

Artur BAL

POLITECHNIKA ŚLĄSKA, INSTYTUT AUTOMATYKI,
ul. Akademicka 16, 44-100 Gliwice

Grafowa metoda segmentacji obiektów w postaci skupisk na przykładzie obrazów kometowych

Dr inż. Artur BAL



Absolwent Politechniki Śląskiej, gdzie pracuje na stanowisku adiunkta w Instytucie Automatyki. Jego zainteresowania badawcze dotyczą m.in. przetwarzania i analizy obrazów, a w szczególności: analizy obrazów biomedycznych, obszarowej segmentacji obrazów, wizualnej archiwizacji obiektów, odwzorowania barw. Jest kierownikiem Laboratorium Wizualnej Archiwizacji Cyfrowej utworzonego w Instytucie Automatyki Politechniki Śląskiej.

e-mail: artur.bal@polsl.pl

Streszczenie

W pracy przedstawiono propozycję metody segmentacji obiektów będących skupiskami — przykładem takich obiektów są tzw. komety, które są wynikiem jednokomórkowej elektroforezy żelowej. Opracowana metoda działa dwuetapowo: etap 1. to segmentacja służąca wyznaczeniu elementów składowych należących do obiektów, etap 2 wykorzystuje minimalne drzewo rozpinające do określenia zbioru elementów tworzących poszczególne obiekty, obszar poszczególnych obiektów wyznaczany jest jako otoczka wypukła odpowiedniego drzewa rozpinającego.

Słowa kluczowe: analiza obrazów, segmentacja obrazów, obiekty w postaci skupisk, minimalne drzewo rozpinające, test kometowy, jednokomórkowa elektroforeza żelowa.

Graph-based aggregate object image segmentation method on example of comet assay images

Abstract

This paper deals with a problem of segmentation of aggregate objects, that is objects which are formed by a set of unconnected elements smaller than the object itself. Images of such a type of objects are very difficult for segmentation. An example of this type of objects are “comets” (Fig. 1, left column) from *Single Cell Gel Electrophoresis* images (also called *comet assay* images). In comet assay images the comet region is formed by unconnected fragments of DNA. Because of not satisfying results of comet segmentation with use of the standard methods, a new method for segmentation of such images was developed. The new method works in two stages. The first stage is the image segmentation—for comets the Bernsen binarization method (Eqs. (1) and (2)) with median filtering of the obtained results was chosen—the result of this stage is a set of comet elements e_i which represent DNA fragments (Fig. 1, the 2nd column). In the second stage the minimum spanning trees T_p are created (Fig. 1, 3rd column)—graph vertexes v_i represent elements e_i , and length d_{ij} of edge e_{ij} between vertexes v_i and v_j is equal to the closest distance between pixels of elements e_i and e_j —then for each connected tree T_p its convex hull which defines the region of comet K_p (Fig. 1, the 4th column) is created. In case of defects appearing in comet images, the incorrect region can be rejected e.g. by use of geometrical or photometrical features of the regions.

Keywords: image analysis, image segmentation, aggregate objects, minimum spanning tree, comet assay, single cell gel electrophoresis.

1. Wprowadzenie

Coraz częściej obraz lub sekwencja obrazów jest ważnym, a czasem jedynym źródłem informacji o badanym obiekcie lub procesie. Pozyskanie tych informacji z obrazów wymaga wykorzystania odpowiednich metod z zakresu przetwarzania i analizy obrazów. Jednym z najważniejszych i zarazem najtrudniejszych procesów jakie należy przeprowadzić w celu wyekstrahowania z obrazów takich informacji jest proces segmentacji obrazów.

