

Segmentasi Citra Berbasis Clustering Menggunakan Algoritma Fuzzy C-Means

I Made Budi Adnyana, I Ketut Gede Darma Putra, dan I Putu Agung Bayupati

Abstract— Clustering based image segmentation in this study using Fuzzy C means algorithm with Xie Beni Index as an objective function. Preprocessing applied in this model using Statistical Region merging. Spatial function applied in Fuzzy C means method to reduce noise in clustering. The system evaluation is done by measuring cluster validity value (Xie Beni Index), execution time, and number of iteration. Experimental results on three test images illustrates the proposed method able to perform image segmentation well.

Index Term— image segmentation, fuzzy c-means, clustering

Abstrak— Segmentasi citra berbasis clustering pada penelitian ini menggunakan metode Fuzzy C Means dengan menerapkan fungsi objektif Xie Beni Index. Preprocessing diterapkan pada model yang dikembangkan ini menggunakan metode Statistical Region Merging. Spatial function diterapkan pada metode Fuzzy C Means untuk mengurangi noise pada saat clustering. Evaluasi sistem dilakukan dengan pengukuran nilai cluster validity (Xie Beni Index), waktu eksekusi, dan jumlah iterasi. Hasil pengujian pada tiga buah citra uji menunjukkan metode yang diusulkan mampu melakukan segmentasi citra dengan baik..

Kata kunci— Segmentasi citra, fuzzy c-means, clustering

I. PENDAHULUAN

SEGMENTASI citra (*image segmentation*) merupakan teknik untuk membagi suatu citra menjadi beberapa daerah (*region*) di mana setiap daerah memiliki kemiripan atribut. Salah satu teknik dalam segmentasi citra adalah dengan clustering. Segmentasi citra berbasis cluster menggunakan data multidimensi untuk mengelompokkan piksel citra ke dalam beberapa cluster. Pada umumnya piksel di-cluster berdasarkan kedekatan jarak antar piksel[1]. Terdapat beberapa algoritma clustering yang dapat digunakan dalam menyelesaikan permasalahan segmentasi citra seperti algoritma K-means, Fuzzy C-Means, ISODATA, dan Snob. Fuzzy C-Means merupakan algoritma yang populer digunakan dalam teknik fuzzy clustering. Pengembangan algoritma Fuzzy C-Means untuk segmentasi citra ini pun telah dilakukan

dengan mengkombinasikannya dengan beberapa algoritma metaheuristik atau algoritma optimasi seperti Genetic Algorithm [2] dan Particle Swarm Optimization [3][4]. Perkembangan teknik-teknik segmentasi citra ini dilakukan untuk memperoleh hasil yang lebih bagus dan berkualitas. Namun, terkadang segmentasi citra memberikan hasil yang kurang memuaskan pada suatu citra tertentu. Hal ini dapat dipengaruhi oleh beberapa faktor variasi yang terdapat pada citra itu sendiri seperti pencahayaan, citra yang mengalami degradasi atau citra yang warnanya kabur, dan citra yang berisi noise. Tujuan dari penelitian ini adalah mengkombinasikan dan mengimplementasikan algoritma Fuzzy C Means dengan Statistical Region Merging untuk proses segmentasi citra berbasis clustering, serta melakukan uji coba untuk mengetahui unjuk kerja sistem segmentasi citra yang diusulkan ini.

Metode Statistical Region Merging (SRM) merupakan metode yang sejenis dengan teknik region growing and merging. Pada region growing, region adalah kumpulan piksel yang homogen dan secara iterasi merupakan hasil gabungan dari region yang lebih kecil. Statistical Region Merging mengikuti urutan tertentu dari pemilihan region.[5] Penelitian ini menerapkan algoritma Fuzzy C Means dengan Spatial Function untuk proses segmentasi citra berbasis clustering. Metode SRM digunakan sebagai preprocessing untuk menyederhanakan kompleksitas warna yang terdapat pada citra uji agar mudah disegmentasi. Citra yang digunakan untuk uji coba model yang dikembangkan ini adalah beberapa citra uji dari Weizmann Segmentation Dataset[6]. Pengujian dilakukan 9 kali untuk setiap citra uji, masing-masing dengan menggunakan 2 sampai 10 jumlah cluster, untuk mengetahui jumlah cluster optimal dari setiap citra uji..

