

**Построение эффективных алгоритмов
обработки и распознавания изображений гистологических срезов
на основе эволюционных моделей**

У статті запропонований новий підхід до побудови ефективних алгоритмів сегментації кольорових зображень гістологічних зрізів на основі еволюційних моделей. Алгоритм обробки зображення у вигляді окремої хромосоми представлений спрямованим ациклічним графом, де нетермінальні вузли являються стандартними операторами обробки зображень, а кінцеві вузли - вихідними площинами зображення. Якість процесу сегментації визначається за допомогою оцінки цільової функції. Розроблено функціональну та термінальну множини, проблемно-орієнтовані оператори кросингвера і мутації. Зазначений підхід підтвердив свою ефективність у процесі експериментів на базі лабораторії фундаментальних досліджень ІНВХ АН України при проведенні гістологічних досліджень.

Введение.

Существует ряд заболеваний, которые характеризуются наличием различного рода новообразований. Для диагностирования и прогнозирования развития новообразований обязательным является гистологическое исследование атипичных клеточных структур. Различные по тканевой принадлежности опухоли строго специфичны на тканевом, клеточном и на структурном уровне. Однако имеется ряд морфологических и цитологических методов, позволяющих определить доброкачественность или злокачественность исследуемой опухоли: размеры клеток, размеры ядра клеток, ядерно-цитоплазматическое соотношение и т.д [1].

Применение автоматизированных систем обработки и анализа изображений (АСОАИ) позволяет существенно повысить достоверность обнаружения изменений внутренних органов при онкологических заболеваниях и ускорить процесс обработки препаратов, а также расширяет возможности профилактики и предупреждения злокачественных новообразований за счет повышения уровня автоматизации обработки медицинских изображений. Основной функцией АСОАИ является извлечение требуемой, значимой информации, определяющей эффективность работы систем обработки и распознавания медицинских изображений (СОРМИ).

Масштабы применения методов и средств цифрового анализа изображений в СОРМИ в настоящее время ограничены из-за низкой точности классификации и отсутствия эффективного автоматического подбора параметров сегментации. Для преодоления этих недостатков необходимо использование эффективных методов и средств цифрового анализа изображений.

Краткая постановка задачи построения эффективных алгоритмов обработки и распознавания изображений гистологических срезов.

Задача построения эффективного алгоритма сегментации изображений заключается в том, чтобы по изображениям обучающего множества G определить такую последовательность операций обработки изображений с соответствующим набором параметров, при которой критерий оценки качества распознавания $F(G)$ стремится к экстремуму.

Входное изображение I_k подвергается сегментации с помощью программы S с набором параметров P_j^S , в результате работы которой получается набор изображений M_{kj}^S :

$$M_{kj}^S = \text{segm}(S, P_j^S, I_k). \quad (1)$$

Поскольку сегментация изображений гистологических срезов подразумевает определение положительных и отрицательных клеток (представляющих здоровые и раковые клетки ткани), имеющих различные цветовые характеристики, то перед выполнением программы S исходное изображение I_k необходимо разбить на набор цветовых плоскостей:

$$I_k = [C_k, Red_k, Green_k, Blue_k, H_k, Light_k, S_k, L_k, a_k, b_k, u_k, v_k]. \quad (2)$$

Каждая цветовая плоскость соответствует значениям пикселя изображения в одной из цветовых схем (ЦС): RGB, HLS, Lab и Luv.

Многомерную пороговую сегментацию цветных изображений гистологических срезов для каждого из классов клеток, можно представить в виде последовательного выполнения двух операций обработки изображения:

1. $M_{k1}^s = \text{segm}(S, P_1^s, I_k)$, где $S = \text{ClipLo}$ – программа порогового ограничения, при котором все значения пикселей в выбранной ЦС ниже значения P_1^s устанавливаются в 0, остальные сохраняют свои значения.

