IMPLEMENTASI SEGMENTASI CITRA CTSCAN PARU-PARU PADA *BRONCHUS* MENGGUNAKAN *K-MEANS CLUSTERING* DAN *WATERSHED*

Bagus Perdana[1]

[1] Teknik Informatika - Universitas Komputer Indonesia
Jl. Dipatiukur 112 - 114 Bandung
E-Mail : Bagusperdanaactive@gmail.com

ABSTRAK

Beberapa Teknologi yang digunakan di bidang kesehatan adalah Rontgen / foto rontgen, MRI (*Magnetic Resonance Imaging*) dan CT Scan. Banyak dokter yang memanfaatkan CT Scan sebagai salah satu alat untuk memindai penyakit atau gejala yang ada pada organ tubuh tertentu.

Salah satu teknik pemrosesan citra yang digunakan untuk mengatasi tingkat derau yang tinggi adalah metode *watershed*. Metode *watershed* digunakan pada citra CT scan *bronchus* untuk mendeteksi pola pada citra agar bisa mengenali *bronchus* dalam kedaan normal atau tidak normal. Metode *watershed* digunakan untuk mendeteksi ciri-ciri apa saja yang terbentuk dan menenentukan apakah pola yang terbentuk dari hasil metode *watershed* adalah *bronchus* normal atau tidak normal. Metode *k-means clustering* akan digunakan untuk mendeteksi pola tersebut dan akan dikategorikan menjadi dua kategori yaitu normal dan tidak normal.

Dari hasil penelitian, analisis pengolahan citra, pelatihan dan pengujian metode *watershed* dan *k-means clustering* untuk mengidentifikasi pola citra CT scan *bronchus* untuk mendapatkan informasi keadaan *bronchus* normal dan tidak normal. Adapun akurasi dari penerapan metode adalah 76.6% dan dapat mengenali citra CT scan dengan baik.

1. PENDAHULUAN

Bronchus merupakan saluran nafas yang terbentuk dari belahan dua trakea pada ketinggian kira-kira vertebrata torakalis kelima, mempunyai struktur serupa dengan trakea dan dilapisi oleh jenis sel yang sama. *Bronchus* berjalan ke arah bawah dan samping menuju paru dan bercabang menjadi dua, yaitu *bronchus* kanan dan *bronchus* kiri [1].

Beberapa Teknologi yang digunakan di bidang kesehatan adalah Rontgen / foto rontgen, MRI (*Magnetic Resonance Imaging*) dan CT Scan. Banyak dokter yang memanfaatkan CT Scan sebagai salah satu alat untuk memindai penyakit atau gejala yang ada pada organ tubuh tertentu. Hasil CT scan mempunyai beberapa kelemahan diantaranya, memiliki derau yang tinggi hal ini disebabkan karena paparan radiasi sinar X yang digunakan sekitar 4% dari radiasi sinar X yang digunakan pada radiologi / foto rontgen sehingga mengurangi akurasi gambar dan ada gambar yang seharusnya tidak ada menjadi terekam diakibatkan pasien bergerak saat perekaman gambar.

1.1. Pengolahan Citra

Untuk melakukan pengenalan suatu objek berdasarkan tekstur yang dimiliki objek tersebut maka dibutuhkanlah suatu teknik untuk melakukannya, teknik tersebut adalah teknik pengolahan citra. Pengolahan citra merupakan proses pengolahan dan analisis citra yang banyak melibatkan persepsi visual. Proses ini mempunyai ciri data masukan dan informasi keluaran yang berbentuk citra. Istilah pengolahan citra digital secara umum didefinisikan sebagai pemrosesan citra dua dimensi dengan komputer. Dalam definisi yang lebih luas, pengolahan citra digital juga mencakup semua data dua dimensi.

1.2. Elemen Citra

Dalam pengolahan citra objek yang digunakan adalah sebuah citra dari sebuah objek tertentu yang mana citra tersebut mengandung sejumlah elemen dasar. Elemen dasar tersebut di manipulasi dalam pengolahan citra, elemen tersebut adalah warna, kecerahan, kontras, kontur, bentuk, tekstur, waktu dan pergerakan dan deteksi dan pengenalan.

