

SEGMENTASI CITRA SEL TUNGGAL SMEAR SERVIKS MENGGUNAKAN METODE RADIATING NORMALLY BIASED GENERALIZED GRADIENT VECTOR FLOW SNAKE

Martini Dwi Endah Susanti¹⁾, Handayani Tjandrasa²⁾, dan Chastine Fatichah³⁾

^{1,2,3)}Department of Informatics, Institut Teknologi Sepuluh Nopember,
Kampus ITS, Sukolilo, Surabaya

e-mail: mdesusanti@gmail.com¹⁾, handatj@its.ac.id²⁾, chastine@cs.its.ac.id³⁾

ABSTRAK

Proses segmentasi nukleus dan sitoplasma dibutuhkan dalam sistem diagnosa otomatis pemeriksaan kanker serviks karena dapat mempengaruhi keakuratan sistem. Sebuah algoritma energi eksternal untuk active contour (snake) dengan menggabungkan metode Normally Biased Gradient Vector Flow Snake dan Generalized Gradient Vector Flow Snake diusulkan pada penelitian ini. GGVF dapat memperbesar jangkauan konvergensi kontur aktif, meningkatkan kinerja konvergensi dan lebih handal terhadap noise, serta NBGVF dapat memberikan solusi untuk masalah perlindungan tepi yang tipis. Oleh karena itu, penelitian ini menggabungkan GGVF dan NBGVF untuk mengusulkan versi energi eksternal yang baru yang disebut Normally Biased Generalized Gradient Vector Flow Snake (NBGGVFS). Radiating Edge Map dimanfaatkan sebagai fungsi edge map snake pada penelitian ini. Penelitian ini bertujuan untuk segmentasi sitoplasma dan nukleus citra sel tunggal smear serviks menggunakan metode Radiating Normally Biased Generalized Gradient Vector Flow Snake. Metode usulan terdiri atas tiga proses utama, yaitu pra proses, proses segmentasi awal dan segmentasi akhir. Metode segmentasi awal yang digunakan pada penelitian ini adalah mean shift clustering. Dataset yang digunakan adalah dataset Herlev. Proses uji coba dilakukan dengan cara memvariasikan nilai bandwidth metode segmentasi awal dan dengan membandingkan nilai akurasi dan ZSI metode usulan dengan metode penelitian sebelumnya. Berdasarkan uji coba, nilai rata-rata Zijdenbos Similarity Index (ZSI) dan akurasi untuk segmentasi nukleus adalah 96,96% dan 90,68%, dan untuk sitoplasma adalah 86,78% dan 89,35%. Hasil pengujian menunjukkan metode yang diusulkan mampu mendeteksi area nukleus dan sitoplasma pada citra sel tunggal smear serviks lebih optimal dibandingkan metode sebelumnya sehingga dapat digunakan sebagai metode segmentasi pada sistem identifikasi kanker serviks.

Kata Kunci: Radiating Normally Biased Generalized Gradient Vector Flow Snake; Radiating Edge Map; Segmentasi, Citra sel tunggal smear serviks.

ABSTRACT

Nucleus and cytoplasm segmentation is needed in an automatic diagnosis system for cervical cancer examination because it can affect the system accuracy. An external energy algorithm for active contour by combining Normally Biased Gradient Vector Flow Snake and Generalized Gradient Vector Flow Snake methods is proposed in this study. GGVF can increase the convergence range of active contour and more reliable against noise. NBGVF can provide solutions for weak edge protection problem. Therefore, this study combines that two methods to propose a new external energy version called Normally Biased Generalized Gradient Vector Flow Snake. Radiating Edge Map is used as an edge map snake function in this study. The study aims to segment cytoplasm and nucleus of single cell cervical smear images using Radiating NBGGVFS method. The proposed method consist of three main process, namely preprocessing, initial segmentation and final segmentation. The initial segmentation method used is main shift clustering. The dataset used is Herlev dataset. The experiment process is carried out by varying bandwidth value of initial segmentation and by comparing the accuracy and ZSI of proposed method with the previous research. Based on the experiments, the average value of ZSI and accuracy for nucleus segmentation is 96.96% and 90.68%, and for cytoplasm is 86.78% and 89.35%. The results show that the proposed method is capable of detecting the nucleus and cytoplasm area in single cell cervical smear image more optimal than the previous method so it can be used as a segmentation method in cervical cancer identification system.

Keywords: Radiating Normally Biased Generalized Gradient Vector Flow Snake; Radiating Edge Map; Single cell cervical smear images, Segmentation.

I. PENDAHULUAN

Kanker serviks merupakan jenis kanker kedua yang menyerang wanita dan mengakibatkan kematian. Teknik pemeriksaan bernama Pap tes merupakan sebuah teknik yang sederhana, terjangkau, dan efektif untuk mendiagnosa kanker serviks. Analisis hasil pemeriksaan pap smear secara manual memiliki kelemahan yaitu membutuhkan banyak tenaga ahli di bidang patologi, memakan banyak waktu, dan rawan terhadap kesalahan. Sebuah sistem penyaringan otomatis dan sistem diagnosa yang akurat untuk hasil pemeriksaan pap smear akan sangat bermanfaat untuk mengatasi kelemahan-kelemahan tersebut. Langkah yang paling utama dan paling penting

dari sistem tersebut adalah proses segmentasi sel nukleus dan sitoplasma pada citra hasil pemeriksaan pap smear, karena dapat memengaruhi keakuratan sistem.

Algoritma berbasis kontur aktif atau *snake* merupakan salah satu metode untuk segmentasi citra smear serviks untuk mendapatkan bagian nukleus, sitoplasma dan area latar belakang [1]. Kelemahan *snake* adalah kontur awal harus ditempatkan dekat dengan objek dan konvergensi *snake* sangat lemah terhadap batas kecekungan objek [2]. Untuk mengatasi masalah tersebut *snake* dikembangkan menjadi *Gradient Vector Flow Snake* (GVFS) sebagai pengganti energi eksternal *snake*. Fungsi energi eksternal ini digunakan untuk mengatasi kesulitan dalam mendapatkan batas kecekungan. Namun, GVF memiliki konvergensi yang lemah dan sensitif terhadap inisialisasi *snake* yang disebabkan karena penghalusan yang berlebihan pada area batas objek sehingga belum bisa mendapatkan hasil segmentasi kontur sel nukleus dan sitoplasma dari citra smear serviks secara akurat [14].

Kemudian kombinasi metode *spatial k-means* dan *Radiating Gradient Vector Flow Snake* (RGVFS) diusulkan untuk melakukan segmentasi kontur sitoplasma dan nukleus citra smear serviks [3]. Penelitian tersebut mengusulkan perhitungan *edge map* pada metode GVFS menggunakan *Radiating Edge Map* (REM). Penelitian tersebut menggunakan metode *spatial k-means* untuk segmentasi awal dan menyebabkan beberapa citra tidak dapat tersegmentasi dikarenakan ukuran nukleus yang berukuran sangat kecil dan terdapat banyak *noise*. Metode RGVFS juga digunakan pada penelitian sebelumnya dalam mendeteksi kontur sitoplasma dan nukleus citra smear serviks yang dikombinasikan dengan metode *Fuzzy C Means* (FCM) [4]. Namun metode tersebut ternyata juga belum berhasil untuk mendapatkan kontur sitoplasma dan nukleus secara akurat karena metode perhitungan REM dimanfaatkan pada metode GVFS, dimana GVFS memiliki konvergensi yang lemah dan sensitif terhadap inisialisasi *snake* [13].

