

Spatial Fuzzy C-means dan Rapid Region Merging untuk Pemisahan Sel Kanker Payudara

Desmin Tuwohingide¹, Chastine Fatichah²

Abstract—Segmentation and overlapped cells separation are important phases in microscopic image processing of breast cancer, because the accuracy of overlapped cells separation result determines the accuracy of breast cancer cell calculation. The amount of breast cancer cells is considered by doctor in determining the action towards patients. Two of the most common topics discussed in previous studies are the problem of increasing the accuracy of overlapped cancer cell separation result by calculating the number of cancer cell and over-segmentation problem. Compared to watershed method, clustering method produces higher accuracy in separating overlapped cancer cells. In this paper, a combination of Spatial Fuzzy C-Means (SFCM) and Rapid Region Merging (RRM) method is proposed to separate the overlapped cells and handling the over-segmentation problem. The input image of overlapped cells separation phase is the result of breast cancer cell identification by Gram-Schmidt (GS) method, while the clustered cancer cells are overlapped cancer cells which are detected based on the area of geometric feature. 40 microscopic breast cancer cells image of benign and malignant type is used as the datasets. The average value of Mean Square Error (MSE) for cell identification is 0.07 and the average accuracy of overlapped cells separation using SFCM and RRM is 78.41%.

Intisari—Segmentasi dan pemisahan sel bertumpuk merupakan tahapan penting dalam proses pengolahan citra mikroskopis sel kanker payudara, karena akurasi hasil pemisahan sel bertumpuk akan menentukan akurasi perhitungan sel kanker payudara. Hasil perhitungan jumlah sel kanker payudara merupakan salah satu faktor pertimbangan bagi dokter untuk melakukan tindakan terhadap pasien. Masalah peningkatan akurasi hasil pemisahan sel kanker bertumpuk berdasarkan perhitungan jumlah sel kanker dan masalah *over-segmentation* adalah topik yang paling umum dibahas pada penelitian-penelitian sebelumnya. Metode *clustering* yang digunakan untuk pemisahan sel kanker darah bertumpuk menghasilkan nilai akurasi yang tinggi dibandingkan metode *watershed*. Pada makalah ini diusulkan kombinasi metode *Spatial Fuzzy C-Means* (SFCM) dan *Rapid Region Merging* (RRM) untuk pemisahan sel bertumpuk dan penanganan masalah *over-segmentation*. Citra masukan pada tahapan pemisahan sel bertumpuk adalah citra hasil identifikasi sel kanker payudara berdasarkan metode *Gram-Schmidt* (GS), sedangkan sel kanker yang akan dipisahkan adalah sel kanker

yang dideteksi bertumpuk berdasarkan ciri (*feature*) geometri area. *Data set* yang digunakan sebanyak 40 citra mikroskopis sel kanker payudara jenis *benign* dan *malignant*. Pada makalah ini dihasilkan nilai rata-rata *Mean Square Error* (MSE) identifikasi sel sebesar 0,07 dan rata-rata akurasi hasil pemisahan sel bertumpuk menggunakan metode SFCM dan RRM sebesar 78,41%.

Kata Kunci— Segmentasi Sel Bertumpuk, Citra Mikroskopis, Sel Kanker Payudara, *Spatial Fuzzy C-Means*, *Rapid Region Merging*

I. PENDAHULUAN

Pada umumnya, kanker payudara didiagnosis melalui pemeriksaan rutin atau ketika pasien menyadari gejala-gejala tertentu. Tetapi pemeriksaan fisik saja tidak cukup untuk mengonfirmasi diagnosis kanker payudara, sehingga dilakukan perangkaian prosedur pemeriksaan seperti mamografi, biopsi, dan lain sebagainya untuk memperoleh informasi terkait diagnosis dan kebutuhan terapi pasien. Salah satu data yang informasinya digunakan untuk diagnosis penyakit kanker payudara adalah data sel kanker berupa citra mikroskopis yang diambil dari tubuh pasien menggunakan kamera khusus. Citra mikroskopis tersebut yang kemudian akan dianalisis oleh tim medis untuk mengonfirmasi kesehatan pasien.

Penerapan teknik pengolahan citra untuk menganalisis citra mikroskopis sel kanker payudara dilakukan untuk mempermudah diagnosis penyakit kanker payudara. Pada penerapannya, terdapat berbagai permasalahan dalam implementasi pengolahan citra, antara lain teknik segmentasi tradisional tidak bisa secara maksimal melakukan segmentasi dan pemisahan sel bertumpuk pada citra mikroskopis yang memiliki beragam karakteristik dari segi warna, bentuk, ukuran dan tekstur. Keberagaman ini yang menjadi tantangan dalam pengolahan citra mikroskopis, sehingga banyak dilakukan penelitian dan pengembangan metode terkait pengolahan citra mikroskopis sel kanker payudara.

Masalah identifikasi sel kanker payudara pada citra mikroskopis sel kanker payudara telah dipaparkan pada beberapa penelitian [1], [2]. Permasalahan lainnya adalah proses pemisahan sel bertumpuk pada citra mikroskopis sel kanker payudara. Proses pemisahan sel bertumpuk merupakan salah satu tahapan terpenting dalam proses segmentasi citra mikroskopis sel kanker payudara dikarenakan keakuratan hasil pemisahan sel bertumpuk akan menentukan akurasi perhitungan jumlah sel kanker payudara. Hasil perhitungan sel kanker payudara merupakan salah satu faktor pertimbangan bagi tim medis untuk melakukan tindakan terhadap pasien.

