

А.В. Григорьев, Т.В. Мартыненко

Донецкий национальный университет, г. Донецк
Донецкий национальный технический университет, г. Донецк

УСОВЕРШЕНСТВОВАНИЕ МЕТОДА ЭВОЛЮЦИОННОГО СИНТЕЗА ДЛЯ ПОВЫШЕНИЯ СКОРОСТИ ОБРАБОТКИ ИЗОБРАЖЕНИЙ ГИСТОЛОГИЧЕСКИХ СРЕЗОВ

Публікацію присвячено розвитку методу еволюційного синтезу алгоритмів обробки зображень гістологічних зрізів. Було здійснено модернізацію розробленої фітнес-функції, що дозволяє враховувати швидкість кожної операції обробки зображення; та покращення локальних рішень, що ґрунтується на зміні деяких фрагментів синтезованих алгоритмів їх більш ефективними аналогами. Результати проведених комп'ютерних експериментів показали, що для трьох розглянутих видів гістологічних досліджень перехід до оптимізованого еволюційного методу дозволяє на порядок підвищити швидкість обробки зображень гістологічних зрізів без втрати точності.

The paper is devoted to the evolutionary synthesis method enhancement. The modernization of the fitness-function is carried out. It allowed raise a speed of synthetic image processing algorithms by accounting each operator's speed. The local decision optimization allowed raise synthetic algorithm speed by changing groups of operators with its more rapid analogues. The results of the computer experiments appear the optimized evaluation method implementation, which allows raise speed of synthetic algorithms degree without it accuracy lack.

Введение

Большинство компьютерных систем медицинской и технической диагностики (КСД), базирующихся на анализе цифровых изображений, обладают низкой точностью устанавливаемого диагноза. Это обусловлено необходимостью обработки больших объемов графической информации, которой присущи низкое качество и высокая вариабельность большинства структур изображений. Основными компонентами КСД являются подсистемы обработки и распознавания изображений. Подсистема обработки осуществляет улучшение качества изображения, т.е. отделение исходного сигнала от внешних воздействий. Подсистема распознавания изображений позволяет произвести анализ обработанного сигнала, а так же представление полезной составляющей сигнала в виде, удобном для специалиста.

В связи с развитием КСД появилась возможность повысить скорость обработки медицинских данных и получить объективную оценку состояния больных онкологическими заболеваниями. При диагностике и прогнозировании развития новообразований обязательным является исследование клеточных структур на основе

анализа изображений гистологических срезов. Эффективность данного процесса может быть повышена путем автоматизации анализа морфологии изображений гистологических срезов.

Современные автоматизированные системы обработки и анализа изображений гистологических срезов осуществляют анализ снимка на основе структурных элементов клетки, оценивают их размер, форму и в некоторых ситуациях контекст.

Анализ существующих КСД, основанных на обработке изображений, показывал, что для данных систем характерны следующие недостатки:

- отсутствие универсальных методов обработки, сегментации и распознавания изображений [1-6];
- отсутствие критериев эффективности обработки и распознавания медицинских изображений;
- отсутствие эффективных методов и процедур классификации [1, 3].

Таким образом, рассмотренные КСД не позволяют осуществлять автоматическую настройку (как на условия получения исходного изображения, так и на особенности медицинско-

го исследования) в полной мере [1, 2, 4-10].

В связи с этим актуальна разработка специализированных инструментальных средств, позволяющих ускорить процесс обработки изображений-препаратов [1, 7, 9], повысить точность определения изменений внутренних органов, и, как следствие, повысить точность установки диагноза.

В [10, 11] описана реализация инструментальных средств, построенная на базе метода эволюционного синтеза, в основе которого лежит модификация эволюционного подхода к формированию алгоритмов обработки изображений.

Постановка задачи

Задан набор цифровых цветных изображений гистологических срезов A_i , $i \in \{1, \dots, N\}$. Для каждого A_i на базе экспертной оценки специалиста в области гистологии сформированы наборы бинарных изображений $B^{(i)} = \langle B_1^{(i)}, \dots, B_K^{(i)} \rangle$. Функция яркости каждого $B_j^{(i)}$ совпадает с характеристической функцией множества пикселей, относящихся, по мнению эксперта, к j -му классу элементов гистологического изображения A_i . Количество классов K зависит от вида гистологического исследования.