Generalized Gradient Vector Flow Snake (GGVFS) diusulkan dengan tujuan untuk meningkatkan konvergensi GVFS [2]. GGVFS didapat dengan mengganti koefisien *smoothing* pada GVFS dengan dua koefisien pembobot pada persamaan GGVFS. Qin et al. (2013) menjelaskan bahwa GGVFS ternyata belum dapat meningkatkan konvergensi batas kecekungan *snake* [5].

Penelitian lain mengusulkan metode baru menggunakan kombinasi metode *Fuzzy C-Means Non Local Spatial* (FCM-NLS) dan *Radiating Component Normalized Generalized Gradient Vector Flow Snake* (RCNGGVFS) [6]. RCNGGVFS memanfaatkan perhitungan REM dalam pencarian *edge map* pada metode CNGGVFS. Metode CNGGVFS memberikan kemampuan yang baik dalam meningkatkan konvergensi batas kecekungan dan tidak sensitif terhadap inisialisasi *snake*. Namun metode tersebut ternyata juga belum berhasil secara optimal pada beberapa citra sel serviks. Citra yang tidak dapat tersegmentasi umumnya memiliki warna sitoplasma yang hampir mirip dengan latar belakang, terdapat banyak *noise*, serta memiliki area nukleus yang berukuran sangat kecil sehingga sering diabaikan dan dianggap sebagai *noise*, sehingga dibutuhkan metode segmentasi awal yang handal dalam menangani ukuran dan bentuk klaster data karena perbedaan ukuran sitoplasma dan nukleus yang kontras, serta yang dapat menangani *noise* pada citra.

Penelitian ini mengusulkan metode baru untuk segmentasi citra sel tunggal smear serviks menggunakan kombinasi metode *mean shift clustering* sebagai metode segmentasi awal dan metode *Radiating Normally Biased Generalized Gradient Vector Flow Snake*. Metode ini memanfaatkan REM untuk menghitung *edge map* dari citra dengan menggabungkan metode GGVFS dan *Normally Biased Gradient Vector Flow Snake* (NBGVFS) [7]. Model GGVFS memperbesar jangkauan konvergensi kontur aktif, meningkatkan kinerja konvergensi dan lebih handal terhadap *noise*. NBGVF memberikan solusi untuk masalah perlindungan tepi yang tipis. Oleh karena itu, metode ini menggabungkan GGVF dan NBGVF untuk mengusulkan versi energi eksternal yang baru. Metode *mean shift Clustering* digunakan sebagai metode segmentasi awal yang akan menghasilkan inisial kontur. *Mean shift Clustering* merupakan metode *clustering* berbasis densitas dan *nonparametric* yang dapat menangani bentuk dan ukuran klaster data. Metode usulan terdiri dari 3 proses utama, yaitu tahap praproses, citra smear serviks RGB asli dikonversi ke dalam ruang warna CIELAB dan layer L* dinormalisasi untuk mendapatkan citra keabuan dan dilakukan penghilangan *noise* dengan metode *median filtering*. Kemudian citra smear serviks dibagi ke dalam tiga bagian (nukleus, sitoplasma dan area latar belakang) menggunakan *mean shift clustering*. Selanjutnya tahap terakhir adalah segmentasi kontur. Pada tahap ini diusulkan metode RNBGGVFS sebagai fungsi pengganti energi eksternal *snake* untuk mendapatkan kontur nukleus dan sitoplasma citra sel tunggal smear serviks.

Uji coba dilakukan dengan menggunakan dataset Herlev [12], yang terdiri atas 917 citra sel tunggal smear serviks. Gambar 2 dan Gambar 3 merupakan contoh citra sel tunggal smear serviks dan *groundtruth* dataset Herlev. Uji coba dilakukan dengan membandingkan nilai rata-rata akurasi dan *Zijdenbos similarity index* (ZSI) dari metode usulan dengan metode penelitian sebelumnya [11]. Metode yang diusulkan diharapkan dapat memberikan hasil segmentasi citra sel tunggal smear serviks yang akurat. Makalah ini dibagi menjadi 3 bagian: bagian II sampai IV. Metode usulan dijelaskan pada bagian II, hasil percobaan dan pembahasan dijelaskan pada bagian III, serta kesimpulan dijelaskan pada bagian IV.

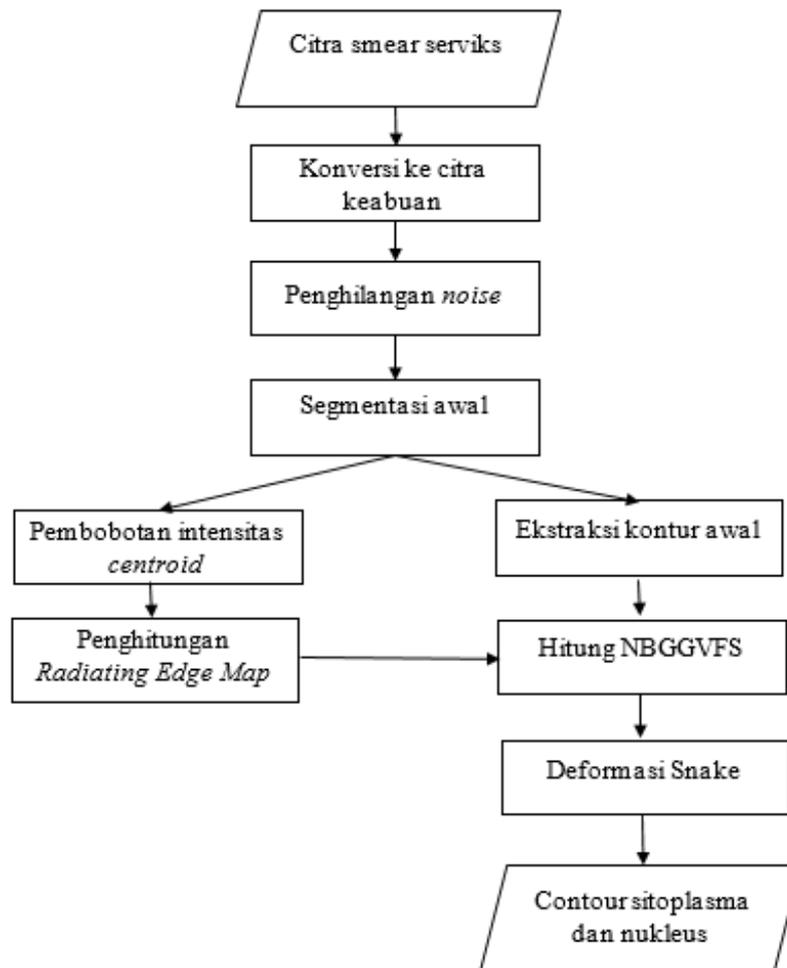
II. METODE PENELITIAN

Gambar 1 menunjukkan proses segmentasi citra sel tunggal smear serviks secara umum. Segmentasi citra sel tunggal smear serviks terdiri atas tiga tahapan utama yaitu tahap pra proses, tahap segmentasi awal dan tahap segmentasi kontur.