1.3 Pola

Pola adalah entitas yang terdefinisi dan dapat diidentifikasi melalui ciri-cirinya (*features*). Ciri-ciri tersebut digunakan untuk membedakan suatu pola dengan pola lainnya. Ciri yang bagus adalah ciri yang memiliki daya pembeda yang tinggi, sehingga pengelompokan pola berdasarkan ciri yang dimiliki dapat dilakukan dengan keakuratan yang tinggi.

1.3. Metode Watershed

Buat M1,M2,M3, ... , MR menjadi kumpulan koordinat titik dalam daerah dengan nilai minimum dari sebuah gambar g(x,y). Gambar di sini pada umumnya sudah merupakan gambar yang telah diproses dengan pre-processing terlebih dahulu. Kemudian buat C(Mi) menjadi kumpulan koordinat pada daerah pengisian yang memiliki hubungan dengan daerah minimum Mi (dianggap daerah pengisian dan daerah minimum membentuk komponen yang saling tersambung). Notasi min dan max digunakan untuk menandai nilai minimum dan nilai maksimum dari g(x,y). Kemudian buat T[n] menjadi kumpulan koordinat (s,t) di mana g(s,t) < n, sehingga dapat didefinisikan:

$$T[n]=\{(s,t)|f(x,y)<n\}$$

Secara geometri, T[n] adalah kumpulan koordinat dari g(x,y) yang berada di bawah daerah g(x,y) = n. Topografi tersebut kemudian diisi dengan penambahan integer mulai dari n = min hingga n = max. Pada setiap penambahan n, algoritma perlu mengetahui jumlah titik yang berada di bawah "kedalaman" n. Pada umumnya, daerah yang berada di bawah g(x,y) = n diberi warna hitam atau nilai 0 dan yang berada di atasnya diberi warna putih atau nilai 1.

Kemudian buat Cn(Mi) menjadi kumpulan koordinat titik pada daerah pengisian yang berhubungan dengan daerah minimum Mi yang diisi pada tahap n. Cn(Mi) dapat dilihat sebagai gambar binary bila menggunakan persamaan:

$$C_n(M_i) = C(M_i) \cap T[n]$$

Dengan kata lain Cn(Mi) = 1 pada lokasi (x,y) jika (x,y) ∈ C(Mi) dan (x,y) ∈ T[n], selain itu Cn(Mi) = 0. Dari sini dapat dikatakan bahwa

$$C[n] = \bigcup_{i=1}^{R} C_n(M_i)$$

dan

$$C[\max+1] = \bigcup_{i=1}^{R} C(M_i)$$

C[n-1] adalah subset dari C[n] dan C[n] adalah subset dari T[n] maka C[n-1] adalah subset dari T[n]. Dari sini didapatkan bahwa tiap komponen yang terhubung dari C[n-1] memiliki 1 komponen yang terhubung dari T[n]. Jika Q adalah kumpulan komponen yang terhubung dalam T[n], maka untuk tiap komponen yang terhubung q ∈ Q[n], maka ada 3 kemungkinan:

1. q ∩ C[n −1] adalah kosong

2. q ∩ C[n −1] mengandung 1 komponen yang terhubung dari C[n-1]

3. q ∩ C[n −1] mengandung lebih dari 1 komponen yang terhubung dari C[n-1]

Jika kondisi c terjadi maka pengisian akan menyebabkan 2 daerah menjadi 1, maka titik pada C[n-1] menjadi milik dam atau watershed line.

1.4. Parameter Nilai Citra

1.4.1. Area

Untuk memudahkan mendeteksi dan pemahaman mengenai ciri-ciri pada suatu object. Dimisalkan saja pada sebuah citra yang diperiksa terdapat satu object dan langsung diterapkan dengan cara lebih sederhana. Notasi O dalam persamaanpersamaan berikut adalah titik atau piksel milik object dalam citra biner, sedangkan indeks *x* dan *y* adalah lokasi titik tersebut dalam matriks berukuran *m x n.*

$$A = \sum_{x=1}^{m} \sum_{y=1}^{n} O(x,y)$$

1.4.2. Perimeter

Perimeter adalah Bagian Terluar dari suatu object yang bersebelahan dengan Piksel atau puksel-puksel dari latar belakang. Oleh karena itu perimeter mempunyai beberapa definisi yang berbeda namun memiliki maksud yang sama.