Необходимо синтезировать алгоритм F , преобразующий исходные цветные изображения A_i в серии бинарных изображений $O^{(i)} = \langle O_1^{(i)}, \dots, O_K^{(i)} \rangle = F(A_i)$. При этом, различие между изображениями $O^{(i)}$, сформированными алгоритмом F , и $B^{(i)}$, предоставленными экспертом, должно быть минимальным, т.е. необходимо найти:

$$F = \arg \min_F \sum_{i=1}^N \Delta(F(A_i), B^{(i)}), \quad (1)$$

$$\Delta(F(A_i), B^{(i)}) = \Delta(O^{(i)}, B^{(i)}) = \sum_{j=1}^K \delta(O_j^{(i)}, B_j^{(i)}), \quad (2)$$

$$\delta(O_j^{(i)}, B_j^{(i)}) = SP(O_j^{(i)}, B_j^{(i)}) \cdot SV(O_j^{(i)}, B_j^{(i)}). \quad (3)$$

Здесь $SP(A, B)$ и $SV(A, B)$ – соответственно, специфичность и чувствительность изображения A относительно шаблона B , вычисляемые по формулам:

$$SV(A, B) = \frac{\#(A \cap B)}{\#A} \quad (4)$$

$$SP(A, B) = \frac{\#(\bar{A} \cap \bar{B})}{\#A}, \quad (5)$$

где $\#A$ – мощность множества A .

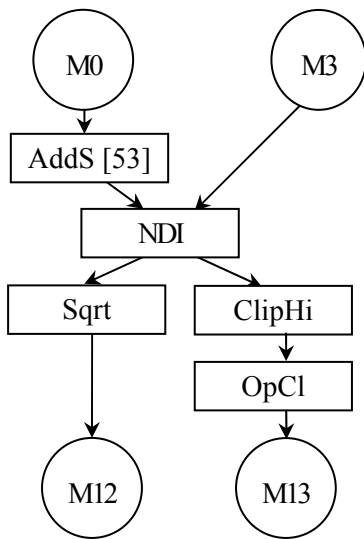
Метод эволюционного синтеза алгоритмов обработки изображений

Для решения поставленной задачи в [10, 11] предложен метод эволюционного синтеза. Данный метод представляет собой модификацию парадигмы генетического программирования с использованием в качестве составляющих генома элементарных операций – основных операций цифровой обработки изображений. Гены хромосом $Chr_i = [m_1, m_2, \dots, m_i]$ в данном методе представлены в виде:

$$m_j = [prop_j, rp_j, wp_j, param_j] \quad (6)$$

где $prop_j$ – стандартная операция обработки изображения; rp_j – вектор входных плоскостей; wp_j – вектор выходных плоскостей; $param_j$ – вектор параметров. Алгоритм обработки изображения, представленный отдельной хромосомой, может быть изображен в виде направленного ациклического графа (рис. 1). На данном графе начальные узлы представляют цветовые плоскости (M_0 - M_{11}) исходного изображения в различных цветовых схемах. Здесь использованы плоскости схем: RGB (M_1 - M_3), HLS (M_4 - M_6), Lab (M_7 - M_9), Luv (M_{10} , M_{11}), а также усреднённая яркость C по каналам RGB (M_0). Промежуточные узлы графа – операции обработки изображения. Функциональное множество, представлено следующими операциями обработки изображений: попиксельное сложение, вычитание, произведение, а так же нечеткое объединение и пересечение плоскостей; градиционные преобразования (негатив, константное, линейное и нелинейное затемнение и осветление изображения, пороговая бинаризация, выделение областей по уровню яркости); линейная фильтрация (фильтры Гаусса, Лапласа, Собеля, Превита); морфологические операции (эрозия, дилатация, размыкание, замыкание, морфологическая фильтрация); операции вычисления характеристик выделенных элементов (площади, периметра, статистических моментов первого и высших порядков, эксцентриситета). Конечные узлы графа – итоговые плоскости $B_j^{(i)}$ (M_{12} – M_{11+K}).

Таким образом, в рамках предложенного представления возможно конструирование достаточно сложных алгоритмов обработки изображений, позволяющих решать поставленную задачу.