II. METODE PENELITIAN

A. Gambaran Umum Sistem

Penelitian ini mencoba menerapkan metode Fuzzy C Means dengan Spatial Function dan Statistical Region Merging sebagai preprocess (FCMSRM) untuk melakukan segmentasi citra. Citra sampel yang digunakan dalam penelitian ini diperoleh dari Weizmann Segmentation Dataset. Jumlah citra uji yang digunakan sebanyak 3 buah. Pengujian dilakukan pada masing-masing citra uji dengan menggunakan algoritma FCMSRM yang diusulkan. Masing-masing pengujian menggunakan 2 sampai 10 cluster untuk mengetahui jumlah cluster yang paling optimal. Hasil uji coba sistem ini

I Made Budi Adnyana adalah Magister Teknik Elektro, Program Pasca Universitas Udayana (e-mail: b.adnyana@yahoo.co.id).

I Ketut Gede Darma Putra adalah dengan Magister Teknik Elektro Program Pasca Sarjana Universitas Udayana (e-mail: ikgdarmaputra@gmail.com).

I Putu Agung Bayupati adalah Magister Teknik Elektro Program Pasca Sarjana Universitas Udayana (e-mail: bayuhelix@yahoo.com).

dievaluasi dengan mengukur nilai cluster validity, waktu eksekusi, dan jumlah iterasi.

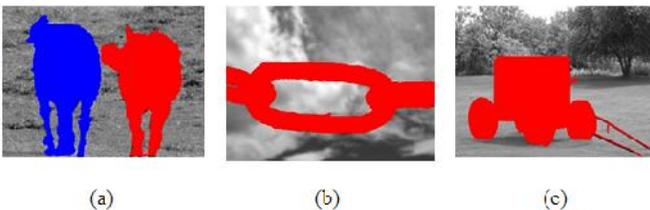
B. Data Uji

Citra uji yang digunakan pada penelitian ini adalah Weizmann Segmentation Dataset yang diperoleh dari situs www.weizmann-usa.org yang disediakan oleh Weizmann Institute of Science [6], dimana situs tersebut menyediakan berbagai macam citra uji yang khusus digunakan untuk proses segmentasi citra. Berbagai macam citra uji dengan beragam variasi dan karakteristik tersebut dibungkus kedalam sebuah dataset yang dapat diunduh secara gratis. Dari sekian banyak citra uji yang tersedia, pada penelitian ini hanya digunakan tiga buah citra uji yang sudah dipilih berdasarkan karakteristik warna yang bervariasi, dapat dilihat pada Gambar 1.



Gambar 1. Citra uji: (a) horses.png, (b) chain.png, (c) carriage.png

Masing-masing citra uji yang terdapat dalam Weizmann Segmentation Dataset ini memuat citra ground truth segmentation, yaitu hasil segmentasi yang diharapkan dari masing-masing citra uji. Citra ground truth ini dibuat secara manual sesuai dengan persepsi manusia. Citra ground truth segmentation dapat dilihat pada Gambar 2.



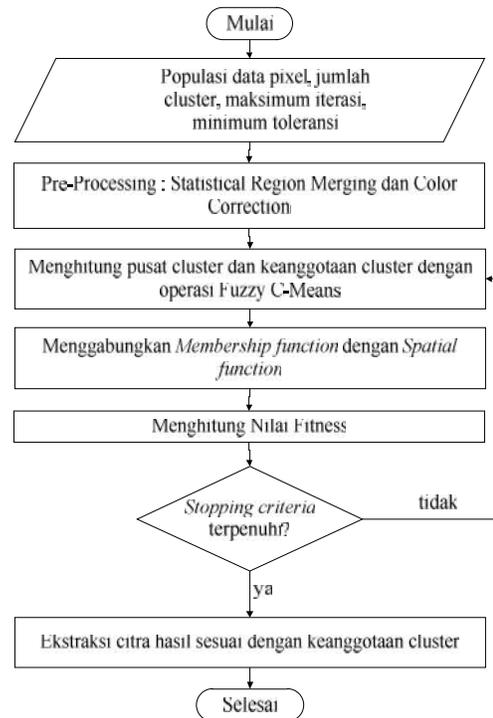
Gambar 2. Citra ground truth segmentation : (a) gt-horses.png, (b) gt-chain.png, (c) gt-carriage.png

C. Segmentasi Citra Menggunakan FCMSRM

Data yang akan diolah pada model algoritma ini adalah nilai piksel yang terdapat pada citra uji dalam ruang warna RGB. Input dari model ini adalah berupa populasi data piksel citra, jumlah cluster (k), maksimum iterasi dan minimum toleransi, serta nilai Q (Qualification Criteria) yang akan digunakan pada proses SRM. Langkah-langkah dari segmentasi citra menggunakan model FCMSRM ini dideskripsikan seperti Gambar 3.