Beberapa penelitian telah melakukan segmentasi dan pemisahan sel bertumpuk pada citra mikroskopis. Sebuah penelitian menerapkan deteksi sel kanker payudara bertumpuk

¹Mahasiswa, Jurusan Teknik Informatika Fakultas Teknologi Informasi Institut Teknologi Sepuluh Nopember, Jln Teknik Kimia, Gedung Informatika, Kampus ITS Sukolilo, Kota Surabaya, Jawa Timur 60111 (telp: 031-5939214; fax: 031-5913804; e-mail: desmin.tuwohingide@gmail.com)

²Dosen, Jurusan Teknik Informatika Fakultas Teknologi Informasi Institut Teknologi Sepuluh Nopember, Jln Teknik Kimia, Gedung Informatika, Kampus ITS Sukolilo, Kota Surabaya, Jawa Timur 60111 (telp: 031-5939214; fax: 031-5913804; e-mail: chastine.fatichah@gmail.com)

menggunakan deteksi *concave point* [3]. Karena bentuk sel kanker payudara yang beragam, dilakukan deteksi titik yang memiliki *concavity* yang tinggi pada daerah *boundaries*. Titik-titik *concavity* ditentukan dengan teknik pendeteksian sudut berdasarkan informasi *global* dan *local curvature* dan metode *Watershed* untuk pemisahan sel bertumpuk. Hasil nilai akurasi yang diperoleh pada penelitian tersebut adalah 89,20% untuk *data set 1* dan 92,70% untuk *data set 2*. Secara berkelanjutan, beberapa penelitian yang menerapkan *Watershed* untuk pemisahan sel bertumpuk berdasarkan *fuzzy active contour model* dan deteksi *concave point* menggunakan konstruksi *concave vertex graph* dan pemilihan jalur garis terpendek untuk inialisasi garis pemisah, dan menghasilkan nilai akurasi sebesar 96,23% [2], [4], [5]. Dalam sebuah penelitian lain, kombinasi metode *Region-Based Active Contour* dan *Bayesian* untuk pemisahan sel bertumpuk menghasilkan nilai rata-rata akurasi sebesar 61,30%, sementara kombinasi metode *Region-Based Active Contour* dan *Watershed* menghasilkan rata-rata nilai akurasi sebesar 65% [6]. Selanjutnya, telah dilakukan juga penelitian deteksi dan pemisahan sel bertumpuk pada citra mikroskopis sel darah menggunakan informasi ciri (*feature*) geometri dan analisis *K-Means Clustering* untuk pemisahan sel kanker bertumpuk [7]. Hasilnya menunjukkan bahwa metode ini memiliki rata-rata akurasi yang lebih tinggi, yaitu 96,46%, dan nilai galat (*error*) yang lebih rendah dibandingkan dengan metode *Classical Watershed* yang memperoleh rata-rata nilai akurasi sebesar 79,87% dan metode *Condition Erosion Watershed* yang menghasilkan rata-rata nilai akurasi sebesar 87,22%.

Berdasarkan penelitian-penelitian tersebut, ada beberapa hal yang sangat untuk dilakukan dan dikembangkan agar dihasilkan segmentasi dan pemisahan sel kanker bertumpuk yang nilai akurasinya tinggi. Identifikasi sel kanker payudara menggunakan metode *Gram-Schmidt* (GS) merupakan salah satu metode yang memberikan nilai *Mean Square Error* (MSE) yang rendah [1]. Deteksi sel kanker payudara perlu dilakukan untuk mengurangi kemungkinan *over-segmentation* dan *under-segmentation* berdasarkan informasi ciri (*feature*) geometri area dengan memberikan asumsi luasan sel kanker tunggal dan luasan sel kanker bertumpuk [8]. Untuk pemisahan sel bertumpuk, metode *clustering* menghasilkan pemisahan sel darah bertumpuk yang baik dan memperoleh nilai akurasi yang tinggi dibandingkan dengan metode *classical watershed* dan metode *condition erosion watershed*, sehingga dapat dikembangkan lagi untuk menghasilkan nilai akurasi yang semakin tinggi [7].

Pada penelitian lain, dipaparkan metode *clustering Spatial Fuzzy C-Means* (SFCM) yang merupakan metode *clustering* berdasarkan informasi spasial dan probabilitas ketetangaan, serta merupakan pengembangan dari metode *Fuzzy C-Means clustering*, karena informasi spasial dan probabilitas ketetangaan memiliki korelasi yang tinggi dan sangat penting, sehingga algoritme ini tidak hanya tegar (*robust*) terhadap derau tapi juga mengurangi kesalahan pengelompokan [9], [10].

Selain beberapa masalah tersebut, *over-segmentation* menjadi salah satu masalah utama yang harus ditangani dalam

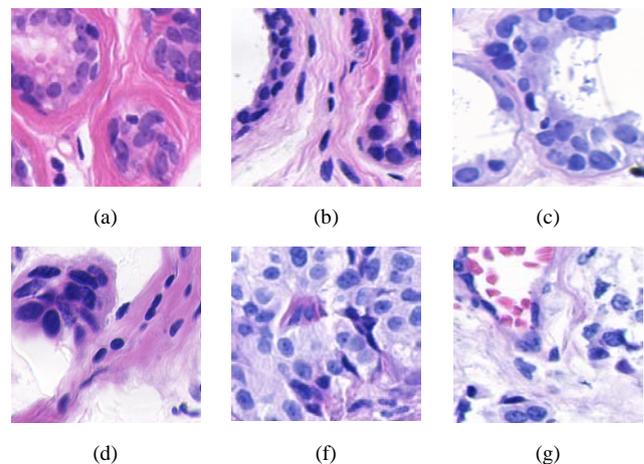
proses pemisahan sel bertumpuk. Metode *Region Merging* adalah metode yang diimplementasikan untuk mengurangi masalah *over-segmentation*. Salah satunya adalah *Rapid Region Merging* (RRM), yang merupakan metode penggabungan daerah berdasarkan kemiripan region, yang mampu meningkatkan nilai akurasi segmentasi karena mampu mengatasi masalah *over-segmentation* [11].