1.4.3. Compactness

Kombinasi beberapa sifat geometri object dapat digunakan untuk menganalisis bentuk object tersebut melalui perhitungan factor bentuk tak berdimensi. Kekompakan (*Compactness*) suatu object dapat diukur melalui analisis factor bentuk tak berdimensi dengan ketidaksamaan *isoperimetric* yang diwakili persamaan berikut

$$C = \frac{4 . \pi . A}{P^2}$$

Dimana C, A dan P masing-masing adalah *compactness*, Area dan Perimeter.

1.5. Metode K-Means Clustering

Data Clustering merupakan salah satu metode Data Mining yang bersifat tanpa arahan (unsupervised). Ada dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu hierarchical (hirarki)

data clustering dan nonhierarchical (non hirarki) data clustering. K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan objective function yang diset dalam proses clustering, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu cluster dan memaksimalisasikan variasi antar cluster.

Data clustering menggunakan metode K-Means ini secara umum dilakukan dengan algoritma dasar sebagai berikut [8] :

1. Tentukan Jumlah Cluster
2. Alokasikan data ke dalam cluster secara random
3. Hitung Centroid/rata-rata dari data yang ada masing-masing cluster
4. Alokasikan masing-masing data ke centroid/rata-rata terdekat
5. Kembali ke step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada Objective Function yang digunakan diatas nilai Threshold yang ditentukan

2. ISI PENELITIAN

2.1. Analisis Sistem

Analisis sistem merupakan suatu cara yang digunakan dalam mengidentifikasi dan mengevaluasi suatu permasalah, hambatan yang dihadapi dan kebutuhan yang diharapkan dalam sebuah sistem. Analisis ini diperlukan sebagai dasar untuk tahapan perancangan sistem. Hasil dari proses analisis sistem ini akan menghasilkan berbagai kesimpulan dan saran yang dapat digunakan untuk dijadikan dasar perancangan aplikasi yang akan dibangun.

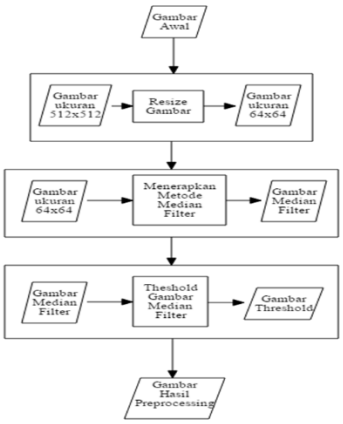
2.2. Analisis Dataset

Penelitian ini menggunakan dataset gambar CT scan bronkus paru-paru manusia yang merupakan basis data citra dengan format JPEG dengan ukuran rata-rata citra adalah 64x64.

2.3. Analisis Metode

Analisis metode digunakan untuk mengetahui alur proses dari sebuah metode yang digunakan dapat diterapkan kedalam aplikasi yang dibangun. Pembangunan aplikasi ini menggunakan metode *watershed* untuk proses segmentasi citra dan *k-means clustering* untuk klasifikasi citra yang telah disegmentasi.

2.3.1. Citra Latih



Gambar 1 Flowchart Pre-Processing Citra

Citra latih adalah citra yang dimasukkan guna mengisi data latih yang telah ada adan menambah kemampuan akurasi dari metode yang digunakan adapun Citra latih yang digunakan adalah citra CT scan bronkus.

Citra latih yang dimasukan akan mendapatkan beberapa proses pengolahan agar *noise* pada citra latih dapat dihilangkan sehingga dapat memudahkan untuk pemrosesan lebih lanjut. Proses tersebut adalah sebagai berikut :

1. Resize
2. Median Filter
3. Thresholding

Hasil dari pengolahan tersebut lalu di gunakan untuk

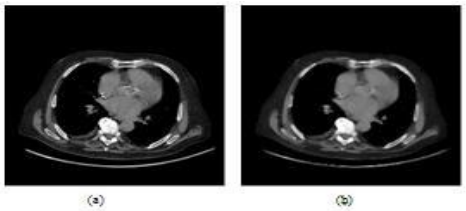
proses segmentasi yang menggunakan metode watershed.