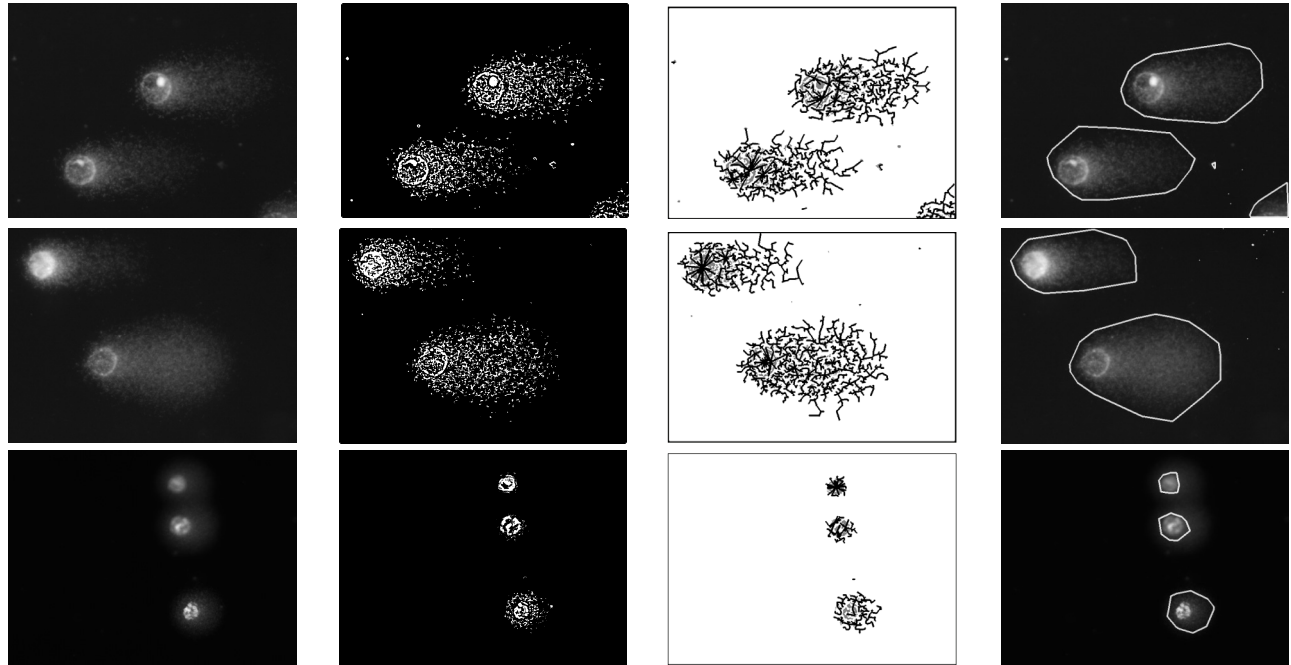
W ogólnym przypadku konieczność realizacji tego procesu zachodzi każdorazowo, gdy celem przetwarzania jest uzyskanie informacji o cechach konkretnego obiektu przedstawionego na analizowanym obrazie. Jakość wyniku segmentacji decyduje w znacznym stopniu o poprawności całego procesu analizy obrazów. Najlepsze wyniki uzyskiwane są, gdy segmentacji podlegają obiekty spójne, dla których reprezentujące je piksele charakteryzują się dużą jednorodnością cech np. jasności lub barwy i dodatkowo istnieje dobrze rozpoznawalna różnica między pikselami reprezentującymi obiekt a innymi pikselami. Niestety w przypadku rzeczywistych danych takie warunki praktycznie nie występują. Poważnym problemem jest segmentacja obiektów niespójnych. Jednym z rodzajów takich obiektów są obiekty w postaci skupisk wielu mniejszych elementów o określonych cechach — przykładem takich obiektów są tzw. komety otrzymywane w wyniku jednokomórkowej elektroforezy żelowej (rys. 1, lewa kolumna).

2. Problem segmentacji obrazów z jednokomórkowej elektroforezy żelowej

Jednokomórkowa elektroforeza żelowa (ang. *single cell gel electrophoresis*, SCGE) określana w skrócie jako test kometowy (ang. *comet assay*) [4, 5, 7] jest metodą oceny stopnia uszkodzenia pojedynczych komórek poddanych działaniu czynników stresujących komórki takich jak np. promieniowanie jonizujące i stosowany jest m.in. do indywidualizowanego planowania leczenia chorych z chorobą nowotworową. Ocena stopnia uszkodzenia komórek polega na określeniu ilości DNA, które pod wpływem pola elektrycznego, znalazło się poza jądrem komórkowym, z którego pochodzi. Ocena całego testu kometowego wyznaczana jest na podstawie wyników oceny stopnia uszkodzenia wszystkich komórek (zwykle wymaga się by ocenie poddano co najmniej 100 komórek) w ramach danej próbki (w przypadku planowania leczenia w onkologii pojedyncza próbka są to komórki pobrane od jednego pacjenta w jednym czasie).

Przeprowadzenie testu kometowego wymaga odpowiedniego przygotowania materiału biologicznego tzn. m.in. odpowiedniego wybarwienia komórek barwnikami fluorescencyjnymi i przeniesienia ich na odpowiednie podłoże; ocena stanu komórek dokonywana jest poprzez obserwację mikroskopową przy pomocy mikroskopu fluorescencyjnego. Ocena ta dokonywana jest zazwyczaj wizualnie przez odpowiednio przeszkolonego laboranta co skutkuje stosunkowo niską dokładnością, powtarzalnością i wiarygodnością otrzymywanych wyników. Dodatkowo taki sposób oceny skutkuje występowaniem dużego zróżnicowania wyników między poszczególnymi laborantami i w praktyce uniemożliwia wiarygodne porównywanie wyników testów kometowych otrzymanych w różnym czasie lub różnych ośrodkach.