2. $M_{k2}^s = \text{segm}(S, P_1^s, I_k)$, где $S = \text{ClipHi}$ – отсекает любой пиксель, значение которого превышает значение параметра P_1^s .

Построение эффективных алгоритмов сегментации на основе эволюционных моделей.

Для поиска полиномиальной функции, аппроксимирующей функцию-классификатор, предложено использовать метод генетического программирования, который позволяет вести случайный направленный поиск в пространстве решений [2-4].

Потенциальное решение задачи – хромосома Chr_i может быть представлена следующим выражением:

$$Chr_i = [m_1, m_2, \dots, m_l], \quad (3)$$

где m_l – гены, кодирующие операции обработки изображений с соответствующими параметрами сегментации, l – длина хромосомы, i – номер особи в популяции.

Для решения задачи сегментации изображений гистологических срезов каждый ген представляется в виде:

$$m_j = [prop_j, rp_j, wp_j, param_j], \quad (4)$$

где $prop_j$ – стандартная операция обработки изображения, rp_j – вектор входных плоскостей, wp_j – вектор выходных плоскостей, $param_j$ – вектор параметров операции обработки изображения.

Параметры могут быть целочисленными, вещественными или определяющими категорию. Входные плоскости $D_1 \dots D_{12}$ получаются представлением исходного изображения в системах координат цвета: RGB, HLS, Lab, Luv и значение контрастности. Выходные плоскости $S_1 \dots S_k$ получаются в результате обработки и могут служить входными плоскостями для других ген.

При формировании множества возможных решений используется набор стандартных операторов обработки изображений (генофонд СОРМИ), который может меняться в зависимости от области применения. В таблице 1 приведен набор стандартных спектральных, пространственных, пространственно-спектральных, логических и пороговых операторов.

Таблица 1 – Стандартные операторы обработки изображений, входящие в состав ГП

№ ГП	Обозначение гена	Операция обработки изображения	Количество		
			входов	выходов	параметров
1.	AddP	Добавляет плоскость	2	1	0
2.	AddS	Добавляет скаляр	1	1	1
3.	SubP	Вычитает плоскость	2	1	0
4.	Diff	Абсолютная разница	2	1	0
5.	NDiff	Нормализованная разница	2	1	0
6.	MultS	Умножение на скаляр	1	1	1
7.	Neg	Инвертирование плоскости	1	1	0
8.	MultP	Умножение плоскостей	2	1	0
9.	SqrtP	Корень квадратный	1	1	0
10.	SqrP	Квадрат	1	1	0
11.	LinScl	Линейное масштабирование	1	1	2
12.	LinComb	Линейное соединение	2	1	1
13.	MinP	Минимум	2	1	0
14.	MaxPs	Максимум	2	1	0
15.	IFLTE	Если меньше, то иначе	4	1	0
16.	ClipHi	Отсечение высоких частот	1	1	1
17.	ClipLo	Отсечение низких частот	1	1	1
18.	Thresh	Порог	1	1	1
19.	SAVAR	Отклонение угла спектра	2-16	1	2
20.	SADIF	Приращение угла спектра	2-16	1	2
21.	SADIST	Разница угла спектра	2-10	1	2-10
22.	SANORM	Нормализация спектрального вектора	2-10	2-10	0

Продолжение таблицы 1

№ ПП	Обозначение гена	Операция обработки изображения	Количество		
			входов	выходов	параметров
23.	Laplac3	Оператор Лапласа 3x3	1	1	0
24.	Laplac5	Оператор Лапласа 5x5	1	1	0
25.	MorphLaplac	Морфологический оператор Лапласа	1	1	1
26.	IsoGrad	Изотропический градиент	1	1	0
27.	Mean	Среднее	1	1	1
28.	Variance	Дисперсия	1	1	1
29.	Skewness	Ассиметрия	1	1	1
30.	Kurtosis	Экссесс	1	1	1
31.	SkewCoeff	Коэффициент асимметрии	1	1	1
32.	KurtCoeff	Коэффициент эксцесса	1	1	1
33.	SD	Среднее отклонение	1	1	1
34.	Range	Ранжирование	1	1	1
35.	Median	Медиана	1	1	1
36.	Erod	Эрозия	1	1	1
37.	Dil	Наращивание	1	1	1
38.	Open	Открытие	1	1	1
39.	Clos	Закрытие	1	1	1