2.3.1.1. Resize

Setelah citra masuk ke lalu akan segera di resize menjadi ukuran 64x64. Lalu akan masuk ke tahap selanjutnya yaitu *median filtering.*

2.3.1.2. Median Filter

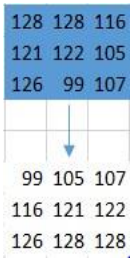
Untuk mengurangi noise yang ada pada citra awal tahap pertama dalam *preprocessing* adalah menerapkan metode median filter.



Gambar 2 Median Filtering (a) Gambar Awal (b) Gambar Hasil Median Filter Dimana Gambar Semakin Halus Dan Noise Hilang

median filter terlihat bahwa proses penghalusannya terjadi pada daerah tepi gambar. Sehingga meski terjadi penghalusan gambar, *median filter* lebih kearah "melembutkan" gambar yang ada. Dapat pula mengurangi komputasi dengan mencari nilai extrem yang perlu untuk diubah, atau dengan membatasi nilai median yang digunakan.

Caranya dengan mengurutkan nilai pada matriks 3x3 yang ada pada citra contohnya gambar 3.4(a) dan mengkonvert gambar tersebut ke dalam matriks dan didapat hasil sebagai berikut :



Gambar 3 Proses Median Filter

2.3.1.3. Threshold

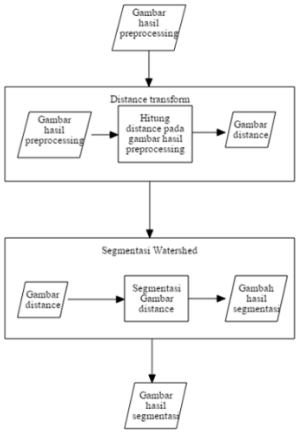
Tahap selanjutnya adalah menerapkan metode *thresholding* untuk mengubah citra awal menjadi bilangan biner. Pada persamaan (2.2)

$$f(x,y) = \begin{cases} 0, & x < 103 \\ 1, & x \geq 103 \end{cases}$$

T = level threshold

yaitu dengan memfilter bilangan pada matriks menjadi dua nilai yaitu 0 dan 1. Level *threshold* diberi nilai 103. Nilai 103 diambil karena nilai tersebut sangat tepat untuk mendapatkan informasi nilai *threshold* apabila menggunakan nilai lebih dari 103 banyak informasi yang hilang dan tingkat akurasi dari metode menurun.

2.3.2. Segmentasi Watershed



Gambar 4 Flowchart Watershed

Setelah nilai citra *median filter* dan melewati tahapan *thresholding* dan nilai telah diproleh, dilakukan proses untuk mencari garis watershed yang mempresentasikan objek yang akan disegmentasikan. Langkah pertama dalam mengoprasikan watershed untuk sebuah citra adalah mengindentifikasi bagian intensitas minimum.

Pertama menghitung nilai *distance* yang ada pada matriks gambar. Dengan metode *cityblock*.

$$d_{ij} = \sum_{k=1}^{n} \left| x_{ik} - x_{jk} \right|$$

0	0	1	1	0
0	1	2	2	1
1	2	3	2	1
1	2	3	3	2
1	2	3	4	3

Gambar 5 Matriks Hasil *Distance* Berdasarkan Contoh Gambar 3.10

Langkah pertama untuk menerapkan metode *watershed* adalah mengindentifikasi nilai minimum dari gambar 3.13, setelah itu dilakukan perbadingan nilai dengan delapan tetangganya dengan menggunakan struktur elemen 3x3. Dimana pusat pixel adalah nilai 1, karena nilai 1 adalah nilai minimum dari semua nilai yang lain.