Berdasarkan uraian di atas, pada makalah ini diusulkan metode untuk pemisahan sel kanker bertumpuk pada citra mikroskopis sel kanker payudara menggunakan SFCM dan RRM. Tahapan pertama adalah melakukan segmentasi sel kanker pada citra mikroskopis sel kanker payudara berdasarkan identifikasi sel menggunakan metode GS, kemudian dilakukan deteksi sel kanker bertumpuk dan inialisasi jumlah sel bertumpuk berdasarkan informasi ciri geometri area. Selanjutnya adalah melakukan pemisahan sel kanker bertumpuk menggunakan *clustering* SFCM, sementara metode RRM mengatasi masalah *over-segmentation*.

II. USULAN METODE

A. Data

Data set yang digunakan adalah data citra mikroskopis sel kanker payudara, sebanyak 40 citra, yang terdiri atas *data set* tipe *malignant* dan *benign*. *Data set* dapat diunduh di www.bioimage.ucsb.edu. Berdasarkan *data set groundtruth* yang diperoleh, diperlukan proses *cropping* citra secara manual untuk memperoleh bagian citra yang sesuai dengan *groundtruth* menggunakan *tools* Adobe Photoshop CS3, seperti pada Gbr. 1.



Gbr. 1 Citra mikroskopis sel kanker payudara (a)(b)(c) Citra data *benign* (d)(e)(f) Citra data *malignant*.

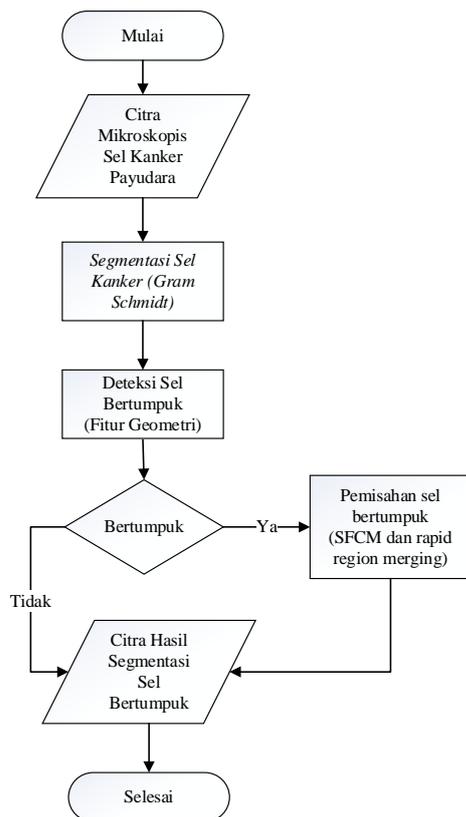
B. Identifikasi Sel Kanker dan Pemisahan Sel Bertumpuk

Secara garis besar, penelitian ini terbagi ke dalam dua tahapan penting. Tahapan pertama adalah identifikasi sel kanker payudara atau tahapan segmentasi awal dan tahapan kedua adalah pemisahan sel kanker bertumpuk, sesuai yang disajikan pada Gbr. 2.

Tahapan pertama adalah tahapan identifikasi sel kanker payudara berdasarkan metode GS [1]. Pada metode ini, dilakukan penentuan tiga vektor warna. Vektor warna pertama

adalah vektor warna yang akan diperkuat atau diinisialisasi sebagai sel kanker, sedangkan vektor warna kedua dan ketiga adalah vektor warna yang dianggap sebagai derau dan *background*. Penentuan vektor warna dilakukan dengan memilih tiga citra *data set* secara acak dan mengambil nilai tengahnya sebagai vektor warna. Kemudian dilakukan *thresholding* untuk memperoleh citra biner. Nilai *thresholding* yang digunakan adalah 0,03. Nilai ini diperoleh berdasarkan uji coba terhadap 20 *data set* yang dipilih secara acak. Proses terakhir pada tahapan identifikasi sel adalah penghapusan derau menggunakan operasi morfologi.

Tahapan kedua adalah pemisahan sel kanker bertumpuk berdasarkan metode SFCM dan RRM. Tahapan ini terdiri atas beberapa proses, yaitu deteksi sel kanker bertumpuk, pemisahan sel kanker bertumpuk menggunakan SFCM, dan penggabungan daerah yang mengalami *over-segmentation* berdasarkan RRM. Data masukan yang digunakan pada tahapan ini adalah citra hasil segmentasi sel kanker payudara berdasarkan metode GS, yang diperoleh pada tahapan pertama. Deteksi sel bertumpuk dilakukan berdasarkan informasi ciri geometri area dengan mengasumsikan setiap sel yang memiliki nilai ciri geometri luasan di atas nilai *threshold* adalah bertumpuk. Nilai *threshold* ditentukan berdasarkan uji coba terhadap 20 citra yang dipilih secara acak. Nilai *threshold* dengan hasil terbaik adalah 340.



Gbr. 2 Diagram alir tahapan yang dilakukan.

1) *Spatial Fuzzy C-Means*: Salah satu karakteristik penting sebuah citra adalah ketetanggaan piksel yang berkorelasi tinggi. Dengan kata lain, ketetanggaan piksel memiliki

kemiripan nilai ciri, dan kemungkinan berada dalam *cluster* yang sama sangat besar. Hubungan spasial adalah hal penting dalam *clustering*, tetapi hal ini tidak digunakan pada *Fuzzy C-Means* (FCM) standar, sehingga FCM standar sangat rentan terhadap derau. Untuk mengatasi masalah tersebut, dapat dimanfaatkan informasi spasial yang dibentuk menggunakan distribusi statistika dari ketetanggaan piksel dan *prior probability* untuk memperoleh sebuah fungsi keanggotaan (*membership function*) baru yang digunakan pada SFCM [9]. Metode ini tidak hanya efektif menghapus derau tetapi juga mengurangi kesalahan pengelompokan piksel.

