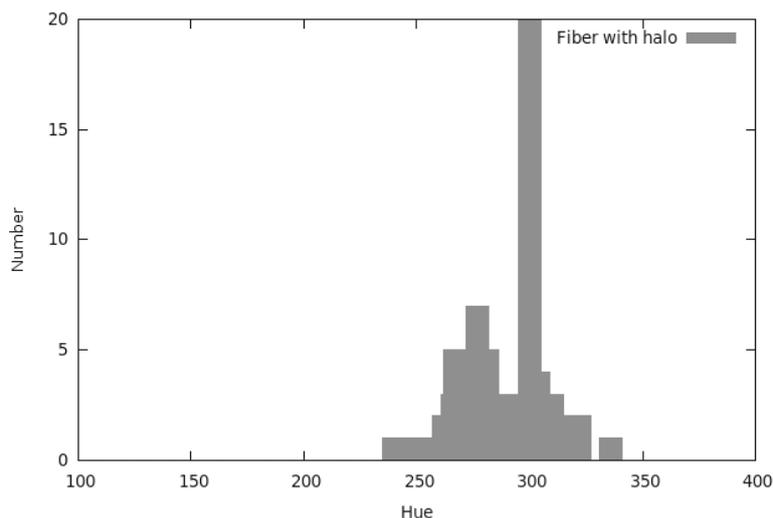


Рис. 2. Гистограмма тона одиночного текстильного волокна с ореолом и фрагментами фона

Определим ширину гистограммы тона объекта ( $W(Hue)$ ) как количество значений тона, присутствующих на данном объекте:

$$W(Hue) = \sum_{i=0}^{359} \begin{cases} 1, \text{ если } N(hue_i) > 0, \\ 0, \text{ иначе.} \end{cases} \quad (1)$$

Следующий интересующий нас параметр — само значение тона протяженного объекта. Значение тона — элемент дискретного конечного циклического кольца чисел без минимального и максимального значения, а также без нуля и единицы [4]. Вследствие этого факта невозможно вычислить среднее значение тона объекта и, таким образом, его дисперсию [5]. Дисперсия показывает, как множество данных распределяется вокруг его среднего значения. Вместо дисперсии и среднеквадратического отклонения здесь предлагается использовать специальный параметр — минимальное среднее отклонение тона ( $d_{min}$ ).

$$d_j = \sum_{i=0}^{359} \left( \sqrt{(h_i - h_j)^2} \right) \frac{n_i}{N} \quad (2)$$

$$d_{min} = \min \{d_0, d_1, \dots, d_{359}\} \quad (3)$$

Здесь  $h$  — значение тона из  $[0; 359]$ ,  $n$  — количество пикселей соответствующего тона.

При вычислении разности  $(h_i - h_j)$  следует учитывать циклическую природу значений тона. Поэтому

$$(h_i - h_j) = \begin{cases} (h_i - h_j), \text{ если } \sqrt{(h_i - h_j)^2} < 180, \\ 359 - (h_i - h_j), \text{ иначе.} \end{cases} \quad (4)$$

Наиболее ценное свойство минимального среднего отклонения заключается в его возможности корректно оперировать элементами в дискретном циклическом кольце чисел без минимального и максимального элементов, а также без нуля и единицы. Аналогично параметру ширины гистограммы тона, чем меньше минимальное среднее отклонение, тем лучше извлечен исследуемый объект.

Одиночное текстильное волокно может быть изогнуто как угодно, в том числе пересекать себя же или другие объекты. Как правило, волокно очень редко пересекает себя или соседние объекты. Поэтому если в извлеченном объекте присутствует множество пересечений скелета, это означает, что объект извлечен некачественно. Таким образом, количество пересечений скелета должно быть минимальным.

Последний параметр — длина скелета объекта. В случае если длина скелета извлеченного объекта невелика, то невозможно будет выполнить статистический анализ выборки пикселей, построенной на основе этого объекта. Поэтому в общем случае чем больше длина скелета,

тем лучше. Однако наиболее частый случай, когда длина скелета извлеченного объекта равна 50. Наименьший возможный размер выборки для анализа статистическими методами — 30 [6–8].

Учитывая рассмотренные выше параметры протяженных объектов становится возможным разработать специальную функцию оценки качества извлеченного протяженного объекта. Эта функция должна:

- 1) расти с ростом длины скелета объекта;
- 2) уменьшаться с увеличением количества пересечений скелета;
- 3) уменьшаться с ростом ширины гистограммы тона;
- 4) уменьшаться с ростом минимального среднего отклонения.