W celu eliminacji tych niekorzystnych zjawisk coraz częściej do oceny wyników testu kometowego stosowane są metody komputerowego przetwarzania i analizy obrazów. Zwykle jednak nie są to metody automatyczne, a jedynie rozwiązania wspomagające laboranta w niektórych czynnościach i wymagające jego nieustannego zaangażowania. Podstawowym problemem występującym przy ocenie testu kometowego jest określenie obszaru komety czyli obszaru zajmowanego przez DNA pochodzące z tego samego jądra komórkowego. Wyznaczenie takiego obszaru ze względu na występujący brak bezpośredniej styczności poszczególnych fragmentów DNA (a zatem również spójności obszaru zajmowanego przez komety) nie jest zadaniem trywialnym.



Rys. 1. Wyniki działania kolejnych etapów metody grafowej; w kolumnach od lewej: oryginalne obrazy kometowe, wyniki 1. etapu działania metody tzn. wyszukiwania elementów e_i (reprezentujących w tym przypadku fragmenty DNA), minimalne drzewa rozpinające T_p rozpięte na elementach e_i , oryginalne obrazy z zaznaczonymi obszarami komet wyznaczonymi przez otoczki wypukłe rozciągnięte na poszczególnych drzewach rozpinających T_p

Fig. 1. Results of different stages of the graph based image segmentation method; in columns from left side: original comet assay images, results after 1. stage it is segmentation of elements e_i (which represents in this case DNA fragments), minimum spanning trees T_p spanned on e_i , comets regions obtained by creation of convex hull for each tree T_p superimposed on original comet assay images

Prawidłową segmentację komet utrudniają ponadto:

- zazwyczaj występująca niewielka różnica jasności między brzegowymi punktami należącymi do komety a jej tłem,
- zbyt bliskie w stosunku do siebie położenie komet w preparacie, a także nakładanie się na siebie fragmentów komet,
- artefakty będące wynikiem niewłaściwego przygotowania preparatu oraz błędy popełnione podczas akwizycji obrazów.

Wszystkie te czynniki powodują, że segmentacja obrazów kometowych jest zadaniem, które jeszcze nie doczekało się skutecznego rozwiązania.

3. Opis grafowej metody segmentacji

Podstawowym założeniem prezentowanej metody jest rezygnacja z przyjmowanego często niejawnie założenia, że obiekty są to spójne, prawie jednorodny, zbiory pikseli leżące na prawie jednorodnym tle. Zamiast tego przyjęto, że każdy obiekt utworzony jest przez pewien zbiór elementów. Przy takim podejściu zadanie segmentacji obiektów sprowadza się do:

- znalezienia w obrazie tych elementów czyli ich segmentacji oraz
- ustalenia przynależności poszczególnych elementów do konkretnych obiektów.

Dla realizacji drugiego z tych zadań istotne jest to, że zwykle odległości między sąsiadującymi ze sobą elementami e_i należącymi do tego samego obiektu O_p są mniejsze niż odległości między sąsiadującymi ze sobą elementami należącymi do różnych obiektów. Wynika z tego, że powinien istnieć próg d_T pozwalający na rozróżnienie od siebie tych odległości, a przez to określenie zbiorów elementów tworzących poszczególne obiekty. Obserwacja tego faktu pozwala zaproponować przedstawiony w tabeli 1 dwuetapowy grafowy algorytm segmentacji obiektów w postaci skupisk. W celu zastosowania przedstawionej metody do konkretnego zastosowania konieczne jest ustalenie:

- jaka metoda segmentacji najlepiej sprawdzi się w zadaniu segmentacji elementów e_i tworzących obiekty,
- metody wyznaczania odległości między tymi elementami.

Tab. 1. Zarys algorytmu grafowej metody segmentacji

Tab. 1. Outline of the graph based image segmentation method

ALGORYTM Segmentacja obiektów w postaci skupisk

DANE: segmentowany obraz

PARAMETRY: etap 1. – parametry metody segmentacji, etap 2. – wartość progu d_T

ETAP 1.

Segmentacja obrazu w celu wydzielenia elementów e_i tworzących obiekty O_p

ETAP 2.