Diasumsikan sejumlah data dalam himpunan data X yang berisi sejumlah n data yang dinotasikan $X = \{x_1, x_2, \dots, x_n\}$, dengan setiap data mempunyai ciri r dimensi $x_{i1}, x_{i2}, \dots, x_{ir}$ yang dinotasikan $x_i = \{x_{i1}, x_{i2}, \dots, x_{ir}\}$. Ada sejumlah *cluster* C dengan titik pusat *cluster* c_1, c_2, \dots, c_k , dengan k adalah jumlah *cluster*. Setiap data mempunyai derajat keanggotaan pada setiap *cluster*, yang dinyatakan dengan u_{ij}^* , seperti ditunjukkan pada (1), dengan i menyatakan data x_i dan j menyatakan c_j . Jumlah nilai derajat keanggotaan setiap data x_i selalu sama dengan 1, seperti pada (5) [9].

$$u_{ij}^* = u_{ij}^p h_{ij}^q / \sum_{k=1}^c u_{kj}^p h_{kj}^q \quad (1)$$

$$\sum_{j=1}^k u_{ij}^* = 1 \quad (2)$$

Berdasarkan (2) dengan parameter c_j adalah *centroid cluster* ke- j , $D()$ adalah jarak data dan *centroid*, sedangkan w adalah parameter bobot pangkat (*weighting exponent*). w tidak memiliki nilai ketetapan, biasanya nilai $w > 1$ dan umumnya diberi nilai 2 [9].

$$u_{ij} = D(x_i, c_j)^{-2/w-1} / \sum_{l=1}^k D(x_i, c_l)^{-2/w-1} \quad (3)$$

Perhitungan *spatial function* sesuai (3) dengan x_j merepresentasikan nilai ciri, Ω merepresentasikan sebuah jendela persegi yang titik tengahnya berada pada piksel x_j . h_{ij} merepresentasikan probabilitas piksel x_j menjadi anggota sebuah *cluster*. Pada *spatial function*, c merupakan jumlah *cluster* yang diinginkan, P_{ij} adalah faktor kontribusi dari tetangga x_t , dan bobot u_{it} adalah *membership pattern* x_t dan *prior probability*. Secara umum, *prior probability* dari p_{ij} merupakan hasil perbandingan dari N_j^i jumlah ketetanggaan piksel yang menjadi anggota *cluster* dan N_j jumlah ketetanggaan piksel pada x_j . [9].

$$h_{ij} = 1 + \left(\sum_{t \in \Omega_j} u_{it} p_{ij} / \sum_{k=1}^c \sum_{t \in \Omega_j} u_{kt} p_{kj} \right) \quad (4)$$

$$p_{ij} = N_j^i / N_j \quad (5)$$

Untuk menghitung titik pusat *cluster* c_l pada fitur j , digunakan (6), dengan parameter N adalah jumlah data, w adalah bobot pangkat, dan u_{ij} adalah derajat keanggotaan data x_i ke *cluster* c_l . Sementara, fungsi objektif dihitung sesuai (9) [9].

Penentuan jumlah *cluster* dilakukan berdasarkan perhitungan sesuai dengan (7), dengan a adalah ciri geometri area bertumpuk, b adalah nilai *threshold* bertumpuk, dan c adalah nilai *threshold* sebesar 1,5. Nilai c diperoleh

berdasarkan uji coba terhadap *data set* dari rentang nilai 1,1 sampai 2. Nilai objektif dihitung berdasarkan (8).

$$c_{ij} = \sum_{l=1}^N (u_{il})^w x_{ij} / \sum_{l=1}^N (u_{il})^w \tag{6}$$

$$k = a / (b/c) \tag{7}$$

$$J = \sum_{i=1}^N \sum_{l=1}^k (u_{ij})^w D(x_i, c_l)^2 \tag{8}$$

Tahapan metode SFCM yang dilakukan dapat disimpulkan sebagai berikut.

1. Inisialisasi jumlah *cluster* ($k \geq 2$) berdasarkan asumsi banyaknya sel tunggal yang menumpuk pada daerah yang dianggap sel bertumpuk, mengikuti perhitungan pada (7).
2. Penentuan bobot pangkat ($w \geq 1$), penentuan jumlah iterasi maksimal, penentuan nilai ambang batas perubahan fungsi objektif.
3. Memberikan nilai awal pada matriks *fuzzy pseudo-partition*, dengan syarat seperti pada (2) dan menghitung nilai *centroid* awal menggunakan (6).
4. Menghitung *spatial function* dan *prior probability* ketetanggaaan menggunakan (4) dan (5).
5. Menghitung kembali keanggotaan matriks *fuzzy pseudo-partition* menggunakan (1) dan memperbarui nilai titik pusat *cluster*.
6. Memperbarui perhitungan. Jika memenuhi syarat maksimum iterasi, atau syarat perubahan nilai objektif di bawah nilai ambang batas, maka berhenti. Jika tidak, kembali ke tahapan ke-2.

2) *Rapid Region Merging (RRM)*: Metode RRM merupakan metode penggabungan sebuah daerah yang memiliki *maximum similarity* dengan daerah tetangganya atau daerah yang memiliki *region adjacent* dengan dirinya.