Основные математические операции: AddS добавляет к входу скалярную величину, которая может быть отрицательной, Diff аналогичен оператору SubP, но выход имеет абсолютное значение, NDI аналогичен оператору SubP, но делит результат на сумму двух входов. MultS масштабирует вход на скалярную величину, которая по умолчанию положительная. LinScl аналогичен оператору MultS, но берет экстра-параметр, который добавляется к скалярному входу. Выход LinComb – линейная комбинация двух выходов, в специализированной пропорции по ее одному параметру, который принимает значения между 0 и 1.

Выполнение Min и Max попиксельно нахождение максимума и минимума, эквивалентно AND и OR для бинарных входов. Выход IFLTE является его третьим входом в случае, когда первый вход меньше чем второй вход, в противном случае значение принимает четвертый выход.

Операции порогового ограничения: ClipHi отсекает любой пиксель, значение которого превышает значение параметра, ClipLo наоборот. Thresh устанавливает все значения ниже параметра в 0 и все превышающие – в DataScale.

Спектральные угловые операторы SAVAR и SADIST осуществляют проверку двух круговых соседних областей вокруг каждого пикселя, размер которых определяется соответствующими параметрами. SAVAR возвращает разницу между дисперсией спектральных углов пикселей двух регионов. SADIF возвращает разницу между средним спектральным углом двух регионов. SADIST возвращает разницу спектральных углов между каждым пикселями вектором, определенными как параметры. SANORM нормализация вектора, определена с помощью входов, имеющих величину равную DataScale.

Операторы соседства обрабатывают единственную входную плоскость и получают выходную. Выход каждого пикселя определяется проверкой соседей пикселя. Остальные гены – хорошо известные операторы обработки изображений или морфологические операторы. Большинство из них имеют один параметр, который определяет радиус кругового структурного элемента для определения соседей к которым оператор применяется.

Ограничение на формирование хромосомы: каждый ген m_j в качестве входных данных может использовать только плоскости данных $D_1..D_{12}$, либо промежуточные плоскости $S_1...S_k$, определенные в одном из предшествующих генов.

Алгоритм обработки изображения, представленный отдельной хромосомой может быть изображен в виде направленного ациклического графа (рис. 1), где нетемinalные

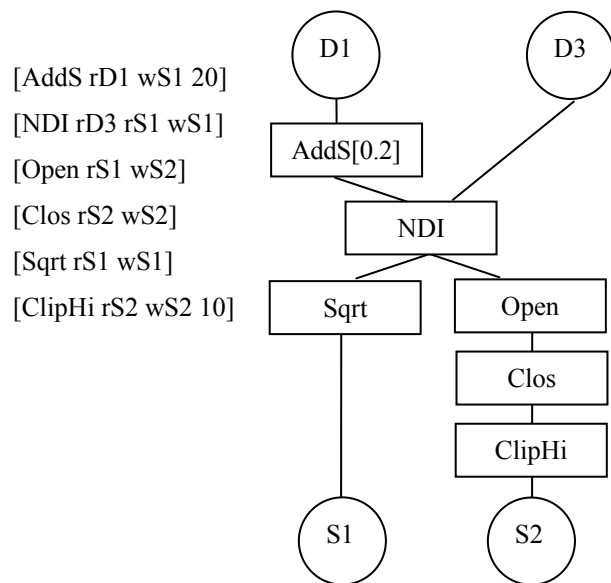


Рис. 1 Представление хромосомы

узлы являются стандартными операторами обработки изображений, а конечные узлы являются выходными плоскостями изображения. Поскольку некоторые узлы используют значения, определенные поддеревьями, то результирующий алгоритм является графом.