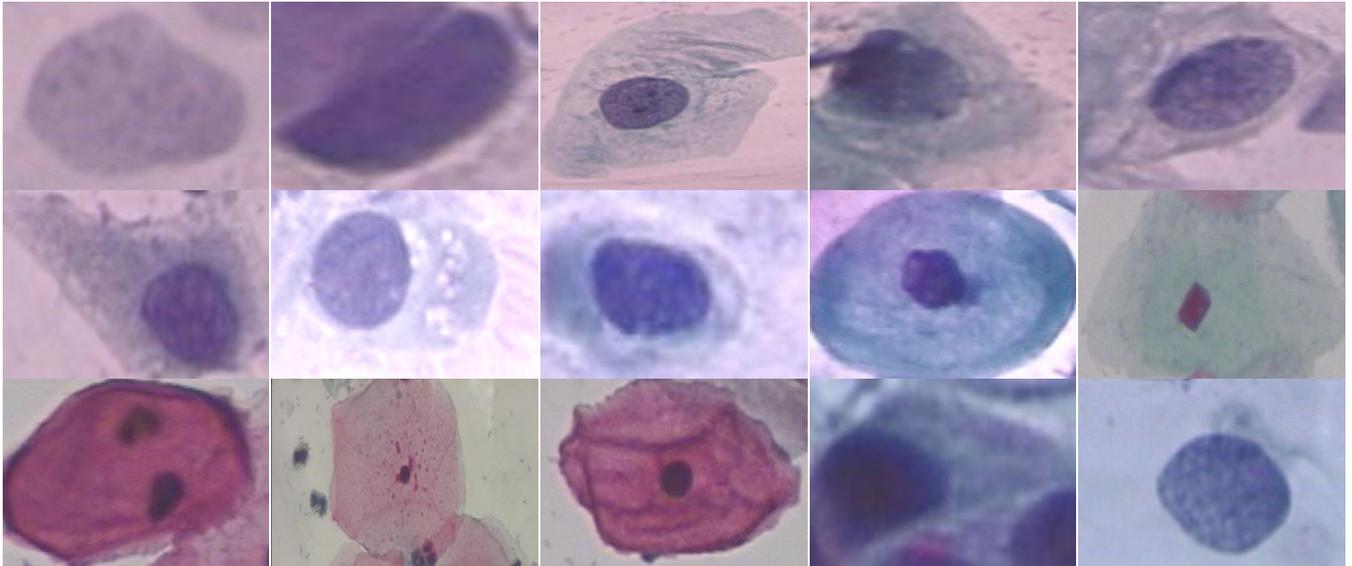
A. Pra proses

Dalam citra smear serviks, terdapat kontras yang rendah, pewarnaan yang tidak seragam serta *noise* yang seringkali akan menghambat proses segmentasi. Diperlukan adanya perbaikan citra untuk meningkatkan akurasi proses segmentasi citra. Citra *smear* serviks (RGB) dikonversi kedalam ruang warna CIELAB. Ruang warna ini terdiri dari tiga dimensi, yaitu L^* dimensi pencahayaan, a^* dan b^* dimensi warna. CIELAB berhasil memisahkan kekontrasan dari warnanya. Setelah kekontrasan dan warna penyusun citra telah terpisah, selanjutnya adalah mengekstraksi dimensi L^* dan menormalisasi ke rentang $[0,255]$ untuk membentuk citra keabuan [9].

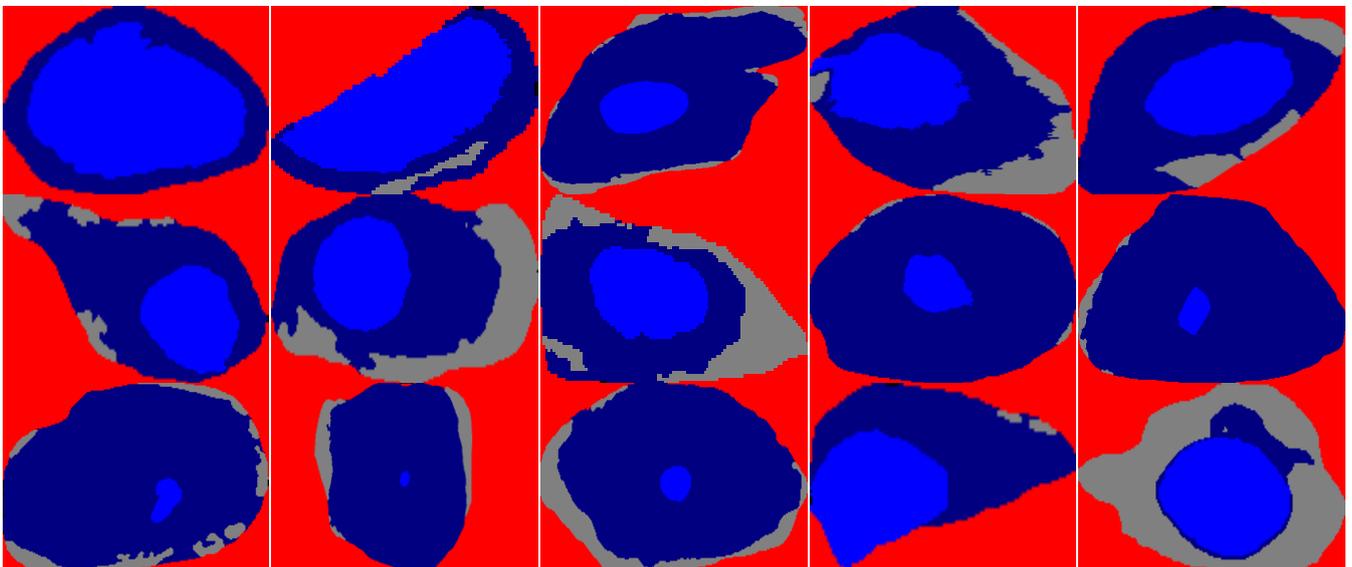
Kemudian yang perlu dilakukan selanjutnya adalah menghilangkan *noise* yang dapat menghambat proses segmentasi dengan metode *median filtering*. Metode tersebut dikenal efektif dalam melakukan filtering penghilangan *noise Gaussian* dan *impulse* pada citra *smear* serviks. Gambar 4 merupakan contoh citra keluaran sistem tahap praproses.



Gambar 1 Diagram alir proses segmentasi citra smear serviks.



Gambar 2 Contoh citra sel tunggal smear serviks dataset Herlev [12].



Gambar 3 Contoh *Ground truth* citra sel tunggal smear serviks dataset Herlev [12].

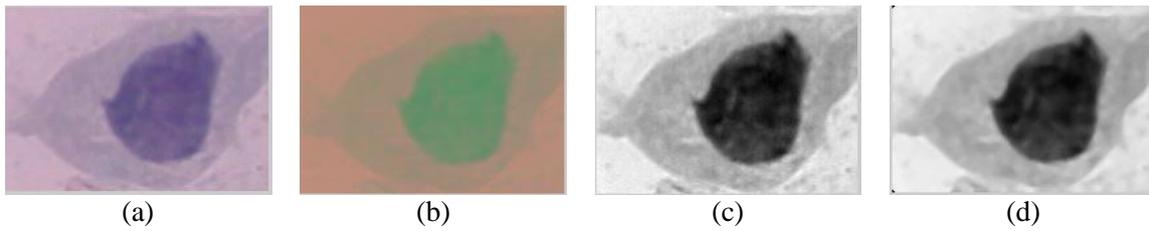
B. Segmentasi awal

Pada citra yang telah dilakukan tahap praproses, dilakukan segmentasi awal yang dibagi menjadi tiga tahap, yaitu

a) Segmentasi *mean shift clustering*

Citra hasil praproses disegmentasi menggunakan metode *mean shift clustering* untuk memisahkan citra tersebut menjadi 3 bagian, yaitu nukleus, sitoplasma dan area latar belakang. Hasil dari segmentasi awal selanjutnya akan digunakan sebagai *initial contour* tahap segmentasi akhir. Pertama-tama menentukan parameter *bandwidth* dan *threshold* konvergensi. Selanjutnya memilih titik secara acak sebagai titik inisial, kemudian menetapkan nilai titik tersebut sebagai *mean* awal. Setelah itu melakukan penghitungan jarak antar piksel dan membandingkan dengan *bandwidth* (radius). Titik dengan jarak yang lebih kecil dari *bandwidth* selanjutnya digabung menjadi klaster yang sama dan dihitung nilai *mean*[8]. Atau tahapan lebih jelasnya sebagai berikut.