В экспериментах по извлечению текстильных волокон со слабоконтрастного зашумленного изображения использовалась формула, в общем виде представленная ниже.

$$R(obj_i) = \frac{a N_f}{\sqrt{b (W(Hue))^x + c (d_{min})^y + d (n_c)^z}}. \quad (5)$$

Здесь  $R$  — относительное значение рейтинга объекта, оценка качества, с которым он был извлечен;  $N_f$  — нечеткая лингвистическая переменная, представляющая собой желаемую длину скелета извлекаемого объекта;  $W(Hue)$  — ширина гистограммы тона;  $d_{min}$  — минимальное среднее отклонение тона;  $n_c$  — количество пересечений скелета;  $a, b, c, d, x, y, z$  — коэффициенты, которые необходимо калибровать относительно каждого типа оборудования ввода изображений. Например, в экспериментах по извлечению текстильных волокон использовался сканер Epson Perfection 3200 на разрешении 2400 dpi. При этом результаты достаточно высокого качества были получены при  $a=b=c=d=1$  и  $x=y=z=3$ . Таким образом, формула (5) приобретает вид

$$N_f = \begin{cases} N^2, & \text{если } N < 50, \\ 50^2, & \text{иначе,} \end{cases} \quad (6)$$

$$n_{cf} = \begin{cases} \frac{10n_c}{N}, & \text{если } n_c > 10, \\ 0, & \text{иначе,} \end{cases} \quad (7)$$

$$R(obj_i) = \frac{N_f}{\sqrt{W(Hue)^3 + (d_{min})^3 + (n_{cf})^3}}. \quad (8)$$

### Извлечение протяженных объектов

Качественное извлечение протяженных объектов является важной задачей, так как от нее зависит точность идентификации этих объектов. Вот почему существующие алгоритмы извлечения объектов должны быть модифицированы таким образом, чтобы показывать максимальное качество на протяженных объектах.

Рассмотрим алгоритм, показывающий наиболее качественные результаты при извлечении монохромных текстильных волокон на слабоконтрастных изображениях. Это алгоритм "Fuzzy Select Tool", встроенный в популярный графический редактор с открытыми исходными кодами GIMP. Алгоритм работает следующим образом.

1. Пользователь выбирает максимально допустимое отклонение от эталона.
2. Пользователь указывает пиксель на изображении.
3. Система получает указанный пиксель.
4. Система высчитывает интервал, в который должны попадать искомые пиксели, как:

$$\begin{aligned} R &\in [R_u - d; R_u + d], \\ G &\in [G_u - d; G_u + d], \\ B &\in [B_u - d; B_u + d], \end{aligned} \quad (9)$$

где  $R_u$ ,  $G_u$ ,  $B_u$  — красная, зеленая и синяя составляющие цвета эталонного пикселя;  $d$  — максимально допустимое отклонение.

5. Система извлекает все пиксели, связанные геометрически с эталонным, которые попадают в вычисленный ранее интервал.

На рис. 3 слева показаны два пересекающихся текстильных волокна, а справа — их бинаризованный полутоновый скелет. На изображении скелета также помечены три окружности, центры которых представляют собой точки интереса. Нижняя точка интереса попала на область между текстильным волокном и фоном, две сверху — на сами текстильные волокна.



Рис. 3. Слева — изображение с пересекающимися текстильными волокнами; справа — их бинаризованный полутоновый скелет

Результаты извлечения протяженных объектов с использованием алгоритма "Fuzzy Select Tool" и максимальным допустимым отклонением, равным 20, показаны на рис. 4.

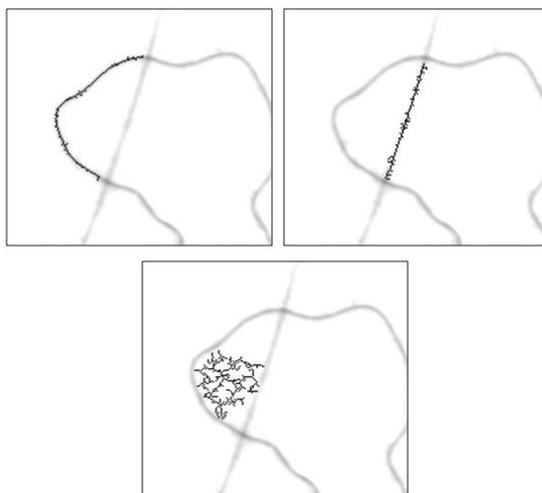


Рис. 4. Извлечение объектов алгоритмом "Fuzzy Select Tool"

Две верхние точки интереса позволили извлечь объекты, близкие к протяженным. Единственный их недостаток — достаточно большое количество мелких ложных ветвей, примыкающих к основному скелету. Однако нижняя точка интереса, не попавшая точно на волокно, стала причиной сбоя алгоритма и извлечения сложного объекта, не являющегося протяженным.