[AddS rM0 wM12 53][NDI rM3 rM12 wM12]
 [Sqrt rM12 wM12][ClipHi rM12 wM13 70]
 [OpCl rM13 wM13]

Рис. 1 Сетевое представление хромосомы

В качестве фитнес-функции для разработанного метода использовано выражение (3) с добавленным мультипликативным штрафом $\beta = (md - 0,5 \cdot d - 0,5) / (md - 1)$, направленным на глубины программного графа. Здесь d – количество узлов текущей особи, md – параметр, устанавливающий максимальную глубину графа.

В качестве операторов репродукции использованы: кроссинговер, представляющий собой обмен подграфов родительских особей, выполняемым по правилам, описанным в [11]; мутация, состоящая в замене случайно выбранной операции или значений одного или нескольких ее параметров.

Предложенный метод показал свою эффективность [10, 11], однако ввиду большого количества обрабатываемой графической информации остается актуальной задача повышения скоростных характеристик синтезируемых алгоритмов обработки.

Повышение скоростных характеристик синтезированных алгоритмов

Проведенный нами анализ описанного метода показал, что повышение скоростных характеристик синтезированных алгоритмов возможно за счет следующих усовершенствований:

- оптимизации локальных решений посредством введения более быстрых способов вычисления определенных групп операций;

- модернизации фитнес-функции (3) за счет учета скоростных характеристик операций обработки из функционального множества.

В основу первого усовершенствования положено два принципа. Первый базируется на том, что ряд операций может быть заменен одной или несколькими альтернативными, имеющими более высокие скоростные характеристики. В частности, несколько последовательно выполняемых над одной плоскостью градиционных преобразований могут быть заменены одной, являющейся их композицией. С другой стороны, при вычислении операции свертки замена одной операции на несколько более простых операций может повысить скорость вычисления. Например, применение двумерного гауссова фильтра может быть сведено к последовательному вычислению одномерных (по строкам и по столбцам изображения).

Второй принцип состоит в использовании промежуточных или конечных результатов, полученных при вычислении одних операций, для других. Так, результат вычисления эрозии по одной из плоскостей может быть использован при вычислении операции размывания или морфологической фильтрации. Значение первого статистического момента может быть использовано при вычислении статистических моментов более высокого порядка.

Модернизация фитнес-функции метода эволюционного синтеза [10,11] состоит в изменении мультипликативной штрафной составляющей β таким образом, чтобы последняя учитывала не только количество операций синтезированного алгоритма, но и скорость их выполнения.

С учётом выше сказанного, была разработана фитнес-функция:

$$f_{mod}(F) = 2^{-\left(\sum_{i=1}^d \tau_i\right)/\sigma} \times \min_{q \in \{1, \dots, N\}} \sum_{j=1}^K SP(F(A_q), B_j^{(q)}) \cdot SV(F(A_q), B_j^{(q)}), \quad (7)$$

где N – мощность обучающего набора; K – количество классов, выделяемых экспертом; τ_i – расчетная скорость выполнения i -й операции алгоритма F ; σ – параметр фитнес-функции.

Значение параметра σ выбирается из следующих соображений. При уменьшении σ фитнес-функция (7) в большей степени оптимизирует скорость синтезированного алгоритма, может быть, в ущерб его точности. С другой

стороны, при увеличении σ большее значение имеет точность алгоритма – в экстремальном случае, при $\sigma \rightarrow +\infty$ фитнес-функция (7) вырождается в целевую функцию (3). Таким образом, параметр σ выбирается экспертом (врачом-морфологом) с учётом рационального баланса между требованиями к точности и скорости, предъявляемыми к синтезируемому алгоритму обработки.

Коэффициенты τ_i выражают среднее количество арифметических операций (умножение, сложение, вычитание, сравнение), производимых в расчёте на один пиксель изображения. При расчёте τ_i учитываются: расчётная сложность операции и наличие либо отсутствие условий, облегчающих её выполнение. Например, при выполнении морфологической операции замыкания облегчающим условием является наличие результата дилатации; при расчёте статистических моментов высоких порядков используются результаты вычисления статистического момента первого порядка.