1. Pilih titik data sebagai inialisasi.
2. Letakkan kernel/*window* dengan radius tertentu di sekitar titik yang diinginkan.
3. Hitung nilai *mean*.
4. Pindahkan pusat kernel ke titik tengah.
5. Ulangi 3 & 4 hingga konvergen. Setiap iterasi akan bergerak "menanjak" pada gradien densitas distribusi data hingga mencapai puncak bukit (maksimum lokal).



Gambar 4 Citra asli dataset Herlev [12] (a), citra dalam ruang warna CIELAB (b), citra *grayscale* (c), dan citra hasil penghilangan *noise* (d).

6. Ulangi 1-5 untuk semua titik data. Titik yang menyatu dengan *local maxima* yang sama adalah anggota dari kelompok yang sama. Jumlah kluster adalah jumlah *local maxima*.

b) *Ekstraksi Kontur Awal*

Setelah didapatkan citra yang tersegmentasi, berikutnya dilakukan ekstraksi kontur awal dengan asumsi bahwa area dengan kecekungan terlalu disebut sitoplasma. Kandidat nukleus yang dapat memiliki beberapa kemungkinan diasumsikan dengan memastikan bahwa nukleus merupakan area yang terdapat di dalam sitoplasma dan kandidat area dengan nilai ambang batas kurang dari nilai ambang batas $\beta\%$ dari citra dianggap sebagai *noise* dan akan diabaikan [3]. Karena ukuran nukleus yang berbeda pada tiap citra, nilai ambang batas yang digunakan adalah 0,5. Berdasarkan statistik uji seperti pada [14], nilai tersebut dianggap dapat menghilangkan area *noise*. Pada penelitian sebelumnya juga digunakan nilai ambang batas yang sama yaitu 0,5. Gambar 5 merupakan contoh citra keluaran tahap segmentasi awal.

c) *Pelabelan centroid dengan bobot intensitas*

Kemudian dilakukan pelabelan terhadap centroid dengan bobot intensitas, proses ini penting untuk membentuk *Radiating Edge Map* pada tahap segmentasi. Diasumsikan area nukleus terdiri dari k titik $(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)$, dan $I(x, y)$ adalah intensitas titik (x, y) . Maka bobot intensitas centroid (x_c, y_c) dapat dilihat pada persamaan (1).

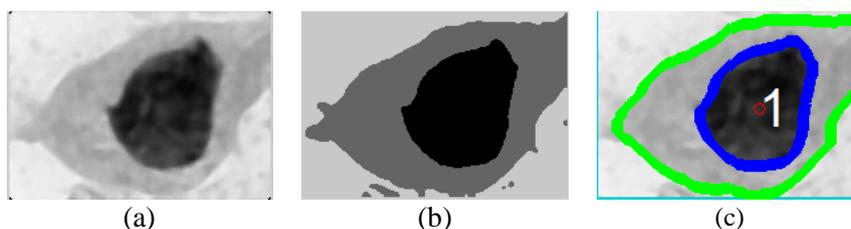
$$\begin{cases} x_c = \frac{\sum_{i=1}^k x_i(255 - I(x_i, y_i))}{\sum_{i=1}^k (255 - I(x_i, y_i))} \\ y_c = \frac{\sum_{i=1}^k y_i(255 - I(x_i, y_i))}{\sum_{i=1}^k (255 - I(x_i, y_i))} \end{cases} \quad (1)$$

C. *Segmentasi Kontur*

Tahap segmentasi kontur menggunakan input citra hasil proses segmentasi awal dan terdiri dari beberapa tahap yaitu perhitungan REM, Perhitungan NBGGVFS, dan deformasi *snake*.

i. *Perhitungan REM*

Edge map (∇f) yang dihitung dengan menggunakan operator deteksi tepi seperti *sobel* tidak dapat menggunakan informasi direksional sepenuhnya. Metode tersebut tidak dapat mendeteksi daerah nukleus dengan benar dikarenakan umumnya nukleus berwarna lebih gelap dibanding sitoplasma, dan sitoplasma berwarna lebih gelap dibanding *background*. Penelitian ini memanfaatkan REM untuk menghitung *edge map* pada metode NBGGVFS. REM merupakan proses penghitungan *gradient edge map* pada sepanjang garis pancaran yang dimulai dari *intensity weighted area centroid* nukleus hasil segmentasi awal. *Radiating Edge Map* (REM) dapat dihitung dengan mengekstraksi setiap *radiating line* dan *radiating difference* sepanjang garis. Tahap-tahap perhitungan *Radiating Edge Map* (REM) dapat dijabarkan seperti berikut [3]:



Gambar 5 Citra hasil preproses (a), citra hasil segmentasi *mean shift clustering* (b), dan hasil ekstraksi inisial kontur-pelabelan centroid (c).

1. Tahap pertama pada perhitungan REM adalah mengekstraksi setiap *Radiating line* menggunakan persamaan $len_b = \lfloor \sqrt{(x_b - x_c)^2 + (y_b - y_c)^2} \rfloor$. (2)

2. Kemudian, menghitung *radiating difference* sepanjang garis menggunakan persamaan $rd(x_i, y_i) = I_0(x_{i-1}, y_{i-1}) - I_0(x_i, y_i)$. (3)

Radiating gradient $rg(x_i, y_i)$ pada titik (x_i, y_i) dapat dihitung dengan $rg(x_i, y_i) = \frac{|rd(x_i, y_i)| + |rd(x_{i+1}, y_{i+1})|}{2}$. (4)

3. *Radiating difference* yang bernilai positif harus diperbaiki, karena intensitas area centroid memiliki nilai kecil dan akan semakin besar jika semakin jauh dari centroid di sepanjang *radiating line*. *Radiating difference* diperbaiki menggunakan persamaan

$$rg(x_i, y_i) = \frac{|Fs(rd(x_i, y_i))| + |Fs(rd(x_{i+1}, y_{i+1}))|}{2} \quad (5)$$

ii. *Perhitungan GGVFS*

GVFS memiliki konvergensi yang lemah dan sensitif terhadap inisialisasi *snake*. Kelemahan ini disebabkan karena *smoothing* yang berlebihan pada area *boundary*, yang diatur oleh koefisien μ pada [5]. Untuk mengatasi masalah ini, ditambahkan dua koefisien pembobot pada persamaan gaya eksternal GVFS [5]. Persamaan gaya eksternal baru tersebut dinamakan *Generalized Gradient Vector Flow Snake* (GGVFS). Fungsi turunan energi dari GGVFS seperti persamaan berikut

$$v_t(x, y, t) = \underbrace{g(|\nabla f|) \nabla^2 v(x, y, t)}_{smoothness\ energy} + \underbrace{h(|\nabla f|) |v(x, y, t) - \nabla f|}_{edge\ energy} \quad (6)$$

dimana

$$g(|\nabla f|) = e^{-|\nabla f|/k}, \quad (7)$$

$$h(|\nabla f|) = 1 - g(|\nabla f|), \quad (8)$$

term pertama pada bagian kanan dari (6) merupakan term *smoothing*, yang akan menghasilkan vektor. Term kedua adalah term data atau disebut juga *edge energy*, yang akan membawa vektor v dekat dengan *gradient edge map* (∇f) dari citra. Parameter k menentukan nilai koefisien pembobot untuk term *smoothing* dan data. Nilai dari k berhubungan dengan kekuatan *noise* dari sebuah citra. Semakin tinggi level *noise*, maka semakin banyak nilai k . Pada persamaan (6), $|\nabla f|$ dapat dihitung menggunakan operator gradien apapun. Jika term data pada persamaan didefinisikan vector v dekat dengan *gradient edge map* (∇f) dari citra, maka koefisien pembobot $h(|\nabla f|)$ harus bertambah. Berdasarkan persamaan, parameter k harus > 0 .