- Tworzenie minimalnych drzew rozpinających T_p — wierzchołki v_i grafu, na którym rozpinane jest T_p , reprezentują elementy e_i , a długość d_{ij} krawędzi e_{ij} łączącej wierzchołki v_i i v_j jest równa najmniejszej odległości między elementami e_i i e_j , przy czym $d_{ij} < d_T$
- Dla każdego spójnego drzewa T_p tworzona jest jego otoczka wypukła, która określa obszar obiektu O_p

W przypadku zadania segmentacji obrazów kometowych elementy e_i odpowiadają fragmentom DNA, a obiekty O_p odpowiadają poszczególnym kometom K_p .

Na podstawie przeprowadzonych testów do segmentacji poszczególnych fragmentów DNA zdecydowano się wykorzystać metodę binaryzacji lokalnej w postaci metody Bernsena [1] i następującej po niej filtracji medianowej jej wyniku (rys. 1, 2. kolumna). Podobnie jak w większości metod lokalnych lokalny próg binaryzacji $T(i, j)$ w metodzie Bersena wyznacza się na podstawie analizy otoczenia piksela (i, j) w postaci, zazwyczaj, prostokątnego okna $W(i, j)$ o wymiarach $m \times n$ (wymiarzy okna są zazwyczaj nieparzyste) i ze środkiem w punkcie (i, j) . Wartość lokalnego progu binaryzacji wyznaczany jest wg zależności

$$T(i, j) = \begin{cases} \frac{1}{2}(I_{max}(i, j) + I_{min}(i, j)) & \text{dla } C(i, j) \geq c \\ \tau_0 & \text{dla } C(i, j) < c \end{cases} \quad (1)$$

gdzie $C(i, j) = I_{max}(i, j) - I_{min}(i, j)$ jest kontrastem w oknie $W(i, j)$, a $I_{min}(i, j)$ i $I_{max}(i, j)$ to odpowiednio najmniejszy i największy poziom jasności w oknie $W(i, j)$, natomiast τ_0 to globalny próg binaryzacji, a c to progowa wartość kontrastu.

Wartość piksela obrazu po binaryzacji jest równa

$$B(i, j) = \begin{cases} 1 & \text{jeżeli } I(i, j) \geq T(i, j) \\ 0 & \text{jeżeli } I(i, j) < T(i, j) \end{cases} \quad (2)$$

gdzie $I(i, j)$ to poziom szarości piksela (i, j) binaryzowanego obrazu. Prezentowane w pracy wyniki zostały uzyskane dla następujących parametrów metody Bernsen: wielkość okna 5×5 , progowa wartość kontrastu $c = 22$, progu globalnego $\tau_0 = \tau_{Otsu} + 0,2$, gdzie τ_{Otsu} jest wartością progu wyznaczoną globalnie metodą Otsu [6] dla danego obrazu (przyjęto, że poziomy szarości obrazów zapisane są w zakresie $(0, 1)$), dla filtracji medianowej także przyjęto okno o wymiarach 5×5 .

Jako odległość d_{ij} między sąsiadującymi elementami e_i i e_j przyjęto najmniejszą odległość między pikselami należącymi do tych elementów — rozwiązanie to wymaga większych nakładów obliczeniowych niż wyznaczenie d_{ij} jako odległości między środkami ciężkości elementów e_i i e_j jednak daje znacznie lepsze wyniki.

Do tworzenia drzewa rozpoczynającego wykorzystano algorytm Kruskala [2, 3] (zarys algorytmu prezentuje tabela 2). Zaletą tego algorytmu w stosunku do innych standardowych algorytmów tworzenia minimalnego drzewa rozpoczynającego jest rozrost drzewa od najkrótszych krawędzi e_{ij} przez co możliwe jest stosowanie ograniczenia $d_{ij} < d_T$ już na etapie tworzenia drzew T_p , a nie dopiero po ich utworzeniu — takie rozwiązanie zmniejsza potrzebny nakład obliczeniowy.

Tab. 2. Algorytm Kruskala zapisany w pseudokodzie

Tab. 2. Kruskal's algorithm written in pseudocode

ALGORYTM Tworzenia minimalnego drzewa rozpoczynającego metodą Kruskala

DANE: graf G o wierzchołkach v_i i krawędziach e_{ij}

1. utwórz las L wierzchołków v_i (każdy wierzchołek tworzy osobne drzewo) oraz zbiór S krawędzi e_{ij}
2. **while** $S \neq \emptyset$
3. z S wybierz e_{ij} o najmniejszej długości i usuń ją z S
4. **if** e_{ij} łączy różne drzewa w L **then**
5. dodaj e_{ij} do L i połącz te drzewa
6. **else**
7. skasuj e_{ij}
8. **end**
9. **end**

Wyniki stosowania metody grafowej do segmentacji obrazów kometowych prezentuje rysunek 1. W celu usunięcia z rozwiązania obszarów, które nie są kometami można metodę uzupełnić o etap przetwarzania końcowego i poprzez porównanie cech geometrycznych lub fotometrycznych otrzymanych obszarów i obszarów wzorcowych można odrzucić obszary nie będące kometami.