Таким образом, разработанная фитнес-функция (7) учитывает как точность обработки изображений гистологических срезов на основе значений чувствительности SV и специфичности SP , так и скорость синтезированного алгоритма за счёт введенной нами штрафной составляющей. Последняя стимулирует использование более быстрых операций обработки (оценка скорости каждой операции заложена в коэффициентах τ_i) и уменьшение общего количества операций синтезированного алгоритма.

Анализ полученных результатов

Для оценки предложенной нами модификации метода эволюционного синтеза был проведен ряд экспериментов, целью которых является сравнение результатов работы лучших алгоритмов, полученных на базе методов эволюционного синтеза, и предложенной модификации.

Известно, что точность и скорость работы алгоритмов, полученных модифицированным методом, зависит от значения параметра σ , и скорость работы алгоритмов, синтезированных обоими методами, зависит от выбора размера обрабатываемых изображений. Таким образом, для достижения поставленной цели необходимо решить следующие задачи:

- 1) Определить значение параметра σ , позволяющее получить наиболее наглядное сравнение синтезированных и модифицированных алгоритмов;
- 2) Определить размер изображений, поло-

женных в основу сравнения двух методов.

Решения поставленных задач предложено искать, исходя из следующих соображений. Поскольку метод эволюционных вычислений направлен, в первую очередь, на оптимизацию точности синтезируемого алгоритма (без возможности подстройки под его скоростные характеристики), значение σ_0 параметра σ для оптимизированного алгоритма выбирается таким образом, чтобы после $\sigma = \sigma_0$, прирост точности вычислений был не значительным в сравнении с приростом величины σ .

При проведении указанных экспериментов были взяты следующие рациональные значения параметров эволюционного алгоритма: мощность популяции $S_{POP}=50$, количество поколений $C_G=8$, вероятность мутации $P_M=0,09$ и адаптивного кроссингвера $P_{AC}=0,5$. Эксперименты проведены на данных, полученных сотрудниками паталого-анатомического отделения Донецкого областного клинического территориального медицинского объединения.

Из графика (рис.2) видно, что насыщение фитнес-функции f_{mod} достигается после $\sigma_0 = \sigma = 1000000$. Зависимость времени работы оптимизированного эволюционного метода от значения параметра σ приведена на рис. 3. Из рис. 2, 3 видно, что внимания заслуживают также значения параметра $\sigma \approx 10$ и $\sigma \approx 1000$, являющиеся точками перелома как первой, так и второй зависимости. Отметим, что наш выбор значения параметра $\sigma = 1000000$ для сравнения двух методов обоснован, поскольку данная точка даёт минимальный проигрыш в точности полученных результатов обработки (зачастую, основополагающая характеристика алгоритма).

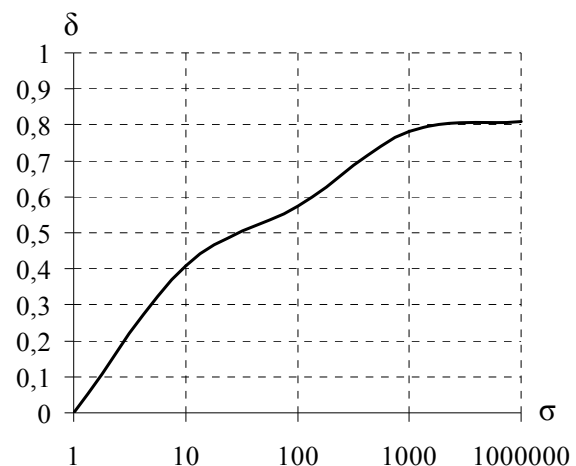


Рис. 2 Зависимость значений целевой функции от коэффициента σ .

Экспериментально установлено, что при $\sigma = 1000000$ алгоритмы, синтезированные оптимизированным эволюционным методом, по точности, в основном, дают решение, идентичное методу эволюционного синтеза. Таким образом, выбор параметра $\sigma = 1000000$ при сравнении скоростных характеристик алгоритмов, получаемых двумя описанными методами, обоснован, поскольку точность этих алгоритмов, фактически, идентична.

Зависимость времени выполнения алгоритмов, синтезированных оптимизированным эволюционным методом, от параметра σ приведена на рис. 3. Результаты приведенных здесь исследований также подтверждают «интересность» точек $\sigma = 10$ и $\sigma = 1000$. Очевидно, данные точки могут дать значительный выигрыш в скорости выполнения синтезированных алгоритмов в сравнении с падением точности последних.