iii. *Perhitungan NBGGVFS*

Model GGVFS memperbesar jangkauan konvergensi kontur aktif, meningkatkan kinerja konvergensi dan lebih handal terhadap *noise*. NBGVF memberikan solusi untuk masalah perlindungan *edge* yang tipis [10]. Oleh karena itu, metode ini menggabungkan GGVF dan NBGVF untuk mengusulkan versi energi eksternal yang baru, yang didefinisikan sebagai berikut:

$$E(V) = \iint g(x, y) (gs(x, y)V_{NN} + hs(x, y)V_{TT}) dx dy + h(x, y) (v - \nabla f) dx dy, \quad (9)$$

$$g(|\nabla f|) = e^{-|\nabla f|/K}, \quad (10)$$

$$h(|\nabla f|) = 1 - e^{-\frac{|\nabla f|}{K}}, \quad (11)$$

$$hs(f) = \begin{cases} 1 & (|e| \geq \tau) \\ \frac{-f^3}{8\tau^3} + \frac{5f}{8\tau} + \frac{1}{2} & (0 < |e| < \tau) \\ 0 & (|e| = 0) \end{cases} \quad (12)$$

$$gs(f) = 1 - hs(f), \quad (13)$$

dimana V_{NN} dan V_{TT} merupakan turunan kedua sepanjang arah normal dan tangen. Sedangkan $g(|\nabla f|)$ dan $h(|\nabla f|)$ merupakan koefisien *smooth* dan *data terms*. Seperti yang didefinisikan pada GGVF, nilai K bertambah dengan intensitas *noise* gambar, tetapi hal ini dapat menyebabkan tepi yang tipis menjadi terlalu halus. Parameter hs dan gs merupakan derajat difusi gambar pada arah normal dan tangen. Tidak seperti koefisien operator difusi normal dan tangen pada NBGVF, kedua koefisien secara langsung bergantung pada intensitas dibanding dengan gradien *edge*, sehingga mengurangi kompleksitas komputasi. Sehingga, persamaan evolusi energi eksternal dapat dituliskan sebagai berikut:

$$v(x, y, t) = g(|\nabla f|) (gs(f)V_{NN}(x, y, t) + hs(f)V_{TT}(x, y, t)) - h(|\nabla f|) [V(x, y, t) - \nabla f]. \quad (14)$$

iv. *Deformasi Snake*

Active contour atau *snake* dapat dihitung dengan meminimasi fungsi energi eksternal dan internal yang dihitung dari data citra. Energi eksternal *snake* telah dihitung dengan menggunakan metode NBGGVFS [10]. Gambar 6 menunjukkan contoh citra hasil keluaran tahap segmentasi kontur. Gambar 7 merupakan perbandingan hasil segmentasi metode usulan dan *ground truth*. Untuk energi internal *snake* dihitung dengan langkah-langkah sebagai berikut:

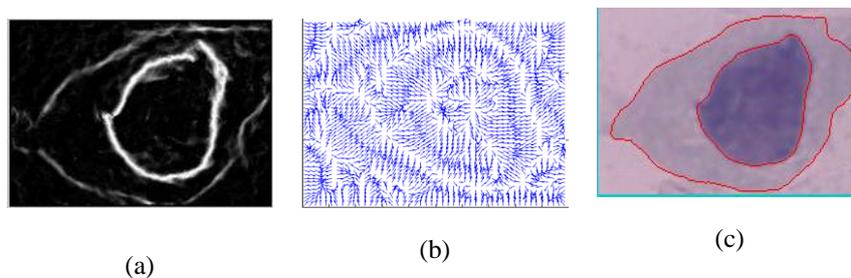
1. Tentukan kontur awal yang telah didapat dari proses segmentasi awal.
2. Tentukan parameter *snake* yang dibutuhkan seperti *alpha* (parameter elastisitas), *beta* (parameter ketegaran), *gamma* (parameter kerekatan) dan *kappa* (bobot tekanan eksternal).
3. Tiap titik *snake* akan bergerak secara berulang mencari lokasi yang terdapat energi *snake* paling minimum.

III. HASIL DAN PEMBAHASAN

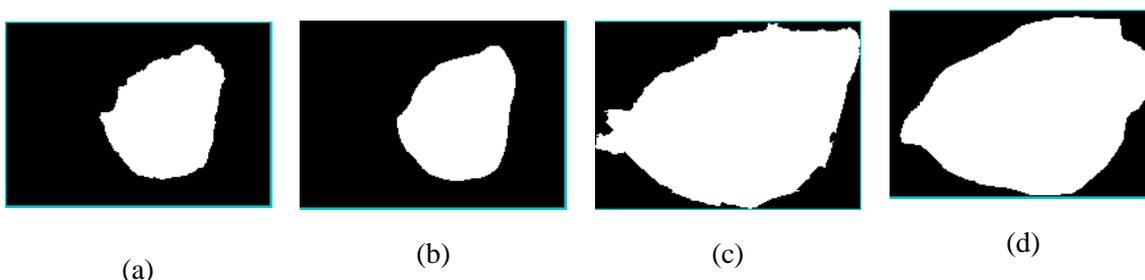
Tahapan penelitian berikutnya setelah tahap implementasi adalah tahap evaluasi dan uji coba metode. Data masukan pada uji coba ini berupa citra mikroskopik sel tunggal smear serviks. Data yang digunakan diperoleh dari Departemen Patologi Rumah Sakit Universitas Teknik Denmark. Dataset Herlev ini terdiri dari 7 kelas yang berbeda dengan jumlah keseluruhan 917 citra sel tunggal smear serviks [12]. Nama-nama kelas pada citra sel tunggal *smear* serviks adalah *Carcinoma in situ*, *Light Dysplastic*, *Moderate Dysplastic*, *Normal Columnar*, *Normal Intermediate*, *Normal Superficial*, dan *Severe Dysplastic*. Setiap satu citra sel tunggal smear serviks memiliki ukuran yang berbeda-beda, berada dalam ruang warna RGB dan berformat BMP. Adapun uji coba dilakukan dengan dua skenario yaitu:

1. Membandingkan hasil segmentasi kontur citra sel tunggal smear serviks dengan menggunakan kombinasi metode *mean shift clustering* dan *Radiating Normally Biased Generalized Gradient Vector Flow Snake* dengan beberapa variasi parameter *bandwidth* pada metode segmentasi awal yaitu 0,2 , 0,25 , dan 0,3.
2. Nilai *bandwidth* yang didapat pada skenario 1 digunakan untuk membandingkan hasil segmentasi metode usulan dengan beberapa metode pada penelitian sebelumnya yaitu *spatial k-means* dan *RGVFS*, *FCM* dan *RGVFS*, serta *FCM_NLS* dan *RCNNGVFS*.

Evaluasi metode dilakukan dengan cara membandingkan citra hasil segmentasi dengan citra hasil segmentasi manual. Evaluasi dilakukan dengan dua cara, diantaranya mencari nilai *Zijdenbos Similarity Index (ZSI)* [11], dan menggunakan *confusion matrix* seperti pada Tabel 1. *ZSI* merupakan metode pengukuran untuk mengetahui



Gambar 6 Radiating Edge Map (a), energi eksternal NBGGVF (b), dan citra hasil segmentasi kontur (c).



Gambar 7 Nukleus *ground truth* (a), nukleus hasil segmentasi metode usulan (b), sitoplasma *ground truth* (c), dan sitoplasma hasil segmentasi metode usulan (d).