Tahapan penggabungan daerah menggunakan algoritme RRM dimulai dengan menentukan terlebih dahulu daerah yang *pseudo-blob* dan daerah yang *real-blob*. Pada makalah ini, daerah *pseudo-blob* ditentukan berdasarkan informasi ciri geometri area. Selanjutnya, daerah yang dianggap sebagai daerah *pseudo-blob* atau daerah yang terbentuk akibat *over-segmentation* akan diproses dengan menghitung nilai kemiripannya dengan daerah *real-blob* yang merupakan tetangganya atau yang memiliki *region adjacent* dengan *pseudo-blob*. Penggabungan daerah *pseudo-blob* dan *real-blob* dilakukan berdasarkan perhitungan nilai *maximum similarity*. Daerah tetangga yang memiliki nilai *maximum similarity* akan ditandai sebagai *winner block P*. *Winner block P* ditandai berdasarkan (9) dan parameter yang digunakan untuk menghitung kemiripan antara *pseudo-blob* dan *real-blob* adalah *gray level contrasts* $G(a, b)$, kemiripan *ridge* $E(a, b)$, dan *region adjacency* $A(a, b)$ [11].

$$Block P = Max(Sim(a, b), Sim(a, c), Sim(a, d)) \tag{9}$$

Penggabungan daerah menggunakan metode RRM dilakukan dengan cara menghitung kemiripan daerah yang dianggap sebagai daerah *over-segmentation* atau *pseudo-blob* dengan sel tetangganya yang dianggap *real-blob*. Untuk menghitung kemiripan antar *cluster*, parameter yang digunakan adalah *gray level contrasts* $G(a, b)$, kemiripan

ridge $E(a, b)$, dan *region adjacency* $A(a, b)$ mengikuti (10). *Gray level contrasts* $G(a, b)$ dihitung berdasarkan jumlah piksel di region a, R_a , dan region b, R_b , serta rata-rata keabuan pada region a, μ_a , dan region b, μ_b , mengikuti (11).

$$Sim(a, b) = 1/G(a, b) \cdot E(a, b) \cdot A(a, b) \tag{10}$$

$$G(a, b) = (R_a - R_b / R_a + R_b) (\mu_a - \mu_b)^2 \tag{11}$$

$$E(a, b) = (|\pi_a - \pi_b| + 1) / (\sigma_a + \sigma_b + 1) \tag{12}$$

Kemiripan *ridge* $E(a, b)$ dihitung mengikuti (12), dengan σ adalah rata-rata keabuan piksel di daerah *boundary* pada kedua sisi *ridge* dan π adalah variasi keabuan piksel di daerah *boundary* pada kedua sisi. Jika *ridge* adalah *boundary*, *gray level* piksel pada kedua sisi akan menunjukkan perubahan yang signifikan, $|\pi_a - \pi_b|$ akan menjadi lebih besar, sedangkan σ_a dan σ_b akan menjadi lebih kecil. Aturan *region adjacency* $A(a, b)$ mengikuti (13) [11].

$$A(a, b) = \begin{cases} 1 & a \text{ dan } b \text{ adjacent} \\ \infty & a \text{ dan } b \text{ tidak adjacent} \end{cases} \tag{13}$$

C. Skenario Uji Coba dan Metode Evaluasi

Pengukuran unjuk kerja metode yang diusulkan perlu dilakukan untuk mengetahui tingkat keberhasilan. Dilakukan dua skenario uji coba untuk mengukur keberhasilan metode usulan yang diterapkan pada tahapan identifikasi sel kanker dan tahapan pemisahan sel kanker bertumpuk.

Skenario pertama adalah membandingkan hasil segmentasi awal atau hasil identifikasi sel pada citra mikroskopis sel kanker payudara menggunakan metode yang diusulkan dengan metode *Region Based Active Contour (RBAC)* [6]. Skenario kedua adalah membandingkan hasil pemisahan sel bertumpuk berdasarkan perhitungan jumlah sel menggunakan metode yang diusulkan dan dengan menggunakan metode *K-Means*.

Evaluasi skenario uji coba pertama dilakukan dengan menghitung nilai MSE atau rata-rata galat untuk mengukur keberhasilan metode yang diusulkan dan metode pembandingan dalam melakukan identifikasi sel berdasarkan citra *groundtruth*. Nilai MSE merupakan hasil perbandingan nilai selisih piksel-piksel pada citra asal $f(x, y)$ dengan piksel-piksel pada citra hasil $g(x, y)$ di posisi piksel yang sama, yang secara matematis dinyatakan dalam (14). Semakin kecil nilai MSE yang diperoleh, maka semakin tinggi akurasi identifikasi sel yang diperoleh.

$$MSE = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N [f(x, y) - g(x, y)]^2 \tag{14}$$

TABEL I
CONFUSION MATRIX

Confusion Matrix		Predicted	
		Positive	Negative
Actual	Positive	TP	TN
	Negative	FP	FN

Evaluasi skenario uji coba kedua dilakukan dengan menghitung nilai akurasi untuk mengukur keberhasilan metode yang diusulkan dan metode pembandingan. Sel kanker yang diidentifikasi benar adalah sel kanker yang terpisah sesuai dengan jumlah sel bertumpuk pada citra asli dengan

memperhatikan ketepatan garis pemotong pada kondisi tertentu. Selain itu, jumlah sel yang mengalami *over-segmentation* dihitung berdasarkan jumlah *cluster*, sedangkan *under-segmentation* dihitung berdasarkan jumlah *cluster* yang kurang dari jumlah sel bertumpuk yang sebenarnya. Nilai akurasi yang dihasilkan oleh metode yang diusulkan akan dibandingkan dengan nilai akurasi metode *K-Means Clustering* [7]. Nilai akurasi dihitung berdasarkan perbandingan jumlah sel benar dan jumlah sel pada citra asli sesuai dengan (15) berdasarkan *confusion matrix* pada Tabel I, dengan *true positive (TP)* adalah jumlah sel yang diidentifikasi dengan benar dan *true negative (TN)* adalah jumlah sel benar yang tidak teridentifikasi.

$$\text{Akurasi} = TP / (TP + TN) \times 100 \quad (15)$$

III. HASIL DAN ANALISIS

A. Hasil Uji Coba Skenario 1

Pada Tabel II, ditunjukkan hasil identifikasi sel menggunakan metode GS dan metode RBAC.