0	0	1	1	0	0	0
0	1	2	2	1	0	0
1	2	3	2	1	0	0
1	2	3	3	2	1	0
1	2	3	4	3	2	1
1	2	3	4	4	3	2

Gambar 6 Proses Pembetukan Dam Watershed

Langkah selanjutnya pada pusat pixel memutuskan mencari nilai minimum dari delapan tetangganya, misalnya untuk mencari nilai minimum dari delapan 1 tetangganya yang terhubung akan dibandingkan dengan pusat pixel, jika kondisi nilai pusat pixel lebih rendah dari nilai citra yang lain, maka melakukan pencarian nilai citra yang hampir mendekati nilai pusat pixel, untuk menentukan nilai minimum dan garis watershed.

Setelah ditemukan nilai minimunnya, maka akan terjadi garis dari nilai 1 ke nilai 1 yang bertetangga. Konsep yang terjadi pada watershed ini yaitu:

1. Tentukan daerah (region) R1, R2,.., Rn, dengan merepresentasikan objek yang memiliki nilai minimum.

2. Dilakukan perbadingan titik pusat pixel dengan delapan tetangganya dengan menggunakan struktur elemen 3x3.

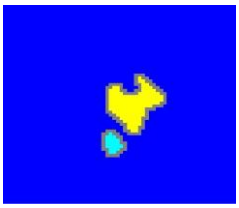
3. Bentuklah dam (garis) pada posisi dimana dua daerah terhubung.

4. Ulangi langkah tiga sehingga semua daerah terhubung, dan didapatkan hasilnya

Hasil yang diperoleh setelah memproses citra threshold, menerapkan distance dan menggunakan segmentasi watershed.



(a) (b)

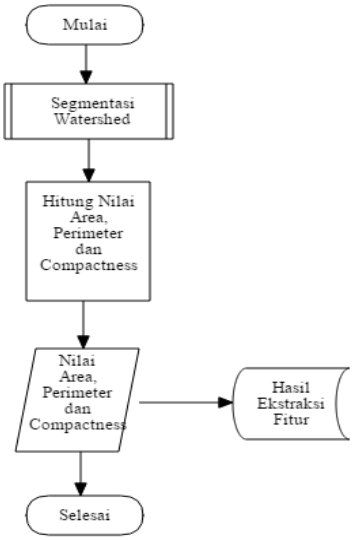


(c)

Gambar 7 Proses Segmentasi Watershed (a) Gambar Threshold (b) Setelah Menerapkan Distance (c) Segmentasi Watershed

2.3.3. Ekstraksi Fitur

Tahapan selanjutnya setelah menentukan segmentasi pada gambar selanjutnya menentukan nilai nilai fitur yang ada pada gambar. Nilai fitur tersebut adalah Area, Perimeter, dan Compactness alur kerja proses *k-means clustering* adalah



Gambar 8 Alur Kerja Ekstraksi Fitur

2.3.3.1. Nilai Area

Notasi O dalam persamaan-persamaan berikut adalah titik atau piksel milik object dalam citra biner, sedangkan indeks *x* dan *y* adalah lokasi titik tersebut dalam matriks berukuran *m x n.*

$$A = \sum_{x=1}^{m} \sum_{y=1}^{n} O(x, y)$$

2.3.3.2. Nilai Perimeter

Perimeter adalah Bagian Terluar dari suatu object yang bersebelahan dengan Piksel dari latar belakang.

2.3.3.3. Nilai Compactness

Kekompakan (*Compactness*) suatu object dapat diukur menggunakan persamaan.

$$C = \frac{4 . \pi . A}{P^2}$$

2.3.4. Pelatihan K-means Clustering

Hasil dari ekstraksi Fitur yang telah di hitung akan dilatih data tersebut sehingga data latih tersimpan. Berikut adalah proses pelatihan k-means:



Gambar 9 Flowchart Pelatihan K-Means

2.3.5. Pengujian

Tahapan pengujian adalah tahapan yang terakhir yaitu untuk menguji ketika gambar atau citra dimasukan untuk dikenali hasil klasifikasinya baik citra yang sudah dilatih ataupun citra yang belum dilatih. Langkah-langkah pada tahap pengujian adalah sebagai berikut :