TABEL 1
CONFUSION MATRIX

		<i>Hasil Segmentasi</i>	
		<i>Negatif</i>	<i>Positif</i>
<i>Ground Truth</i>	<i>Negatif</i>	TN	FP
	<i>Positif</i>	FN	TP

performa metode segmentasi berbasis region. ZSI digunakan untuk menghitung seberapa besar rasio overlap diantara dua citra. ZSI didefinisikan sebagai berikut.

$$ZSI = 2 \frac{|A_M \cap A_G|}{|A_M| + |A_G|} \tag{15}$$

dimana A_M merupakan citra hasil segmentasi menggunakan metode yang diusulkan, sedangkan A_G merupakan citra segmentasi manual. *Confusion matrix* digunakan untuk mencari nilai akurasi dari hasil segmentasi.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{16}$$

Dengan menggunakan *true positive* (TP), *true negative* (TN), *false negative* (FN), dan *false positive* (FP), nilai akurasi segmentasi bisa didapatkan. Nilai akurasi segmentasi mendefinisikan seberapa besar yang terklasifikasi dengan benar secara keseluruhan.

a. Hasil ujicoba skenario 1

Uji coba skenario pertama dilakukan segmentasi dengan metode usulan yaitu menggunakan metode fungsi pengganti gaya eksternal RNBGGVFS yang memanfaatkan perhitungan REM pada pencarian *edge map* pada metode NBBGVFS dengan metode segmentasi awal *mean shift clustering* (RNBGGVFS & *mean shift*). Uji coba skenario 1 ini membandingkan hasil segmentasi dengan menggunakan variasi nilai parameter pada segmentasi awal, yaitu metode *mean shift clustering*. Parameter yang digunakan adalah parameter *bandwidth* atau radius. Parameter *bandwidth* merupakan parameter jarak antar titik dalam sebuah kernel yang menentukan suatu titik merupakan bagian dari suatu kluster atau tidak. Pada *mean shift clustering*, perhitungan jarak antar piksel dibandingkan dengan *bandwidth* (radius), dimana titik dengan jarak yang lebih kecil dari *bandwidth* selanjutnya digabung menjadi kluster yang sama dan dihitung *mean*nya.

Pada penelitian ini tidak semua citra pada dataset berhasil dianalisis. Ada beberapa citra yang gagal disegmentasi dikarenakan beberapa hal. Citra pada dataset Herlev memiliki karakteristik yang berbeda, masing-masing kelas memiliki ukuran nukleus yang berbeda pula. Banyak *noise* yang terdapat pada citra serta kontras yang rendah juga memengaruhi hasil segmentasi. Beberapa citra juga ditemukan memiliki jumlah sel lebih dari satu serta ada pula citra dengan sel yang overlap. Hal tersebut juga memengaruhi keberhasilan proses segmentasi. Pada percobaan skenario pertama, dilakukan uji coba dengan menggunakan variasi nilai *bandwidth*. Nilai yang diuji coba yaitu 0,2 , 0,25 , dan 0,3. Gambar 8 dan Tabel 2 merupakan hasil uji coba segmentasi pada beberapa nilai *bandwidth*. Hasil tersebut meliputi nilai rata-rata akurasi dan ZSI pada nukleus dan sitoplasma. Nilai akurasi didapatkan dengan membandingkan jumlah luas area putih dan area hitam pada citra *groundtruth* dan citra hasil segmentasi sistem. Nilai index ZSI merupakan seberapa besar area yang *overlap* antara citra hasil segmentasi dan *groundtruth*. Berdasarkan uji coba, diketahui bahwa nilai rata-rata akurasi dan ZSI cenderung lebih tinggi pada saat *bandwidth* atau radius bernilai 0,2 dibandingkan pada saat bernilai 0,25 dan 0,3 yaitu 96,3% dan 89,2% pada segmentasi nukleus serta 83,6% dan 87,2% untuk segmentasi sitoplasma.

b. Hasil uji coba skenario 2

Pada penelitian sebelumnya, citra yang dapat dianalisis cenderung lebih sedikit dibanding dengan metode usulan. Pada metode Radiating Gradient Vector Flow Snake (RGVFS) dan Fuzzy C-Means (FCM) [4] serta RGVFS dan *spatial k-means* [3], sedikitnya citra yang berhasil dianalisis dikarenakan salah satu penyebabnya adalah metode *denoising* yang digunakan belum handal, sementara citra yang ada memiliki banyak *noise*. Hal tersebut berdampak pada proses segmentasi awal citra sel tunggal smear serviks. Kelas dengan jumlah citra paling sedikit yang berhasil disegmentasi adalah kelas *normal superficiel*. Kelas tersebut memiliki karakteristik ukuran nukleus yang relatif kecil. Nukleus tersebut tidak berhasil disegmentasi pada proses segmentasi awal, sehingga proses pencarian kontur tidak dapat dilanjutkan ke proses selanjutnya yaitu segmentasi akhir. Kemudian pada metode *Radiating Component Normalized Generalized Gradient Vector Flow Snake* (RCNBBGVFS) dan *Fuzzy C-Means_Non Local Spatial* (FCM-NLS) [6] jumlah citra yang dapat dianalisis meningkat, dikarenakan proses segmentasi awal yang memiliki metode penghilangan *noise* yang cukup handal didalamnya. Pada metode usulan, digunakan metode segmentasi awal yaitu *mean shift clustering*. Metode ini diketahui dapat menangani ukuran serta bentuk kluster data. Metode ini juga tidak sensitif terhadap *noise*, sehingga hasilnya adalah lebih banyak citra yang berhasil disegmentasi.

Pada metode usulan, digunakan kombinasi metode RNBGGVFS dan *mean shift clustering* sebagai segmentasi awal. Metode *mean shift* merupakan metode *clustering* berbasis densitas dan *nonparametric* yang mampu menangani ukuran serta bentuk data. Dengan metode ini diharapkan akan mampu meningkatkan jumlah citra yang dapat disegmentasi. Uji coba skenario kedua dilakukan dengan membandingkan metode usulan yaitu kombinasi *mean shift clustering* dan RNBGGVFS dengan metode penelitian sebelumnya. Metode tersebut diantaranya kombinasi *spatial k-means* dengan *Radiating Gradient Vector Flow Snake* (RGVFS) [3], kombinasi FCM dan RGVFS [4], serta kombinasi FCM_NLS dan RCNNGVFS [6]. Pada hasil menggunakan metode usulan terlihat bahwa kontur yang dihasilkan semakin mendekati kontur pada citra asli. Sementara pada metode lain masih terlihat bahwa ada area bukan nukleus yang dideteksi sebagai area nukleus, begitu juga pada area sitoplasma. Hasil evaluasi skenario 2 dapat dilihat pada Tabel 3. Hasil segmentasi metode usulan dan metode penelitian sebelumnya dapat dilihat pada Gambar 9.

c. Analisis dan Pembahasan

Dataset pada penelitian ini adalah citra sel tunggal smear serviks yang diperoleh dari dataset Herlev. Dataset Herlev terdiri dari 917 citra sel tunggal smear serviks yang diperoleh dari Departemen Patologi Rumah Sakit Universitas Herlev dan Departemen Automasi Universitas Teknik Denmark. Dataset citra tersebut dikumpulkan oleh ahli *cytology* menggunakan mikroskop yang terhubung ke kamera digital. Ahli *cytology* mengklasifikasikan dataset citra sel tunggal smear serviks menjadi 7 kelas secara manual.