Preprocessing

Metode Statistical Region Merging (SRM) ini sejenis dengan teknik region growing and merging. Pada region growing, region adalah kumpulan piksel yang homogen dan secara iterasi merupakan hasil gabungan dari region yang lebih kecil. Statistical Region Merging mengikuti urutan tertentu dari pemilihan region.



Gambar 3. Segmentasi citra menggunakan Algoritma FCMSRM

Algoritma ini digunakan untuk mengevaluasi nilai-nilai dalam sebuah rentang regional dan dikelompokkan bersama berdasarkan kriteria merging sehingga menghasilkan list yang lebih kecil. Pada bidang pengolahan citra digital berfungsi mengelompokkan piksel-piksel yang bertetangga berdasarkan shades yang jatuh dalam sebuah threshold tertentu (Qualification Criteria). Parameter yang diperlukan dalam proses ini adalah Q yaitu kompleksitas segmentasi[5].

Menghitung Pusat dan keanggotaan Cluster

Algoritma Fuzzy C-Means menempatkan piksel citra kedalam masing-masing cluster dengan menggunakan membership fuzzy. Misalkan $X=(x_1, [x]_2, \dots, x_N)$ menunjukkan citra dengan N buah piksel yang akan dipartisi kedalam c buah cluster. Algoritma ini merupakan optimasi iteratif yang meminimalisasi cost function, dimana cost function yang digunakan dalam penelitian ini adalah fungsi objektif Xie Beni Index.

Cost function diminimalisasi jika piksel berada dekat dengan pusat clusternya akan diberikan nilai keanggotaan cluster yang tinggi, sebaliknya nilai keanggotaan yang rendah diberikan jika piksel berada jauh dari pusat clusternya. Membership function menunjukkan probabilitas suatu piksel dapat dimiliki oleh suatu cluster tertentu. Nilai probabilitas pada algoritma Fuzzy C-Means sangat tergantung pada jarak suatu piksel ke masing-masing pusat cluster pada domain fitur. Membership function dan pusat cluster diperbaharui dengan persamaan berikut :

$$\mu_{ik} = \frac{[\sum_{j=1}^m (P_i - V_k)^2]^{-1}}{\sum_{K=1}^c [\sum_{j=1}^m (P_i - V_K)^2]^{-1}} \quad (1)$$

dan

$$v_k = \frac{\sum_{i=1}^N \mu_{ik}^m p_i}{\sum_{i=1}^N \mu_{ik}^m} \quad (2)$$

Dimulai dengan menghitung pusat cluster pada inisialisasi, Fuzzy C-Means akan menyempitkan atau konvergensi ruang solusi v_k yang menunjukkan local minimum. Konvergensi dapat dideteksi dengan membandingkan perubahan yang terjadi pada membership function atau pusat cluster pada dua buah iterasi. Tahap-tahap yang dilakukan pada langkah ini dapat diuraikan detailnya sebagai berikut:

- a) Menentukan inisialisasi nilai awal pada komponen-komponen Fuzzy C-Means, meliputi banyaknya cluster yang diinginkan, pangkat atau pembobot, maksimum iterasi, minimum toleransi, fungsi objektif awal, dan iterasi awal.
- b) Membangkitkan matriks partisi U_{ik} dengan komponen i = banyaknya data; dan k = banyak cluster (pertama kali dibangkitkan secara bebas atau random, dengan kisaran nilai dari 0 sampai 1).
- c) Setelah matriks partisi terbentuk, langkah selanjutnya adalah menghitung posisi pusat cluster dengan Persamaan (2).
- d) Setelah diperoleh pusat-pusat cluster yang baru, tahap selanjutnya adalah menghitung fungsi objektif atau cost function dengan menerapkan Persamaan (4).
- e) Tahap terakhir dari langkah ini adalah memperbaharui matriks partisi μ dengan menggunakan Persamaan (1).