Качество процесса сегментации определяется с помощью оценки целевой функции (ЦФ) F_{kj}^S , которую в зависимости от специфики вида гистологического исследования предложено вычислять двумя способами:

1. Если анализ гистологического исследования не зависит от взаиморасположения объектов на изображении, то в качестве ЦФ предложено использовать минимизацию отклонения полученного количества положительных и отрицательных клеток от количества клеток, определенного по обучающей выборке (с известными результатами классификации).

$$F(Chr_i) = |N_{iP} - K_{iP}| + |N_{iN} - K_{iN}|, (F(Chr_i) \rightarrow \min), \quad (5)$$

где N_{iP} – количество положительных клеток, полученное с помощью многомерной пороговой сегментации $N_{iP} = \text{segm}(I_i, Chr_i)$, N_{iN} – количество отрицательных клеток $N_{iN} = \text{segm}(I_i, Chr_i)$, K_{iP} – количество положительных клеток, полученное из обучающей выборки, K_{iN} – количество отрицательных клеток.

2. F_{kj}^S определяется путем сравнения сегментированного изображения M_{kj}^S с изображением из обучающей выборки G_l .

$$F_{kj}^S(Chr_i) = \text{fitness}(M_{kj}^S, G_l), (F_{kj}^S(Chr_i) \rightarrow \max). \quad (6)$$

Для поставленной задачи целевая функция может быть представлена:

$$f_c = c_1 - c_2 \times \frac{\sqrt{(1 - (SP \times (1 - \alpha)))^2 + (1 - (SV \times \alpha))^2}}{\sqrt{2}}, \quad (7)$$

где c_1 и c_2 константы нормализации, α (0..1) позволяет осуществить подстройку систематической ошибки поиска в зависимости от наивысшей чувствительности SV (высокое значение α) и от наивысшей специфичности SP (низкое значение α).

$$SV = \frac{TP}{TP + FN}, \quad (8)$$

$$SP = \frac{TN}{TN + FP}, \quad (9)$$

где TP - количество правильно найденных значений позитива, FP – количество неправильно найденных значений позитива, TN - количество правильно найденных значений негатива, FN - количество неправильно найденных значений негатива.

Использование предложенной ЦФ позволяет получать адекватные эффективные алгоритмы обработки и распознавания изображений гистологических срезов.

Чтобы избежать случайного выбора точек кроссинговера и сохранить правильные решения в процессе эволюции, предлагается ставить в соответствие каждому ребру программного графа весовой коэффициент связности K_c , определяющий насколько сильно два узла должны быть связаны друг с другом.

Работа алгоритма начинается с выполнения оператора кроссинговера. Точка кроссинговера в первом поколении выбирается случайным образом. Далее для сохранения связи в точке кроссинговера для двух узлов с максимальным значением ЦФ, ребру задается меньшее значение коэффициента связности. Это осуществляется с сохранением значения коэффициента связности ребра и возрастанием значений K_c для всех остальных ребер дерева на 1, т. е. значение коэффициента связности других ребер будет соответствовать номеру поколения. На следующем шаге эволюционного процесса, указанная связь не будет участвовать в выборе точки кроссинговера и эти два связанных узла относятся к эффективной части программного кода. При продолжении эволюционного процесса, количество узлов, обладающих сильной связью возрастает, и эффективные блоки программного кода становятся больше и соединяются друг с другом. В случае, если все связи имеют равное значение коэффициента связности $K_c = \text{Gen}-1$ (где Gen – номер текущего поколения), начинается новый цикл понижения коэффициентов связности.

Конечное программное решение является эффективным программным кодом с минимальными значениями коэффициентов связности и разработанный подход полностью моделирует биологический выбор скрещивания генов, что улучшает эффективность оператора кроссинговера.