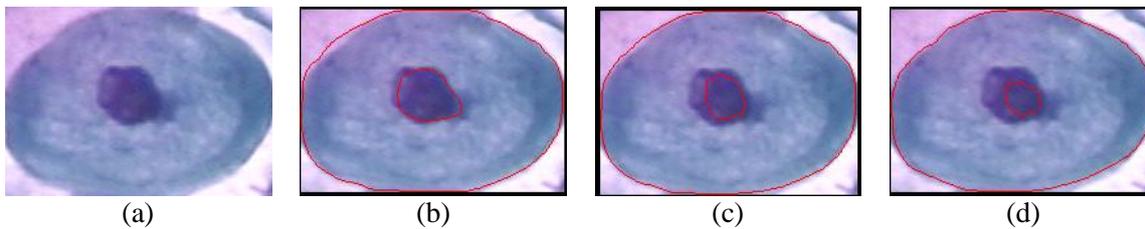
Kelas tersebut terdiri dari kelas *Carcinoma in situ*, *Light Dysplastic*, *Moderate Dysplastic*, *Normal Columnar*, *Normal Intermediate*, *Normal Superficial*, dan *Severe Dysplastic*. Dataset Herlev juga terdapat *groundtruth* atau hasil segmentasi manual. *Groundtruth* citra sel tunggal smear serviks didapatkan dengan menggunakan *software* CHAMP dan tersegmentasi menjadi 3 bagian, yaitu sitoplasma, nukleus, dan *background* [12].

Dari semua kelas citra sel tunggal smear serviks yang berjumlah 917 citra, tidak semua citra berhasil dianalisis. Ada beberapa citra yang gagal disegmentasi dikarenakan beberapa hal. Citra pada dataset Herlev memiliki karakteristik yang berbeda, masing-masing kelas memiliki ukuran nukleus yang berbeda pula. Banyak *noise* yang terdapat pada citra serta kontras yang rendah juga memengaruhi hasil segmentasi. Beberapa citra juga ditemukan memiliki jumlah sel lebih dari satu serta ada pula citra dengan sel yang *overlap*. Hal tersebut juga memengaruhi keberhasilan proses segmentasi.

Metode segmentasi awal yang digunakan pada penelitian ini adalah *mean shift clustering*. Segmentasi awal berfungsi untuk mendapatkan kontur awal yang akan digunakan sebagai data masukan pada proses segmentasi kontur menggunakan metode *Radiating Normally Biased Generalized Gradient Vector Flow Snake* (RNBGGVFS). Metode *mean shift clustering* merupakan metode *clustering* berbasis densitas dan *nonparametric* yang dapat menangani ukuran dan bentuk kluster data, karena pada beberapa karakteristik citra dataset memiliki perbedaan ukuran data yang sangat mencolok, contohnya pada kelas *normal superficial* yang memiliki ukuran nukleus yang sangat kecil yang tidak berhasil dideteksi dengan metode segmentasi awal dari penelitian sebelumnya. Hal ini bermanfaat pada proses segmentasi awal untuk meningkatkan jumlah citra yang berhasil dianalisis dan dapat dilanjutkan ke tahap segmentasi berikutnya.

Edge map yang dihitung dengan menggunakan operator deteksi tepi seperti *sobel*, belum dapat mendeteksi daerah nukleus dari citra smear serviks dengan benar [3]. Hal ini disebabkan karena area nukleus umumnya lebih gelap dari pada area sitoplasma dan area sitoplasma lebih gelap dibanding dengan area *background*. Penelitian ini mengusulkan metode RNBGGVFS untuk segmentasi kontur citra sel tunggal smear serviks. Metode RNBGGVFS merupakan metode *active contour* yang menggunakan NBGGVFS sebagai fungsi energi eksternal dan memanfaatkan perhitungan *radiating edge map* (REM) untuk pencarian *edge map*-nya. REM merupakan proses perhitungan *gradient edge map* pada sepanjang garis pancaran atau *radiating line* yang dimulai dari *intensity-weighted area centroid* nukleus hasil segmentasi awal.

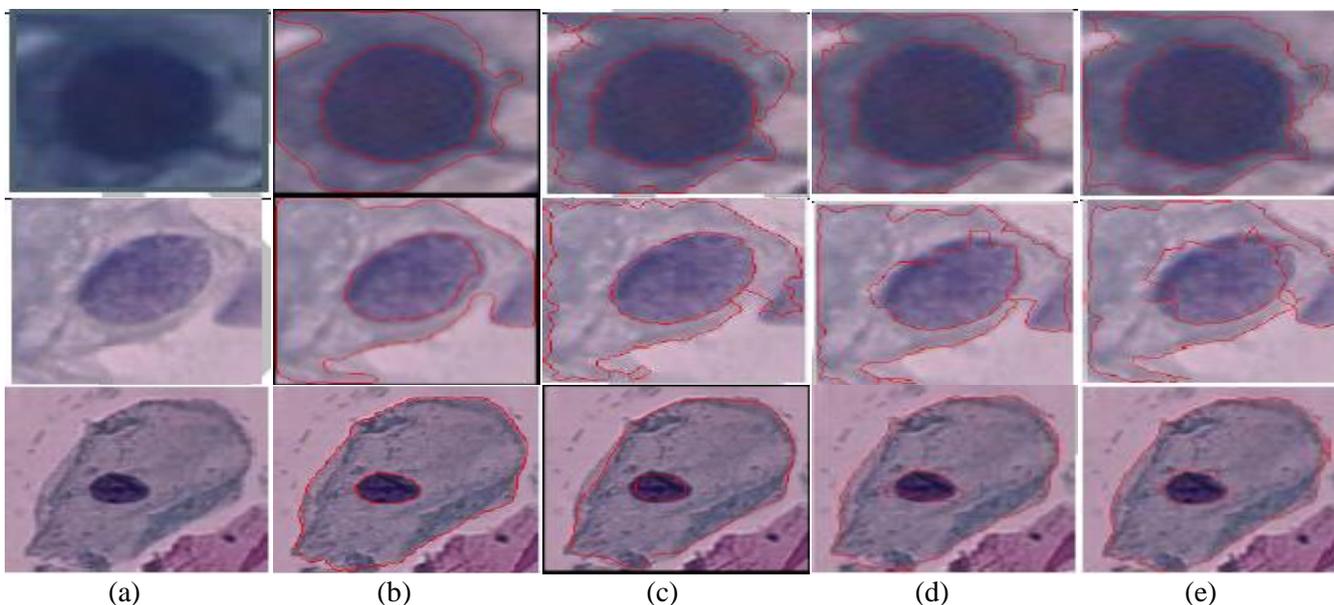
Metode perhitungan REM yang dimanfaatkan pada metode GVFS belum berhasil untuk mendapatkan kontur sitoplasma dan nukleus secara akurat. Hal ini disebabkan karena GVFS memiliki konvergensi yang lemah dan sensitif terhadap inisialisasi *snake* [3]. Penelitian ini menggunakan metode RNBGGVFS, yaitu metode yang memanfaatkan perhitungan REM dalam pencarian *edge map* pada kombinasi metode *Generalized Gradient Vector Flow Snake* (GGVFS) dan *Normally Biased Gradient Vector Flow Snake* (NBGVFS). Model GGVFS memperbesar jangkauan konvergensi kontur aktif, meningkatkan kinerja konvergensi dan lebih handal terhadap *noise*. Berdasarkan metode NGVF, yang memiliki efisiensi difusi yang lebih tinggi, NBGVF memberikan solusi untuk masalah perlindungan tepi yang tipis. Oleh karena itu, metode ini menggabungkan GGVF dan NBGVF untuk mengusulkan versi energi eksternal yang baru. Sehingga, dengan menggunakan perhitungan REM sebagai



Gambar 8 Citra asli (a), Hasil uji coba dengan bandwidth = 0,2 (b) , 0,3 (c), dan 0,4 (d).