Uji coba skenario pertama dilakukan dengan membandingkan hasil identifikasi sel kanker dari metode yang diusulkan dengan metode RBAC dalam melakukan segmentasi awal atau identifikasi kanker berdasarkan perhitungan nilai MSE. Berdasarkan perhitungan nilai MSE yang ditampilkan pada Tabel II, dengan metode GS menghasilkan nilai MSE yang berada dalam rentang nilai 0,0398 sampai 0,1220 dan nilai rata-rata kesalahan segmentasi sebesar 0,0745, nilai MSE terendah dihasilkan oleh citra ke-11 dan nilai MSE tertinggi dihasilkan oleh citra ke-39. Identifikasi menggunakan metode RBAC menghasilkan nilai MSE dalam rentang nilai 0,0728 sampai 0,9578, dengan nilai MSE terendah dihasilkan oleh citra ke-12 dan nilai MSE tertinggi berada pada citra ke-10, dengan rata-rata kesalahan segmentasi sebesar 0,6505.

B. Hasil Uji Coba Skenario 2

Uji coba skenario kedua dilakukan untuk mengetahui perbandingan nilai akurasi tahapan pemisahan sel menggunakan metode usulan dan metode *K-Means* berdasarkan perhitungan jumlah sel yang dipisahkan secara benar.

Sel benar dihitung berdasarkan jumlah *cluster* yang sama dengan jumlah sel pada citra asli dan dengan memperhatikan garis pemisah. Sel *over-segmentation* dihitung berdasarkan sel yang memiliki jumlah *cluster* lebih dari jumlah sel pada citra asli. Sel *under-segmentation* adalah sel yang memiliki jumlah *cluster* kurang dari jumlah sel pada citra asli. Nilai akurasi diperoleh berdasarkan perbandingan jumlah sel yang benar dengan jumlah sel kanker pada citra asli.

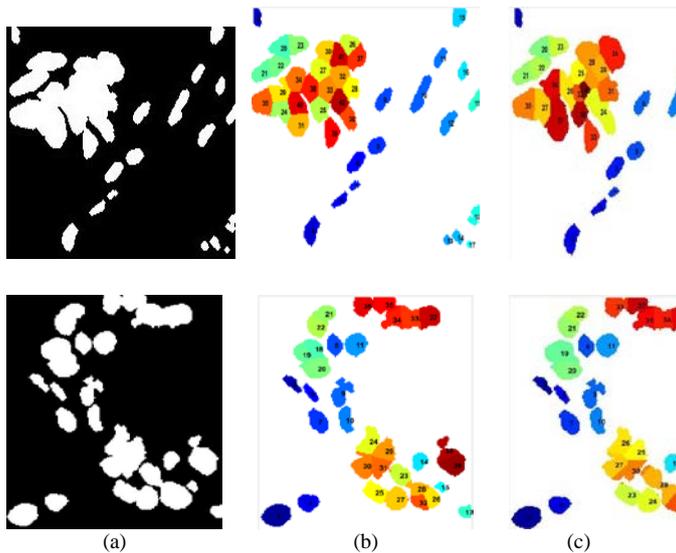
Berdasarkan hasil pemisahan sel bertumpuk pada 40 citra mikroskopis sebagai data uji yang disajikan pada Gbr. 3, dapat dilihat perbedaan hasil pemisahan sel kanker bertumpuk menggunakan metode usulan dan metode *K-Means*.

Pemisahan sel kanker bertumpuk menggunakan metode *K-Means* menghasilkan lebih banyak sel yang *over-segmentation* dibandingkan dengan metode yang diusulkan. Berdasarkan

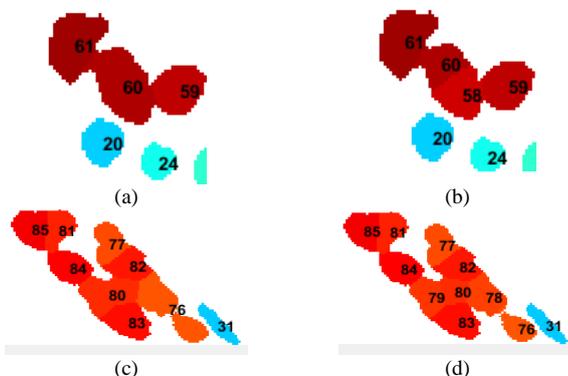
hasil tersebut dapat dihitung jumlah sel yang benar, jumlah sel *over-segmentation*, dan jumlah sel *under-segmentation* seperti yang ditampilkan pada Gbr. 4. Sel benar dihitung berdasarkan jumlah *cluster* dan hasil *clustering* yang sesuai dengan jumlah sel pada citra asli seperti yang ditampilkan pada Gbr.4(a). Sel *over-segmentation* berarti jumlah *cluster* lebih banyak dari jumlah sel pada citra asli, seperti yang ditampilkan pada Gbr. 4(b), ditandai dalam sel yang diberi label 58 dan 60, dan pada Gbr. 4(d) ditandai dalam sel yang diberi label 80 dan 78. Sel *under-segmentation* dihitung berdasarkan jumlah *cluster* yang lebih sedikit dari jumlah sel pada citra asli, seperti pada Gbr. 4(c) di sel yang dilabeli angka 76.

TABEL II
HASIL PERHITUNGAN NILAI MSE PADA UJI COBA SKENARIO 1

Citra Ke	Nilai MSE	
	<i>Gram-Schmidt</i>	RBAC
1	0,0622	0,8855
2	0,0719	0,0816
3	0,0822	0,8646
4	0,1054	0,9002
5	0,0989	0,9158
6	0,0850	0,9119
7	0,0458	0,0610
8	0,0990	0,9045
9	0,0552	0,9346
10	0,0461	0,9578
11	0,0398	0,9177
12	0,0541	0,0728
13	0,0844	0,1384
14	0,0827	0,1706
15	0,0604	0,1383
16	0,0854	0,1552
17	0,0879	0,8384
18	0,0418	0,9061
19	0,1035	0,8658
20	0,0859	0,7977
21	0,0851	0,8573
22	0,0552	0,1959
23	0,0899	0,1626
24	0,0642	0,2253
25	0,0794	0,8363
26	0,0589	0,8411
27	0,0585	0,8836
28	0,0960	0,7324
29	0,0409	0,7505
30	0,0662	0,8508
31	0,0782	0,9105
32	0,0785	0,0991
33	0,1080	0,8859
34	0,0368	0,8955
35	0,0785	0,9123
36	0,0647	0,8650
37	0,0958	0,1352
38	0,0770	0,8879
39	0,1225	0,8311
40	0,0335	0,8428
Rata-rata	0,0736	0,6505



Gbr. 3 Hasil pemisahan sel bertumpuk (a) Citra *groundtruth* (b) Pemisahan sel bertumpuk menggunakan *K-Means* (c) Pemisahan sel bertumpuk menggunakan kombinasi SFCM dan RRM.