Menggabungkan Membership Function dengan Spatial Function

Salah satu karakteristik dari citra adalah bahwa piksel-piksel tetangga sangat berhubungan atau berkaitan. Dengan kata lain, piksel yang bertetangga memiliki nilai fitur yang hampir mirip, dan memiliki probabilitas yang besar mereka berada dalam cluster yang sama. Hubungan spasial ini sangat penting dalam *clustering*, tapi hal ini tidak diperhitungkan dalam algoritma Fuzzy C-Means standar. Untuk menggali informasi spasial, fungsi spasial dapat dirumuskan sebagai berikut :

Dimana $NB(P_i)$ adalah *square window* yang berpusat pada piksel P_i pada domain spasial. Misalnya *window* berukuran 5x5 akan digunakan dalam penelitian ini, maka perhitungan dan pengecekan dengan fungsi spasial ini akan dilakukan hanya pada piksel *window* itu saja. Sama seperti *membership function*, *spatial function* h_{ij} ini merepresentasikan probabilitas piksel x_j dimiliki oleh cluster ke- i . *Spatial function* dari suatu piksel terhadap suatu cluster bernilai besar jika sebagian besar piksel-piksel tetangganya dimiliki oleh cluster yang sama. *Spatial function* ini digabungkan dengan *membership function* menjadi persamaan sebagai berikut :

$$u'_{ik} = \frac{u_{ik}^p * h_{ik}^q}{\sum_{k=1}^c u_{ik}^p * h_{ik}^q} \quad (3)$$

Dimana p dan q adalah parameter yang mengontrol keperluan yang bersifat *relatif* dari fungsi tersebut. Dalam region yang homogen, *spatial function* tersebut dapat dengan mudah membatasi fungsi membership yang asli, dan hasil

clustering tidak akan berubah. Namun pada piksel yang mengandung *noisy*, formula ini mengurangi bobot *noisy* pada cluster dengan memberikan label pada piksel-piksel tetangganya. Hasilnya, *misclassified* piksels dari region yang mengandung *noisy* dapat dengan mudah diperbaiki [7][8].

Menghitung Nilai Fitness

Fitness function atau *cost function* yang digunakan pada model algoritma ini adalah Xie-Beni Cluster Validity Index[9], yang mendefinisikan fungsinya sebagai berikut :

$$V_{xb} = \frac{-\sum_j \sum_i u_{ij} \|X_j - V_i\|^2}{N \left(\min_{i \neq j} \{ \|V_j - V_i\|^2 \} \right)} \quad (4)$$

V_i adalah pusat cluster ke- i dengan jumlah cluster adalah c . X_j merupakan nilai warna dari piksel ke- j dan N adalah jumlah total piksel yang dimiliki oleh citra. Konsep dari fungsi validasi cluster ini adalah *clustering* yang bagus adalah yang menghasilkan data sampel yang dipadatkan kedalam sebuah cluster dan sampel-sampel tersebut dipisahkan antara satu cluster dengan cluster yang lainnya. Kualitas *clustering* yang bagus diperoleh dengan meminimalkan nilai V_{xb} .

Ulangi dari Langkah 2 selama stopping criteria belum terpenuhi

Stopping criteria yang digunakan pada metode ini dibagi menjadi dua, yaitu menggunakan maksimum iterasi dan minimum toleransi. Nilai maksimum iterasi yang digunakan adalah 100 iterasi, artinya ketika segmentasi citra sudah mencapai 100 iterasi maka prosesnya akan dihentikan. *Stopping criteria* yang kedua adalah batas minimum toleransi. Nilai toleransi adalah selisih antara nilai fitness pada satu iterasi dengan iterasi sebelumnya. Batas minimum toleransi yang digunakan pada penelitian ini adalah 0.0000001. Jika nilai toleransi lebih kecil dari batas minimumnya maka proses segmentasi citra akan berhenti. Jika *stopping criteria* ini sudah tercapai maka akan dilanjutkan ke Langkah No.6, sebaliknya jika belum terpenuhi maka proses akan diulangi dari Langkah No.2 dan masuk ke iterasi selanjutnya ($i++$).