4. Podsumowanie

Przeprowadzone do tej pory testy wskazują, że jakość segmentacji komet przy zastosowaniu prezentowanej metody jest znacznie lepsza niż ma to miejsce w przypadku stosowania standardowych

metod segmentacji i co ważne dla obrazów komet pozyskanych w podobnych warunkach nie było konieczności każdorazowej zmiany parametrów metody wraz ze zmianą obrazu testowego. Cecha ta powinna znacząco ułatwić stosowanie tej metody w praktyce. Planowane są badania mające na celu określenie zasad doboru parametrów metody na potrzeby segmentacji obrazów kometowych co może otworzyć drogę do automatyzacji segmentacji obrazów kometowych tą metodą.

Z punktu widzenia możliwych innych zastosowań metody grafowej niż tylko segmentacja komet istotne jest to, że do poprawnej pracy algorytmu konieczne jest spełnienie jedynie dwóch, jak się wydaje, niezbyt restrykcyjnych wymagań tj.:

1. spełnienie musi być założenie o możliwości stosowania progu d_T do rozróżnienia poszczególnych obiektów O_p na podstawie odległości między należącymi do nich elementami e_i , oraz
2. etap 1. powinien zakończyć się znalezieniem reprezentatywnego zbioru elementów e_i tworzących obiekty O_p .

Wymagania te mają charakter uniwersalny — w żaden sposób nie bazują na specyficznych cechach obrazów kometowych oraz na własnościach zastosowanej metody segmentacji elementów tworzących obiekty. Dzięki takim własnościom prezentowana metoda powinna znaleźć zastosowanie do segmentacji innych obiektów w postaci skupisk. Dostosowanie działania metody do innego typu obrazów wymaga jedynie ustalenia parametrów binaryzacji (ewentualnie wyboru innej metody segmentacji lepiej sprawdzającej się w danym zastosowaniu), określenia sposobu wyznaczania odległości między elementami tworzącymi obiekty oraz określenia wartości progu d_T . Dodatkowym atutem metody grafowej jest duża możliwość zrównoleglenia jej działania.

Uniwersalny charakter przyjętych w nowej metodzie rozwiązań powoduje, że podejście to może także zostać wykorzystane na etapie przetwarzania końcowego innych metod segmentacji do poprawy jakości wyników uzyskiwanych tymi metodami.

Praca została częściowo sfinansowana ze środków na działalność statutową Instytutu Automatyki Politechniki Śląskiej (BK 214/RAu1/2011 temat 4). Część badań wykorzystanych w pracy została przeprowadzona w Centrum Biotechnologii, Bioinżynierii i Bioinformatyki ŚLĄSKA BIO-FARMA w Gliwicach (<http://www.biofarma.polsl.pl/>).

5. Literatura

- [1] Bernsen J.: Dynamic thresholding of gray-level images, Proceedings 8th International Conference on Pattern Recognition, Paris, str. 1251–1255, 1986.
- [2] Cormen T. H., Leiserson Ch. E., Rivest R.: Wprowadzenie do algorytmów, WNT, 2004.
- [3] Kruskal J. B.: On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem, Proceedings of the American Mathematical Society, vol 7, no. 1, str. 48–50, 1956.
- [4] Liao W., McNutt M. A., Zhu W.G.: The comet assay: A sensitive method for detecting DNA damage in individual cells, Methods San Diego Calif, vol. 48, no. 1, str. 46–53, 2009.
- [5] Olive P. L.: Review. Impact of the comet assay in radiobiology, Mutation Research/Reviews in Mutation Research, vol. 681, no. 1, ss.: 13–23, 2009.
- [6] Otsu N.: A threshold selection method from gray-level histograms, IEEE Trans. Sys., Man., Cyber., vol. 9, no. 1, ss. 62–66, 1979.
- [7] Östling O., Johanson K. J.: Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells, Biochemical and Biophysical Research Communications, vol. 123, no. 1, str. 291–298, 1984.

otrzymano / received: 16.01.2012

przyjęto do druku / accepted: 02.03.2012

artykuł recenzowany / revised paper