Gbr. 4 (a) Sel Benar, (b) Sel *Over-segmentation*, (c) Sel *Under-segmentation*, (d) Sel *Over-segmentation*.

Dari hasil perhitungan sel secara manual berdasarkan citra hasil pemisahan sel bertumpuk pada 40 citra *data set*, didapatkan hasil uji coba skenario 2 seperti disajikan pada Tabel III dan Tabel IV, berturut-turut menggunakan metode usulan dan metode *K-Means*. Berdasarkan hasil perhitungan manual yang disajikan pada Tabel III, kombinasi metode SFCM dan RRM menghasilkan nilai akurasi terendah sebesar 42,11% pada citra ke-14 dan nilai akurasi tertinggi sebesar 95,83% pada oleh citra ke-20, serta menghasilkan rata-rata akurasi sebesar 78,41%. Hasil perhitungan manual berdasarkan hasil pemisahan sel bertumpuk menggunakan metode *K-Means*, seperti yang ditampilkan pada Tabel IV, menunjukkan nilai akurasi terendah sebesar 21,05% yang dihasilkan oleh citra ke-14 dan nilai akurasi tertinggi sebesar 86,57% pada citra ke-5, dengan rata-rata akurasi sebesar 68,73%. Perbandingan hasil perhitungan jumlah sel kanker benar, sel *over-segmentation*, dan sel *under-segmentation* dari metode yang diusulkan dan metode *K-Means* terhadap 40 citra mikroskopis sel kanker payudara dapat dilihat pada Tabel V. Jumlah total sel kanker pada 40 citra mikroskopis sebanyak 1856. Jumlah sel yang diperoleh dengan metode *K-Means*

adalah sebanyak 1333 sel benar, 382 sel *over-segmentation*, dan 35 sel *under-segmentation*. Jumlah sel yang diperoleh dengan metode usulan adalah sebanyak 1486 sel benar, 197 sel *over-segmentation*, dan 72 sel *under-segmentation*.

TABEL III
PERHITUNGAN JUMLAH SEL BERDASARKAN CITRA HASIL PEMISAHAN SEL BERTUMPUK MENGGUNAKAN METODE USULAN PADA UJI COBA SKENARIO 2

Citra ke	Sel di Citra Asli	Sel Benar	Sel Over	Sel Under	Akurasi (%)
1	47	34	10	3	70,21
2	44	34	8	1	75,00
3	34	26	5	0	85,29
4	46	32	10	0	69,57
5	67	54	1	5	80,60
6	57	44	1	1	78,95
7	42	32	2	1	76,19
8	28	16	13	0	55,56
9	27	19	7	0	70,73
10	29	17	8	1	67,86
11	32	27	3	1	84,38
12	47	40	2	1	89,63
13	62	52	2	2	83,87
14	19	9	9	0	42,11
15	27	19	8	2	70,37
16	40	34	6	0	85,00
17	74	61	0	5	86,49
18	36	27	1	3	86,11
19	71	56	0	4	78,87
20	26	23	1	0	95,83
21	82	67	1	7	82,93
22	42	34	2	2	80,95
23	31	20	10	1	64,52
24	23	20	4	0	81,82
25	46	41	3	0	89,13
26	40	34	4	3	82,50
27	50	35	12	1	74,00
28	61	50	9	1	81,97
29	43	40	0	2	93,02
30	39	30	2	4	82,05
31	74	62	6	6	83,78
32	89	72	0	4	84,27
33	53	36	12	0	69,81
34	26	20	6	0	69,23
35	47	36	3	1	76,60
36	43	36	3	1	83,72
37	54	41	9	0	84,48
38	31	20	8	0	68,00
39	108	94	1	9	87,04
40	26	22	4	0	84,62
Rata-rata akurasi					78,41

C. Analisis Hasil Uji Coba

Berdasarkan hasil uji coba pada skenario 1, metode yang diusulkan menghasilkan nilai MSE yang lebih baik dibandingkan metode RBAC. Rata-rata kesalahan segmentasi metode yang diusulkan adalah 0,0745, sedangkan rata-rata kesalahan segmentasi metode RBAC adalah 0,6505. Hal ini menunjukkan bahwa metode yang diusulkan mampu memberikan hasil identifikasi sel yang lebih baik dibandingkan dengan metode RBAC.

TABEL IV
PERHITUNGAN JUMLAH SEL BERDASARKAN CITRA HASIL PEMISAHAN SEL
BERTUMPUK MENGGUNAKAN METODE *K-MEANS CLUSTERING* PADA UJI
COBA SKENARIO 2