Применение адаптивного оператора кроссинговера на основе коэффициента связности приведено на рис. 2. Теоретически количество итераций n может быть достаточно велико, в случае, если точка эффективного оператора кроссинговера не может быть найдена. Например, программа-кандидат на рис. 2 уже имеет достаточно эффективный программный блок и в данной популяции не обнаружено программы, обеспечивающей более высокое значение ЦФ. В данном случае, чтобы предотвратить заикливание необходимо ограничить количество итераций. Если после выполнения определенного количества итераций эффективная программа не была обнаружена, то необходимо остановить процесс поиска и поместить копию программы в следующее поколение.

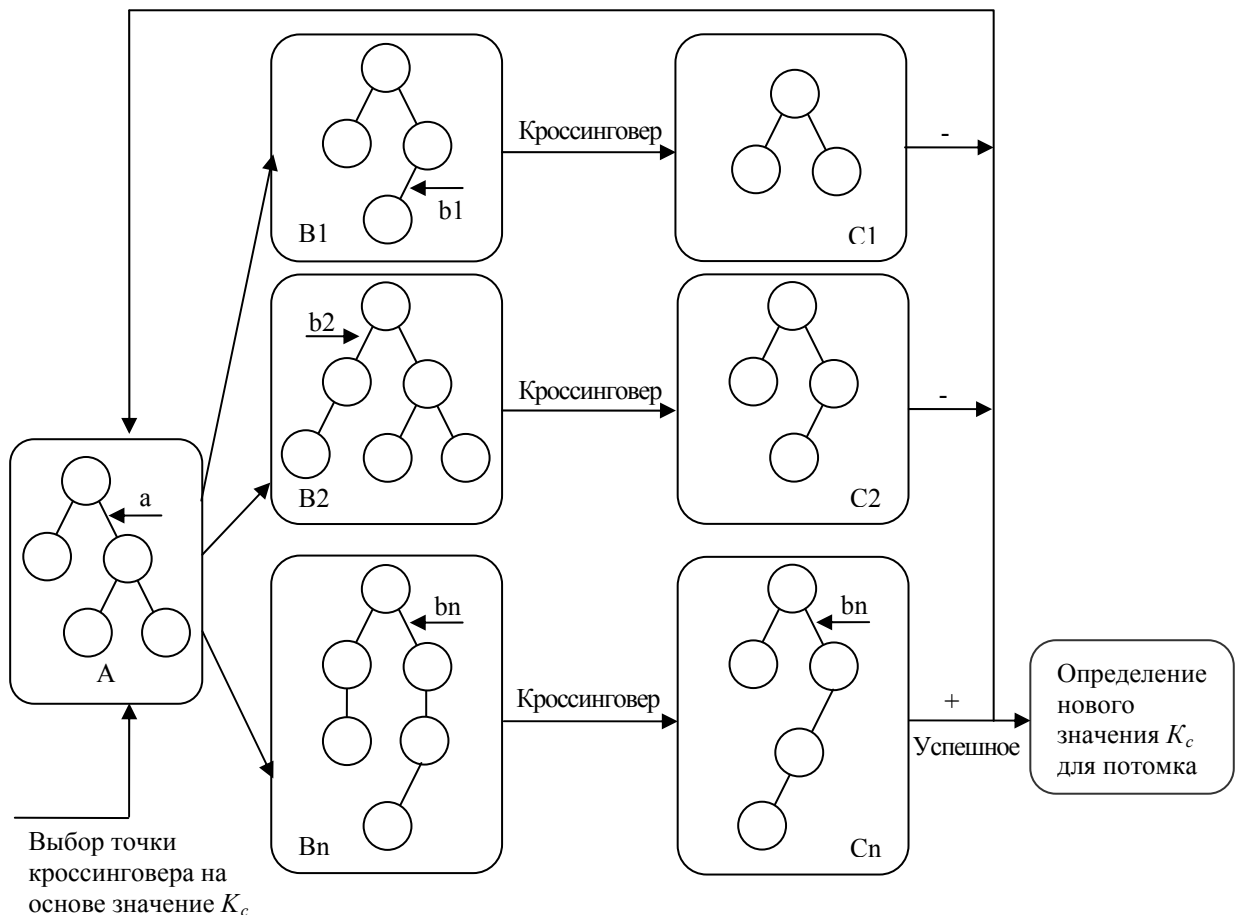


Рис. 2 Применение адаптивного оператора кроссинговера на основе значения коэффициента связности