TABEL 2
HASIL UJI COBA SKENARIO 1

Bandwidth	Nukleus		Sitoplasma	
	Akurasi	ZSI	Akurasi	ZSI
0,2	0,963	0,892	0,836	0,872
0,25	0,962	0,891	0,831	0,871
0,3	0,959	0,891	0,863	0,872



Gambar 9 Citra asli (a), hasil segmentasi metode RNBGGVFS & mean shift clustering (metode usulan) (b), hasil segmentasi metode RCNGGVFS & FCM_NLS (c), hasil segmentasi metode RGVFS & FCM (d), dan hasil segmentasi metode RGVFS & spatial k-means (e).

pengganti perhitungan konvensional *edge map* pada metode NBBGVFS dapat memberikan hasil segmentasi kontur yang akurat pada citra sel tunggal smear serviks.

Uji coba dilakukan untuk membuktikan performa kombinasi metode RNBGGVFS dan *mean shift clustering* dibanding dengan metode pada penelitian sebelumnya, yaitu RGVFS dan *spatial k-means* [3], RGVFS dan FCM [4] dan RCNGGVFS dan FCM_NLS [6]. Tabel 3 menunjukkan bahwa metode usulan memiliki rata-rata akurasi dan ZSI pada semua kelas citra sel tunggal smear serviks yang lebih tinggi dibandingkan dengan metode pada penelitian sebelumnya. Rata-rata akurasi dan ZSI nukleus pada metode usulan dihasilkan sebesar 96,96% dan 90,68% sedangkan pada sitoplasma dihasilkan 86,78% dan 89,35%. Metode RCNGGVFS dan FCM_NLS dihasilkan rata-rata akurasi dan ZSI nukleus sebesar 95,34% dan 88,06% sedangkan pada sitoplasma dihasilkan 83,48% dan 87,14%. Kemudian pada metode RGVFS & FCM dihasilkan rata-rata akurasi dan ZSI nukleus sebesar 94,88% dan 85,69%, sedangkan pada sitoplasma dihasilkan 81,83% dan 86%. Metode RGVFS & *spatial k-means* dihasilkan rata-rata akurasi dan ZSI nukleus sebesar 94,49% dan 86,58% sedangkan untuk sitoplasma dihasilkan sebesar 81,76% dan 85,88%.

TABEL 3
HASIL UJI COBA SKENARIO 2

Metode	Rata-rata akurasi nukleus	Rata-rata ZSI nukleus	Rata-rata akurasi sitoplasma	Rata-rata ZSI sitoplasma
<i>RNBGGVFS & mean shift</i>	0,969	0,907	0,868	0,893
<i>RCNNGVFS & FCM_NLS</i>	0,953	0,881	0,835	0,871
<i>RGVFS & FCM</i>	0,949	0,857	0,818	0,860
<i>RGVFS & spatial k-means</i>	0,945	0,866	0,818	0,8588

IV. KESIMPULAN

Segmentasi kontur citra sel tunggal smear serviks dengan metode *Radiating Normally Biased Generalized Gradient Vector Flow Snake* (RNBGGVFS) berhasil dengan nilai rata-rata akurasi dan ZSI dari semua kelas untuk segmentasi nukleus adalah 96,96% dan 90,68%. Nilai rata-rata akurasi dan ZSI dari semua kelas untuk segmentasi sitoplasma adalah 86,78% dan 89,35%. Tahap pra proses berhasil dilakukan untuk mendapatkan citra keabuan dengan cara konversi citra dari ruang warna RGB ke ruang warna CIELAB layer L. Tahap segmentasi awal dengan metode *mean shift clustering* berhasil untuk mendapatkan hasil segmentasi yang baik sebagai data masukan proses segmentasi kontur serta meningkatkan jumlah citra yang dapat dianalisis. Berdasarkan hasil uji coba, metode usulan lebih optimal dalam mendeteksi kontur nukleus dan sitoplasma citra sel tunggal smear serviks dibanding metode RCNNGVFS dan FCM_NLS, RGVFS dan FCM dan RGVFS dan *spatial k-means*.

Pada penelitian ini masih terdapat beberapa citra yang belum berhasil dianalisis dikarenakan karakteristik citra yang saling overlap, maupun terdapat banyak sel dalam satu gambar. Sehingga, untuk penelitian selanjutnya disarankan mencari metode yang dapat menangani permasalahan tersebut.

DAFTAR PUSTAKA

- [1] Kass, M., Witkin, A., & Terzopoulos, D., 1987, "Snakes: Active contour models", *Int. J. Comput. Vis.*, vol. 1, 321–331.
- [2] Xu, Chenyang., Prince, Jerry L., 1998. "Generalized Gradient vector flow external forces for active contours", *Signal Processing*, 71, 131-139.
- [3] Li, Kuan., Lu, Zhi., Liu Wenyin & Jianping Yin, 2012. "Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake", *Pattern Recognition*, 45, 1255-1264.
- [4] Susanti M.D.E., Tjandrasa H. & Faticah C., 2015. "Segmentasi Nukleus dan Sitoplasma pada Citra Smear Serviks menggunakan Kombinasi Metode Fuzzy C- Means Clustering dan Radiating Gradient Vector Flow Snake", Tugas Akhir, Institut Teknologi Sepuluh Nopember, Surabaya.
- [5] Qin, L., Zhu, C., Member, S., Zhao, Y., Member, S., Bai, H., & Tian, H., 2013, "Generalized Gradient Vector Flow for Snakes : New Observations, Analysis, and Improvement", vol.23(5), 883–897.
- [6] Husain, Nursuci Putri., Faticah, C., 2017, Radiating Component Normalized Generalized Gradient Vector Flow Snake untuk Segmentasi Citra Sel Tunggal Smear Serviks, Segmentasi Citra Sel Tunggal Smear Serviks Menggunakan Radiating Component Normalized Generalized GVFS, *JNTETI* 6 (1), 107-114.
- [7] Wang, Yuanquan., 2009. "NBGVF: Normally Biased Gradient Vector Flow External Force for Active Contours", *IEEE Signal Processing Letters*, 17(10), 1-4.
- [8] D. Comaniciu and P. Meer., 2002, Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 24:603–619.
- [9] Ganesan, P. et al., 2010, "Segmentation and Edge Detection of Color Images Using CIELAB Color Space and Edge Detectors", *IEEE On Emerging Trends in Robotics and Communication Technologies (INTERACT)*, 393-397.
- [10] Zhang, Rui., Zhu, Shiping., & Zhou, Qin., 2016. "A novel gradient vector flow snake model based on convex function for infrared image segmentation", *Sensors*.
- [11] A. Zijdenbos, B. Dawant, R. Margolin and A. Palmer, 1994, "Morphometric analysis of white matter lesions in MR images: method and validation," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 716-724.
- [12] Jantzen J. & Dounias G., The Pap-Smear Benchmark, 2008. Website: <http://mde-lab.aegean.gr/index.php/downloads>. diakses tanggal 12 Mei 2018.
- [13] Chuanyun, Xu., Yang, Zhang., & Sen, Wang (2013). "Cell Segmentation in Cervical smear images using polar coordinates GVF Snake with Radiating Edge Map", *Journal of Multimedia*, 8(3), 213-219.
- [14] Kale, A., & Aksoy, S., 2010, "Segmentation of Cervical Cell Images", *IEEE on International Conference Pattern Recognition (ICPR)*.
- [15] Guan, Tao., Zhou, Dongxiang & Liu, Yunhui (2015). "Accurate Segmentation of Partially Overlapping Cervical Cells Based on Dynamic Sparse Contour Searching and GVF Snake Model". *IEEE Journal of Biomedical and Health Informatics*, 19(4), 1494-1504.