Указанный выше недостаток стал причиной разработки специальной модификации алгоритма "Fuzzy Select Tool", позволяющей адаптировать минимально допустимую величину отклонения в соответствии с качеством извлеченного объекта. Алгоритм "Adaptive Fuzzy Select Tool" работает следующим образом.

1. Для каждого значения максимально допустимого отклонения из заданного интервала (по умолчанию это интервал от 2 до 25) выбрать очередное значение:

- а) извлечь объект с текущим значением максимально допустимого отклонения;
- б) оценить извлеченный объект при помощи функции рейтинга  $R(obj)$ .

2. Из множества извлеченных объектов выбрать тот, рейтинг которого наибольший.

Результаты извлечения протяженных объектов алгоритмом "Adaptive Fuzzy Select Tool" показаны на рис. 5. Некоторые параметры извлеченных объектов представлены в табл. 1.

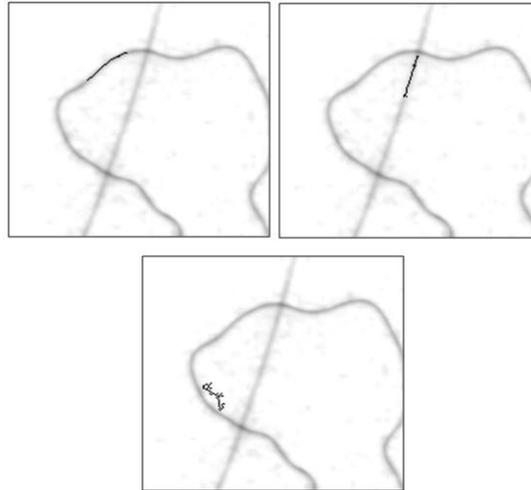


Рис. 5. Извлечение объектов алгоритмом "Adaptive Fuzzy Select Tool"

Таблица 1. Параметры протяженных объектов, извлеченных алгоритмом "Adaptive Fuzzy Select Tool"

Объект	Максимально допустимое отклонение	$R(obj)$
1	13	286,00
2	6	402,95
3	9	100,21

Как показано в табл. 1, последний объект, извлеченный из нижней точки интереса, имеет относительно невысокое значение рейтинга ( $R(obj)=100,21$ ). Такие объекты могут быть вынесены из результатов поиска на основе некоторого порогового значения рейтинга. Конкретные алгоритмы получения порогового значения рейтинга в данный момент находятся на стадии исследования. Из двух верхних точек интереса протяженные объекты были извлечены более качественно, так как в них отсутствуют ложные ветви, примыкающие к скелету. Более высокое качество извлечения протяженных объектов было достигнуто за счет того, что значение максимально допустимого отклонения высчитывалось отдельно для каждого объекта.

### Сравнение протяженных объектов

Существует множество способов сравнения протяженных объектов. Однако для идентификации текстильных волокон необходима функция, которая точно покажет, насколько близки два сравниваемых волокна. Здесь предлагается использовать функцию корреляции [8] для сравнения гистограмм тона текстильных волокон:

$$r(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (10)$$

На рис. 6–8 сравниваются гистограммы тона различных текстильных волокон, а в табл. 2 показаны их числовые характеристики и значение корреляции волокон 2–4 с волокном 1.

Корреляция гистограмм тона хорошо подходит как инструмент сравнения двух текстильных волокон, так как она инвариантна к любым геометрическим искажениям объектов, а также точно показывает, насколько похожи или отличаются два сравниваемых текстильных волокна.

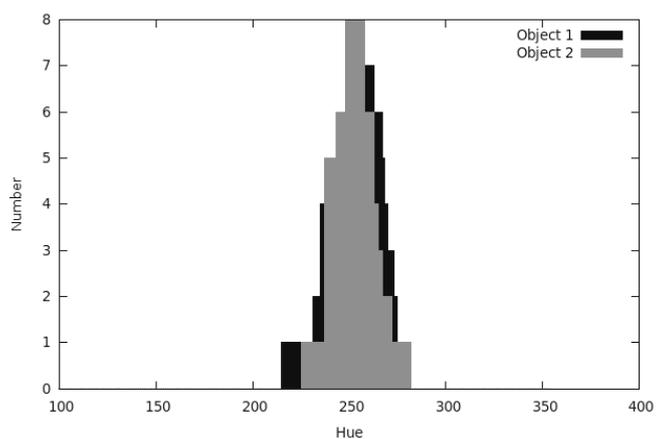


Рис. 6. Гистограммы тона волокна 1 (черное) и 2 (серое)