Gambar 10 Flowchart Pengujian

2.4. Pengujian

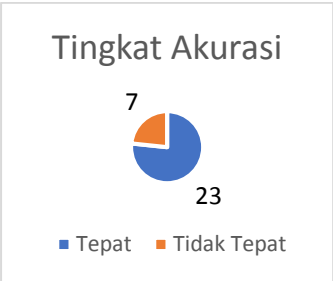
Pengujian metode *k-means clustering* adalah untuk menguji keakurasian dari metode terhadap citra CT scan paru-paru. Penelitian ini menggunakan 100 data, yang akan dibagi menjadi 2, yaitu A1 = 50 data latih dan A2 = 30 data uji. Pengujian metode *k-means clustering* sebagai berikut.

Tabel 1 Pengujian Metode *K-Means Clustering*

Data	Cluster	
	Normal	Tidak Normal
50 data latih	30 data	20 data

Pengujian data terhadap data latih untuk mengetahui citra CT scan *bronchus* normal dan tidak normal.

Tabel 2 Tingkat Akurasi



Gambar 11 Diagram Akurasi

Dari 30 data terdapat 8 data yang tidak sesuai, maka tingkat keakurasian metode *k-means clustering* adalah 76.6%.

3. Penutup

3.1. Kesimpulan

Dari hasil penelitian, analisis pengolahan citra, pelatihan dan pengujian metode *watershed* dan *k-means clustering* untuk mengidentifikasi pola citra CT scan *bronchus* untuk mendapatkan informasi keadaan *bronchus* normal dan tidak normal. Adapun akurasi dari penerapan metode adalah 76.6% dan dapat mengenali citra CT scan dengan baik.

Penerapan metode dapat diimplementasikan dengan baik dan dapat pula diterapkan kasus kasus lain yang berhubungan dengan kesehatan paru-paru dan bagian tubuh lainnya.

3.2. Saran

Dari hasil penelitian, analisis, pengolahan citra, pelatihan dan pengujian terdapat saran-saran yang mungkin akan bermanfaat jika ada yang akan melakukan penelitian yang sejenis, yaitu memastikan validasi dari citra masukan sudah sesuai dengan citra tujuan penelititan. Gambar citra yang baik dan sesuai sangat dibutuhkan untuk mendapatkan tingkat akurasi yang lebih tinggi lagi.

DAFTAR PUSTAKA

[1] Pino P., Kiel dan Suharto., 2013. Pengaruh Lama Waktu Kematian Terhadap Kemampuan Pergerakan Silia Bronkus Hewan Coba Post Mortem Yang Diperiksa Pada Suhu Kamar Dan Suhu Dingin. thesis. Malang : Faculty of Medicine Diponegoro University.

[2] Turnip, Alfian., 2015. Analisis Penggunaan Teknik Post Processing Volume Rendering (Vr) Pada Ct Angiografi Arteri Abdominalis Dengan MSCT. Skripsi. Medan : Universitas Sumatra Utara.

[3] Nalwan, A., 1997. Pengolahan Gambar Secara Digital. Jakarta: Elex Media Komputindo

[4] Afrizal, Syafitri., 2013. Segmentasi Bronchus dan Bronchiolus pada Citra CT Scan Paru – Paru Menggunakan Watershed Filter, Applied Business and Engineering Conference (ABEC), Politeknik Caltex Riau, Riau, hal. 1-7.

[5] Nugroho, Setiawan Qodri., 2015. Pengelompokan Kayu Kelapa Menggunakan Algoritma K-Means Berdasarkan Tekstur Citra Kayu Kelapa Dua Dimensi (2D). Semarang : Universitas Dian Nuswantoro.

[6] Muis, S., 2006. Teknik Jaringan Saraf Tiruan. Yogyakarta: Graha Ilmu

[7] Faradila, 2007. Penyelesaian System Persamaan non Linier dengan metode Jaringan Saraf Tiruan Hopfield. Skripsi. Jakarta : Universitas Negri Hidayatullah.

[8] MacQueen, J. B., 1967. Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1: 281-297.