Ekstraksi citra hasil segmentasi

Setelah proses FCMSRM selesai, maka diperoleh nilai fitness dan pusat-pusat cluster terbaik dari semua iterasi. Selanjutnya nilai dari masing-masing pusat cluster dipetakan pada semua data piksel yang menjadi anggotanya, sehingga semua piksel berubah nilai warnanya sesuai dengan keanggotaan clusternya masing-masing.

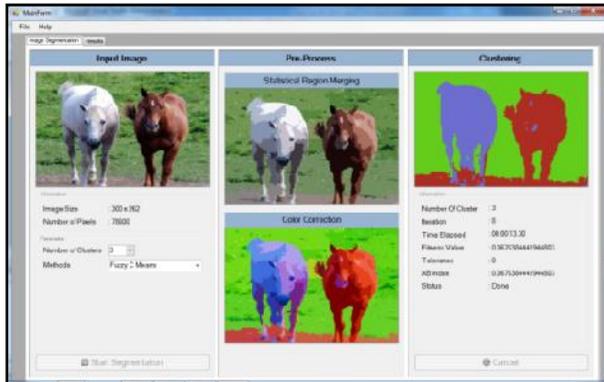
Pengujian dan Evaluasi Hasil

Pengujian dilakukan dengan memproses citra uji menggunakan metode FCMSRM menggunakan 2 sampai 10 cluster. Evaluasi unjuk kerja sistem yang dikembangkan ini dilakukan dengan mengukur nilai *cluster validity index*, rata-rata waktu eksekusi dan jumlah iterasi untuk setiap citra yang diuji. Pengukuran nilai *cluster validity index* bertujuan untuk menilai kualitas hasil *clustering* suatu algoritma. Pengukuran rata-rata waktu eksekusi bertujuan untuk mengetahui seberapa cepat algoritma yang dikembangkan dapat menyelesaikan segmentasi citra.

Untuk menilai kualitas partisi yang dihasilkan oleh suatu algoritma *clustering*, perlu digunakan suatu fungsi yang disebut dengan *cluster validity index* dimana fungsi ini melakukan evaluasi terhadap hasil akhir dari sebuah *clustering*. Fungsi *cluster validity index* yang akan digunakan pada penelitian ini adalah fungsi validitas Xie-Beni Index menggunakan persamaan (4) diatas.

III. HASIL DAN PEMBAHASAN

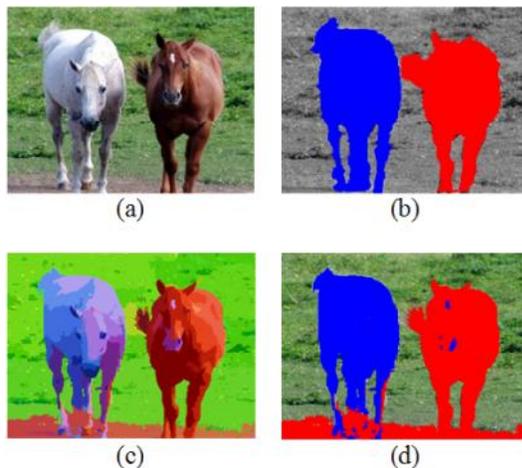
Sistem segmentasi citra ini dibuat menggunakan bahasa pemrograman C# pada Visual Studio 2008, tampilan antarmukanya dapat dilihat pada Gambar 5.



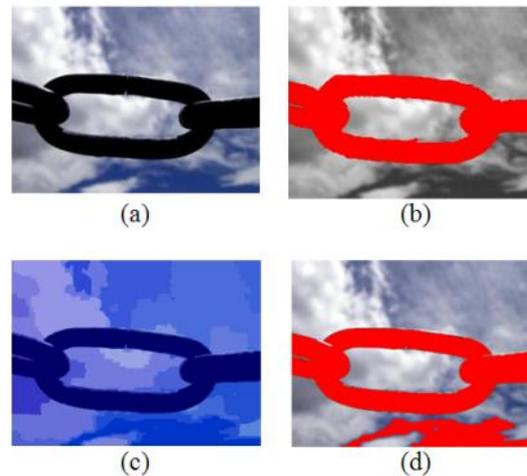
Gambar 4. Tampilan antarmuka sistem

Pengujian sistem ini dilakukan dengan memproses masing-masing citra uji dengan algoritma FCMSRM, masing-masing menggunakan 2 sampai 10 cluster. Parameter atau variabel tetap yang digunakan pada pengujian ini adalah maksimum iterasi = 100, minimum toleransi = 0.0000001, window berukuran 5x5 untuk operasi spatial function, dan nilai $Q = 64$.