В результате экспериментальных исследований получены эффективные алгоритмы для обработки и распознавания изображений гистологических срезов. Применение полученных алгоритмов для сегментации изображения гистологического среза стромы, подвергнутого иммуногистохимическому окрашиванию, представлено на рис. 3.

Выводы.

Предложен новый подход к построению эффективного алгоритма сегментации цветных изображений гистологических срезов на основе эволюционных моделей. В качестве целевой функции для отбора лучших решений предложено использовать оценку систематической ошибки поиска в зависимости от чувствительности и от специфичности. Разработаны функциональное и терминальное множества, проблемно-ориентированные операторы кроссинговера и мутации. Апробация предложенного подхода осуществляется на базе лаборатории фундаментальных исследований ИНВХ АМН Украины при проведении гистологических исследований.

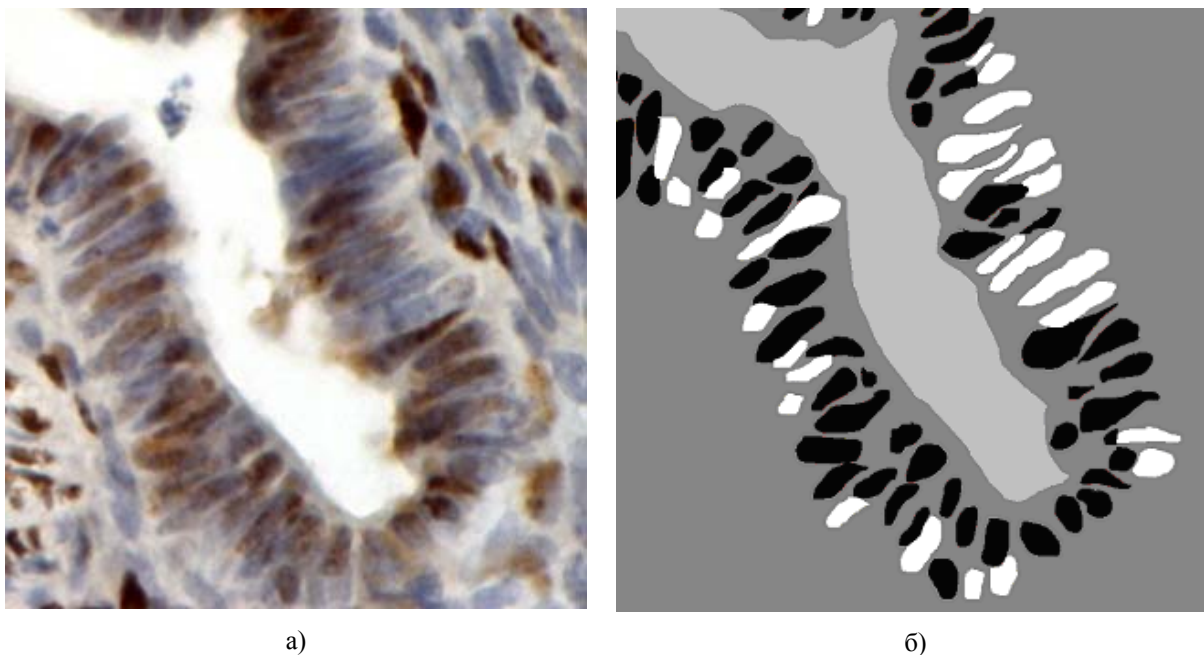


Рис. 3 Результаты эволюционных вычислений
 а) исходное изображение
 б) обработанное изображение, где