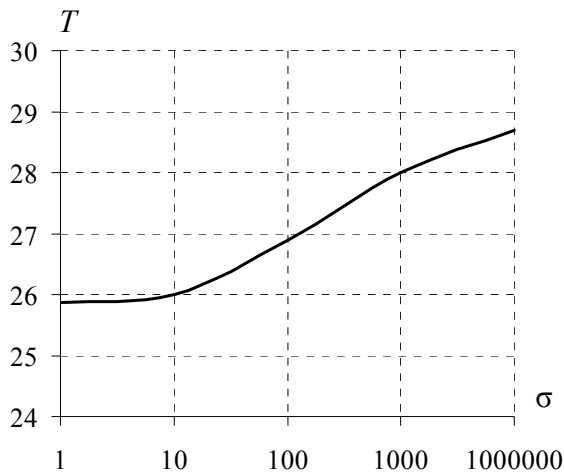


Рис. 3 Зависимость времени работы синтезированного алгоритма от коэффициента σ .

Таким образом, при сопоставлении двух методов целесообразно использовать значение параметра $\sigma = 1000000$.

При решении второй подзадачи более показательным является не значение целевой функции для некоторого выбранного разрешения исходных изображений, а расчетная эффективность алгоритмов, как динамика зависимости скорости обработки от объема исходных данных. Поскольку рассматриваемые зависимости определяются видом гистологического исследования, по данной части было проведено три компьютерных эксперимента, каждый из которых соответствует определенному виду: гистологические срезы железы и стромы печени, а также лимфатического узла. Целью каждого

эксперимента являлось сравнение скоростных характеристик метода эволюционного синтеза и оптимизированного алгоритма для представленных видов гистологических исследований. Из графиков рис. 4-6 видно, что применение оптимизированного эволюционного метода позволяет на порядок увеличить скорость обработки изображений гистологических срезов.

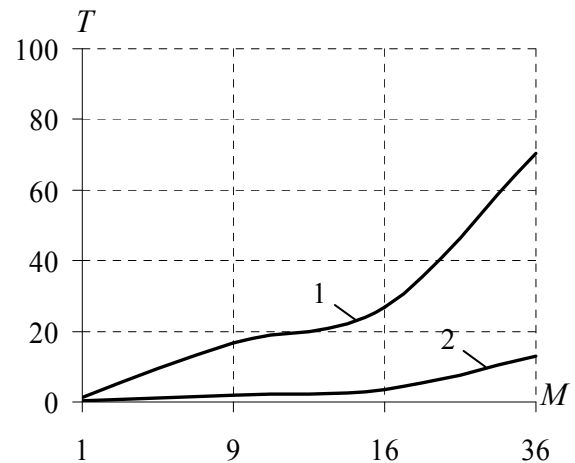


Рис. 4 Зависимость времени работы синтезированного алгоритма от размера изображения по гистологическим срезам железы печени:

1 – обработка изображения синтезированным алгоритмом;
2 – оптимизированный алгоритм обработки изображения.

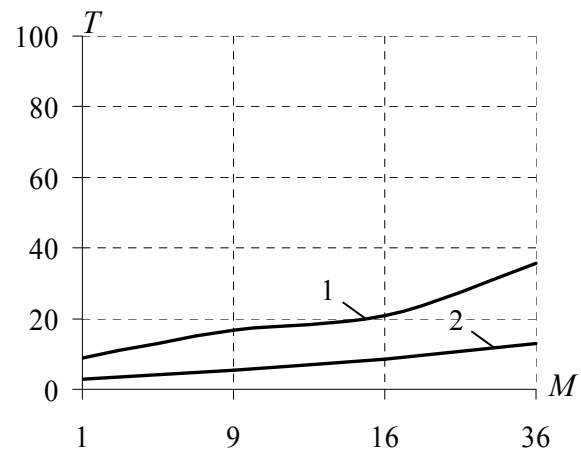


Рис. 5 Зависимость времени работы синтезированного алгоритма от размера изображения по гистологическим срезам стромы печени:

1 – обработка изображения синтезированным алгоритмом;
2 – оптимизированный алгоритм обработки изображения.