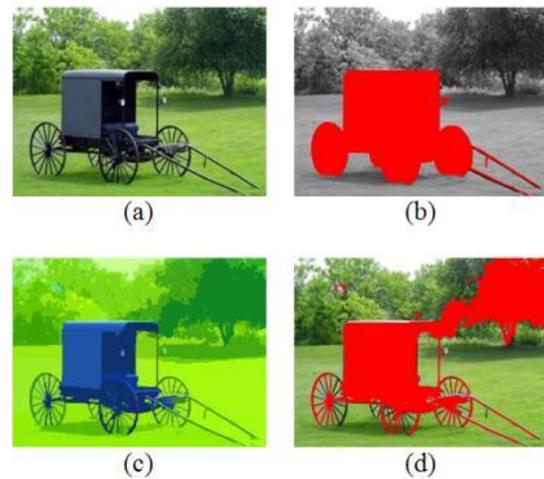
Gambar 5 sampai dengan Gambar 9 menunjukkan citra hasil uji coba segmentasi menggunakan metode FCMSRM. Citra yang ditampilkan adalah original image (citra asli), ground truth segmentation (citra hasil yang diharapkan), preprocessed image (citra hasil preprocessing dengan metode SRM), dan result image (citra hasil segmentasi dengan FCMSRM)



Gambar 5. Hasil segmentasi citra uji "horses.png": (a) original image, (b) ground truth segmentation, (c) preprocessed image, (d) FCMSRM result image



Gambar 6. Hasil segmentasi citra uji "chain.png": (a) original image, (b) ground truth segmentation, (c) preprocessed image, (d) FCMSRM result image



Gambar 7. Hasil segmentasi citra uji "carriage.png": (a) original image, (b) ground truth segmentation, (c) preprocessed image, (d) FCMSRM result image

Berdasarkan gambar hasil pengujian diatas, dapat dilihat metode FCMSRM yang diusulkan ini mampu dengan baik melakukan segmentasi terhadap citra uji "horses" dan "chain" sehingga sudah mendekati citra ground truth segmentation. Namun pada citra "carriage" hasil clustering kurang optimal, dapat dilihat bagian background masuk kedalam cluster objek, hal ini disebabkan oleh warna background hampir sama dengan warna objek.

Pengujian sistem ini menggunakan 2 sampai 10 jumlah cluster, hasil perbandingan jumlah iterasi ditunjukkan pada Tabel I, perbandingan waktu eksekusi ditunjukkan pada Tabel II, dan perbandingan nilai Xie Beni Index ditunjukkan pada Tabel III.

Berdasarkan tabel perbandingan nilai Xie-Beni Index diatas, nilai XB Index paling kecil pada citra "horses.png" adalah pada jumlah cluster = 3. Citra uji lainnya menghasilkan nilai XB Index paling kecil pada jumlah cluster = 2. Nilai XB index paling kecil ini mengindikasikan jumlah cluster yang paling optimal dari masing-masing citra uji. Rata-rata jumlah iterasi, waktu eksekusi dan nilai XB Index dari pengujian setiap citra uji ditunjukkan pada Tabel 4.

TABEL I
PERBANDINGAN JUMLAH ITERASI

Citra Uji	Jumlah Cluster								
	2	3	4	5	6	7	8	9	10
horses	9	15	14	30	12	24	26	34	19
chain	10	20	100	16	100	100	100	100	57
carriage	12	11	9	23	37	100	100	10	100

TABEL II
PERBANDINGAN WAKTU EKSEKUSI (DETIK)

Citra Uji	Jumlah Cluster				
	2	3	4	5	6
horses	0:00:07	0:00:25	0:00:25	0:01:13	0:00:38
chain	0:00:07	0:00:21	0:02:40	0:00:35	0:05:25
carriage	0:00:07	0:00:09	0:00:11	0:00:38	0:01:20

Citra Uji	Jumlah Cluster			
	7	8	9	10
horses	0:01:36	0:02:07	0:03:21	0:02:13
chain	0:06:00	0:07:05	0:09:02	0:05:48
carriage	0:04:32	0:05:34	0:00:40	0:07:57