■ - Класс 1 (железа)

$f_u=0,89$

[ClipHi rS1 wS1 189] [Mean rS1 wS1 45000]

■ - Класс 2 (положительные клетки)

$f_u=0,83$

[ClipLo rD1 wS2 125] [ClipHi rS2 wS2 200] [Open rS2 wS2] [Clos rS2 wS2] [Mean rS2 wS2 2108]

□ - Класс 3 (отрицательные клетки)

$f_u=0,81$

[AddS rD8 rS3 32] [ClipLo rD8 wS3 105] [ClipHi rS3 wS3 131] [Open rS3 wS3] [Clos rS3 wS3] [Mean rS3 wS3 2156]

Литература.

1. Аруин А.И., Кануллер Л.Л., Исаков В.А. Морфологическая диагностика болезней желудка и кишечника – М.: Триада-Х, 1998.
2. Goldberg D.E. Genetic algorithms in search, optimization and machine learning /Addison-Wesley Publishing Company, Inc. – 1989.
3. Koza J. R. Genetic Programming: On the Programming of Computers by Means of Natural Selection (Complex Adaptive Systems). Cambridge, MA: MIT Press, 1992.
4. Brumby S., Theiler J., Perkins S., Harvey N., Mitchell M. Investigation of Image Feature Extraction by a Genetic Algorithm, 1999 [<http://citeseer.ist.psu.edu/340330.html>]

Ю.А. Скобцов, Т.В. Мартыненко

Построение эффективных алгоритмов обработки и распознавания изображений гистологических срезов на основе эволюционных моделей

В статье предложен новый подход к построению эффективных алгоритмов сегментации цветных изображений гистологических срезов на основе эволюционных моделей. Алгоритм обработки изображения в виде отдельной хромосомы представлен направленным ациклическим графом, где нетерминальные узлы являются стандартными операторами обработки изображений, а конечные узлы являются выходными плоскостями изображения. Качество процесса сегментации определяется с помощью оценки целевой функции. Разработаны функциональное и терминальное множества, проблемно-ориентированные операторы кроссинговера и мутации. Указанный подход подтвердил свою эффективность в процессе экспериментальных исследований на базе лаборатории фундаментальных исследований ИНВХ АН Украины при проведении гистологических исследований.

Ю.О. Скобцов, Т.В. Мартыненко

Побудова ефективних алгоритмів обробки та розпізнавання зображень гістологічних зрізів на основі еволюційних моделей

У статті запропонований новий підхід до побудови ефективних алгоритмів сегментації кольорових зображень гістологічних зрізів на основі еволюційних моделей. Алгоритм обробки зображення у вигляді окремої хромосоми представлений спрямованим ациклічним графом, де нетермінальні вузли являються стандартними операторами обробки зображень, а кінцеві вузли - вихідними площинами зображення. Якість процесу сегментації визначається за допомогою оцінки цільової функції. Розроблено функціональну та термінальну множини, проблемно-орієнтовані оператори кроссинговера і мутації. Зазначений підхід підтвердив свою ефективність у процесі експериментів на базі лабораторії фундаментальних досліджень ИНВХ АН України при проведенні гістологічних досліджень.

Y.A. Skobtsov, T.V. Martynenko

Construction Of Effective Images Processing And Recognition Algorithms Of Histologic Cuts Based On Evolutionary Models

In article the new approach to construction of effective color images segmentation algorithms of histologic cuts is offered on the basis of evolutionary models. The algorithm of image processing and recognition is submitted by directed acyclic graph where nonthermal junctions are standard images processing operators and thermal junctions are target planes of the image. Quality of segmentation process is defined with the estimation of criterion function. The functional and terminal sets, the problem-oriented crossover and mutation operators are developed. The specified approach has confirmed the efficiency during experimental researches on the basis of laboratory of basic researches IURS AS of Ukraine at during histologic researches.