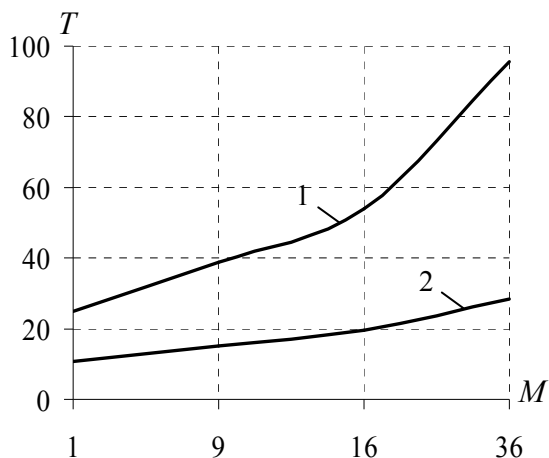


Рис. 6 Зависимость времени работы синтезированного алгоритма от размера изображения по гистологическим срезам лимфатического узла:

1 – обработка изображения синтезированным алгоритмом;
2 – оптимизированный алгоритм обработки изображения.

Выводы

В статье рассмотрены вопросы усовершенствования метода эволюционного синтеза. Предложены: 1) модернизация фитнес-функции метода эволюционного синтеза, позволяющая учитывать не только количество операций синтезированного алгоритма, но и скорость их выполнения; 2) оптимизация локальных решений, позволяющая заменить фрагменты алгоритма-решения их более эффективными аналогами.

Результаты проведенных компьютерных экспериментов показали, что для трёх рассмотренных видов гистологического исследования переход к оптимизированному эволюционному методу позволяют на порядок повысить скорость обработки изображений гистологических срезов без потери точности.

СПИСОК ЛИТЕРАТУРЫ

1. Gunduz C., Yener B., Gultekin S.H. The cell graphs of cancer. // Bioinformatics – 2004. – No. 20. – P. 145-151.
2. Poli R. Hopfield neural nets for the optimum segmentation of medical images. Handbook of Neural Computation. – Oxford University Press. – 1997.
3. Serra J. Image Analysis And Mathematical Morphology. – Academic Press, Inc. – 1982. P. 621.
4. Tasoulis D.K., Spyridonos P., Pavlidis N.G., Cavouras D., Ravazoula P., Nikiforidis G. Urinary bladder tumor grade diagnosis using on-line trained neural networks. // Based Intell. Inform. Eng. Syst. –

2003. P. 199-206.
5. Walker R.F., Jackway P.T., Lovell B. Classification of cervical cell nuclei using morphological segmentation and textural feature extraction. // Proc of the 2nd Australian and New Zealand Conference on Intelligent Information Systems. – 1994. – P. 297-301.
6. Weyn J.B., G. Van de Wouwer, Koprowski M., A. Van Daele, Dhaene K., Scheunders P., Jacob W., E. Van Marck Value of morphometry, texture analysis, densitometry, and histometry in the differential diagnosis and prognosis of malignant mesothelioma. // Pathol. – 1999 No 189. – P. 581-589.
7. Clinical Laboratory Improvement Advisory Committee. Summary Report. U.S. Department Of Health & Human Services Public Health Service. – 1996.
8. Cui G., Koh T., Chen D. Overexpression of Glycine-Extended Gastrin Inhibits Parietal Cell Loss and Atrophy in the Mouse Stomach Cancer Res. – 2004. P. 8160-8166.
9. Demir C., Gultekin S.H., Yener B. Learning the topological properties of brain tumors. // Rensselaer Polytechnic Institute Technical Report TR-04-14. – Troy. – 2004.
10. Скобцов Ю.А., Мартыненко Т.В. Эволюционные методы обработки цветных изображений на примере гистологических срезов // Наукові праці Донецького національного технічного університету. Серія: «Обчислювальна техніка та автоматизація». Випуск 14(129). – Донецьк: 2008. – С.147-155.
11. Скобцов Ю.А., Мартыненко Т.В. Построение эффективных алгоритмов обработки и распознавания изображений гистологических срезов на основе эволюционных моделей // Вестник Херсонского национального технического университета, №4(27). – 2007. – С. 43-48.
12. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – Москва: Техносфера, 2005. – 1072 с.