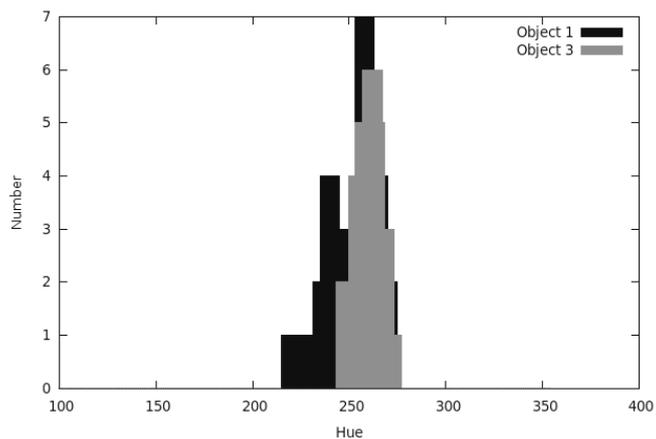


Рис. 7. Гистограммы тона волокна 1 (черное) и 3 (серое)

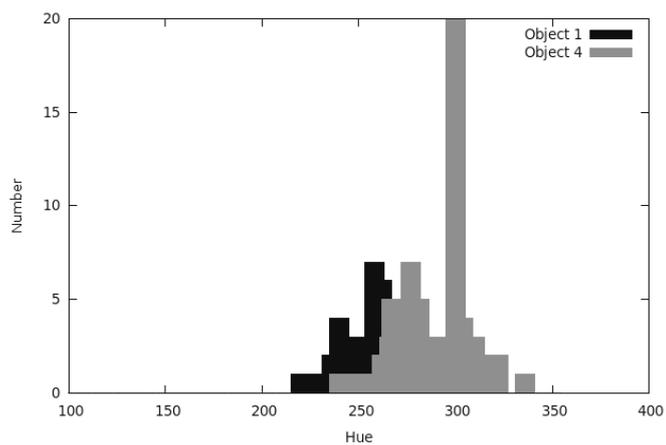


Рис. 8. Гистограммы тона волокна 1 (черное) и 4 (серое)

Таблица 2. Корреляция гистограмм тона

$N$	Среднее значение	Среднее отклонение	Корреляция
1	254	11	1,00
2	260	5	0,84
3	252	8	0,56
4	286	17	0,09

## Заклучение

Идентификация протяженных объектов применяется в различных областях деятельности. Одной из таких областей является криминалистика, в которой важно правильно идентифицировать текстильные волокна. При автоматизации процесса идентификации текстильных волокон возникает ряд проблем, среди которых наиболее сложными можно считать извлечение протяженных объектов с одновременным отделением их от прочих объектов, а также сравнение извлеченных объектов. В данной работе предлагается специально модифицированный алгоритм "Adaptive Fuzzy Select Tool" для извлечения объектов и корреляция гистограмм тона для их сравнения. Указанные алгоритмы были использованы при построении цифровой системы идентификации текстильных волокон.

## METHOD OF EXTENDED OBJECTS IDENTIFICATION ON LOW-CONTRAST IMAGES

D.A. BUSHENKO, R.K. SADYKHOV

### Abstract

Nowadays there is a demand of intelligent systems which are able to perform identification of color low-contrast extended objects such as textile fibers. Since this is a complex problem with no common way to solve it, detailed investigations are needed here. The paper presents a new method to extract and identify the extended objects such as textile fibers. This method uses the modified "Adaptive Fuzzy Select Tool" algorithm for extracting the extended objects and splitting them from the alien objects. Finally, the paper discusses the hue histograms correlation as a tool to compare two extended objects.

### Литература

1. *Sadykhov R., Bushenko D.* // MIPRO'2009. Opatija, Croatia. 2009. P. 297–302.
2. *Bushenko D., Sadykhov R.* // IMCSIT'2009. Mragowo, Poland. 2009. P. 25–32.
3. *Bushenko D., Sadykhov R.* // ICNNAI'2008. Minsk, Belarus. 2008. P. 182–185.
4. *Blahut R.E.* Fast Algorithms for Digital Signal Processing. Addison-Wesley, Wokingham, 1989.
5. *Fanchi J.R.* Math Refresher for Scientists and Engineers, 3<sup>rd</sup> Edn. Willey-Interscience, Hoboken, 2006.
6. *Rumsey D.* Statistics for Dummies. Willey Publishing, Indianapolis. 2003.
7. *Sheldon R.M.* Introduction to Probability and Statistics for Engineers and Scientists, 3<sup>rd</sup> Edn. Elsevier Academic Press, Burlington, 2004.
8. *Bendat J.S., Piersol A.G.* Random Data Analysis and Measurement Procedures. John Willey & Sons, New York, 1986.