TABEL III
PERBANDINGAN NILAI XIE-BENI INDEX

Citra Uji	Jumlah Cluster				
	2	3	4	5	6
horses	0.20718	0.06764	0.11321	0.07253	0.53544
chain	0.05578	0.10344	0.05911	0.30605	0.77657
carriage	0.09657	0.09769	0.20807	0.10056	0.15587

Citra Uji	Jumlah Cluster			
	7	8	9	10
horses	0.34253	0.32665	1.56001	1.49813
chain	0.22451	0.58285	0.06137	0.10779
carriage	0.10936	0.29649	0.11041	0.10371

TABEL IV
RATA-RATA ITERASI, WAKTU EKSEKUSI DAN XB INDEX

Citra Uji	Rata-rata		
	Iterasi	Waktu eksekusi (detik)	XB Index
horses	20	80	0.5248133
chain	37	247	0.2530522
carriage	45	141	0.1420811

IV. KESIMPULAN

Metode FCMSRM yang dikembangkan ini mampu melakukan segmentasi citra dengan menggunakan metode Fuzzy C Means dengan spatial function, serta preprocessing menggunakan metode SRM berfungsi untuk menyederhanakan kompleksitas warna pada citra uji agar lebih mudah untuk disegmentasi. Algoritma Fuzzy C Means pada penelitian yang diusulkan ini menggunakan Xie Beni Index sebagai fungsi objektifnya. Evaluasi hasil menggunakan pengukuran terhadap nilai XB index, waktu eksekusi dan jumlah iterasi. Hasil uji coba sistem menunjukkan segmentasi pada citra uji "horses.png" diperoleh rata-rata jumlah iterasi adalah 20, rata-rata waktu eksekusi 80 detik, rata-rata XB Index 0.5248133, dan jumlah cluster paling optimal adalah 3. Segmentasi pada citra uji "chain.png" diperoleh rata-rata jumlah iterasi adalah 67, rata-rata waktu eksekusi 247 detik, rata-rata XB Index 0.2530522, dan jumlah cluster paling optimal adalah 2. Segmentasi pada citra uji "carriage.png" diperoleh rata-rata jumlah iterasi adalah 45, rata-rata waktu eksekusi 141 detik, rata-rata XB Index 0.1420811, dan jumlah cluster paling optimal adalah 2. Hasil pengujian pada tiga

buah citra uji menunjukkan metode yang diusulkan mampu melakukan segmentasi citra dengan baik

V. DAFTAR PUSTAKA

- [1] D. Putra, "Pengolahan Citra Digital", Yogyakarta : Andi Offset, 2010
- [2] A. Halder, dan S. Pramanik, "Dynamic Image Segmentation using Fuzzy C-Means based Genetic Algorithm", International Journal of Computer Application, Vol. 28, No.6, 2011.
- [3] Mahamed G. H., Omran, Andries P Engelbrecht, Ayed Salman. 2005. Dynamic Clustering using Particle Swarm Optimization with Application in Unsupervised Image Classification. Transactions on Engineering, Computing and Technology, vol. 9 : 199-204
- [4] Swagatam Das, Ajith Abraham. 2010. Spatial Information Based Image Segmentation Using a Modified Particle Swarm Optimization Algorithm. IITA Professorship Program, School of Computer Science and Engineering
- [5] Nock, Richard; Frank Nielsen. 2004. Statistical Region Merging. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (11): 1-7
- [6] S Alpert, M Galun, R Basri, A Brandt. 2007. Image Segmentation by Probabilistic Bottom-Up Aggregation and Cue Integration. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR'07 : 1-8
- [7] Keh-Shih Chuang, Hong-Long Tzeng, Sharon Chen, Jay Wu, Tzong-Jer Chen. 2006. Fuzzy c-means clustering with spatial information for image segmentation. Computerized Medical Imaging and Graphics, vol. 30 : 9-15
- [8] Ahmed, M. N., Yamany, S. M., Mohamed, N., Farag, A. A., Moriarty, T. A. 2002. Modified Fuzzy C-Means Algorithm for Bias Field Estimation and Segmentation of MRI data. IEEE Trans Med Imaging, vol. 21 : 193-9
- [9] Xie XL, Beni GA. 1991. Validity Measure For Fuzzy Clustering. IEEE Trans Pattern Anal Mach Intell Vol.3 : 841-6