Citra ke	Sel di Citra Asli	Sel Benar	Sel Over	Sel Under	Akurasi (%)
1	47	28	17	1	59,57
2	44	25	15	2	56,82
3	34	24	10	0	70,59
4	46	28	13	0	60,87
5	67	58	0	2	86,57
6	57	42	4	1	73,68
7	42	27	7	1	64,29
8	28	14	14	0	51,85
9	27	16	10	0	59,26
10	29	17	9	0	60,71
11	32	22	8	0	68,75
12	47	35	8	0	74,47
13	62	43	12	1	69,35
14	19	4	13	0	21,05
15	27	17	10	0	62,96
16	40	27	13	0	67,50
17	74	61	3	5	82,43
18	36	30	3	2	83,33
19	71	55	0	5	77,46
20	26	18	6	0	75,00
21	82	70	3	3	85,37
22	42	27	11	0	64,29
23	31	18	13	0	58,06
24	23	16	6	0	72,73
25	46	30	14	0	65,22
26	40	31	7	2	77,50
27	50	33	17	0	66,00
28	61	47	13	0	77,05
29	43	35	7	0	81,40
30	39	32	6	0	82,05
31	74	61	9	4	82,43
32	89	72	4	3	80,90
33	53	32	17	0	60,38
34	26	14	10	0	53,85
35	47	33	7	0	70,21
36	43	30	8	2	69,77
37	54	43	15	0	74,14
38	31	10	15	0	40,00
39	108	87	16	1	80,56
40	26	21	5	0	80,77
Rata-rata akurasi					68,73

TABEL V
PERBANDINGAN HASIL PERHITUNGAN SEL METODE YANG DIUSULKAN DAN
METODE *K-MEANS*

	Jumlah Benar	Jumlah Over	Jumlah Under
K-Means	1333	382	35
SFCM dan RRM	1486	197	72

Pada hasil uji coba skenario 2, berdasarkan perhitungan manual jumlah sel benar, sel *over-segmentation*, dan sel *under-segmentation* pada citra hasil pemisahan sel kanker bertumpuk, diperoleh hasil bahwa metode usulan mampu memberikan nilai akurasi yang lebih tinggi dibandingkan metode *K-Means*. Keunggulan metode usulan juga terlihat

dari hasil perhitungan jumlah sel yang mengalami *over-segmentation*, yaitu metode usulan mampu mereduksi jumlah sel yang mengalami *over-segmentation*. Namun, metode usulan memiliki kelemahan, yaitu mengakibatkan meningkatnya jumlah sel yang mengalami *under-segmentation*. Ini disebabkan belum maksimalnya penerapan ciri geometri area dalam penentuan *pseudo-blob*.

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil dan analisis hasil uji coba yang telah dijelaskan, dapat disimpulkan bahwa skema segmentasi dan identifikasi sel kanker berdasarkan metode GS yang dilakukan memberikan hasil yang baik. Hal ini ditunjukkan dengan diperolehnya rata-rata nilai MSE sebesar 0,07. Metode yang diusulkan mampu memberikan hasil pemisahan sel bertumpuk yang lebih tinggi nilai akurasinya dibandingkan dengan metode *K-Means*. Hal ini ditunjukkan dengan dihasilkan nilai rata-rata akurasi sebesar 78,41%. Metode ini juga mampu mereduksi jumlah sel yang mengalami *over-segmentation*, tetapi pada beberapa kasus terjadi peningkatan jumlah sel yang mengalami *under-segmentation*, yang diakibatkan oleh proses penggabungan daerah yang memproses semua sel hasil *clustering*. Untuk penelitian selanjutnya, perlu dilakukan pengembangan metode untuk mendeteksi daerah *over-segmentation* atau *pseudo-blob* sebelum proses penggabungan daerah untuk menghindari masalah *under-segmentation*.

REFERENSI

- [1] C. Fatchah and N. Suciati, "Nuclei Segmentation of Microscopic Breast Cancer Image using Gram-Schmidt and Cluster Validation Algorithm," *ICCSCE*, pp. 27–29, November, 2015.
- [2] A. Mouelhi, M. Sayadi, and F. Fnaiech, "A Supervised Segmentation Scheme Based on Multilayer Neural Network and Color Active Contour Model for Breast Cancer Nuclei Detection," *ICEESA*, 2013.
- [3] S. Xie, L. Chen, J. Chen, and H. Nie, "Image segmentation using iterative watershed and ridge detection," *J. Comput. Appl.*, vol. 29, no. 10, pp. 2668–2670, 2009.
- [4] A. Mouelhi, M. Sayadi, and F. Fnaiech, "Automatic segmentation of clustered breast cancer cells using watershed and concave vertex graph," *2011 Int. Conf. Commun. Comput. Control Appl. CCCA 2011*, no. 1, pp. 2–7, 2011.
- [5] A. Mouelhi, M. Sayadi, F. Fnaiech, and S. Member, "Hybrid Segmentation of Breast Cancer Cell Images Using a New Fuzzy Active Contour Model and an Enhanced Watershed Method", *CoDIT*, no. 1, pp. 382–387, 2013.
- [6] N. Aini, C. Fatchah, and B. Amaliah, "Pemisahan Sel Bertumpuk Citra Sel Kanker Payudara Menggunakan Metode Region-Based Active Contour dan Bayesia," *SCAN-Jurnal Teknologi Informasi dan Komunikasi*, pp. 1–7, 2015.
- [7] C. Fatchah, D. Purwitasari, V. Hariadi, and F. Effendy, "Overlapping White Blood Cell Segmentation And Counting on Microscopis Blood Cell Image," *International Journal on Smart Sensing and Intelligent Systems*, vol. 7, no. 3, pp. 1271–1286, 2014.
- [8] P. Phukpattaranont and P. Boonyaphiphat, "Color Based Segmentation of Nuclear Stained Breast Cancer Cell Images," *Communications*, vol. 5, no. 2, pp. 158–164, 2007.
- [9] Y. Li and Y. Shen, "Fuzzy c-means clustering based on spatial neighborhood information for image segmentation," *Journal of Systems Engineering and Electronics*, vol. 21, no. 2, pp. 323–328, 2010.
- [10] S. Z. Beevi, M. M. Sathik, K. Senthamaraikannan, and J. H. J. Yasmin, "A Robust T Fuzzy Clustering Technique With Spatial Neighborhood Information for Effective Noisy Medical Image Segmentation," *ICCCNT*, 2010.
- [11] Y. Chen and J. Chen, "A Watershed Segmentation Algorithm Based on Ridge Detection and Rapid Region Merging", *ICSPCC*, 2014.