**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNOLOGI INFORMASI**

**INSTITUT TEKNOLOGI SEPULUH NOPEMBER**

**USULAN TUGAS AKHIR**

# IDENTITAS PENGUSUL

**NAMA : RM Reza Rizky Pratama**

**NRP : 5110100181**

**DOSEN WALI : Victor Hariadi, S.Si., M.Kom.**

**DOSEN PEMBIMBING : 1. Dr. Chastine Fatichah, S.Kom., M.Kom.  
 2. Dr.tech. Ir. R.V.Hari Ginardi, M.Sc.**

# JUDUL TUGAS AKHIR

“Identifikasi Sel Bertumpuk Menggunakan *Unsupervised Bayesian Classification* pada Citra Mikroskopik Sel Darah”

# LATAR BELAKANG

Setiap sel yang ada dalam tubuh manusia memegang peranan penting dalam kesehatan manusia. Jumlah sel merupakan salah satu acuan di dalam dunia kedokteran terhadap jenis penyakit, seperti anemia yang disebabkan kurangnya sel darah merah, leukimia yang disebabkan lebihnya sel darah putih, dan lain-lain. Dalam proses ekstrasi sel secara otomatis, salah satu masalah yang ada adalah ekstrasi dari sel yang bertumpuk karena sel yang bertumpuk dapat mengakibatkan ketidakakuratan perhitungan.

Banyak metode ataupun penelitian yang dikembangkan untuk mengatasi masalah sel bertumpuk. Ada dua macam pendekatan dalam mendeteksi sel bertumpuk yakni *region based* dan *boundary based*. Pendekatan *region based* didasarkan pada ukuran (*size*) atau jumlah piksel pada *region*. Sedangkan *boundary based* didasarkan pada bentuk(*shape*) seperti lengkungan atau bentuk kurva [[1](#Jun081)].

Dalam tugas akhir ini, identifikasi sel bertumpuk menggunakan *unsupervised Bayesian classification* pada citra mikroskopik sel darah, pemisahan sel yang bertumpuk nantinya akan dimodelkan sebagai suatu *cluster*, lalu untuk mengamati *cluster* yang telah terbentuk digunakan *unsupervised Bayesian classifier*. Kemudian, *cluster validity index* digunakan untuk mengestimasi jumlah sel yang bertumpuk [[2](#Jun10)]. Diharapkan dengan metode ini dapat mengidentifikasi sel yang bertumpuk dan membuat kontur garis dari sel yang sebelumnya bertumpuk.

# RUMUSAN MASALAH

Rumusan masalah yang diangkat dalam tugas akhir ini adalah sebagai berikut.

1. Bagaimana mengidentifikasi sel yang bertumpuk dari *input* citra sel darah
2. Bagaimana cara melakukan penghitungan jumlah sel yang bertumpuk dari *input* citra sel darah?
3. Bagaimana cara menentukan garis perpotongan dari sel yang bertumpuk sehingga menjadi sel yang berbeda?
4. Bagaimana cara pembentukan kontur sel dari sel yang bertumpuk berdasarkan garis perpotongan yang ditemukan sebelumnya?

# BATASAN MASALAH

Batasan ruang lingkup permasalahan dari pengerjaan tugas akhir ini adalah sebagai berikut.

1. Implementasi menggunakan Matlab 8.1.
2. *Input* citra yang digunakan berasal dari citra mikroskopik sel darah.

# TUJUAN PEMBUATAN TUGAS AKHIR

Tujuan dari pembuatan tugas akhir ini adalah sebagai berikut.

1. Mengidentifikasi sel yang bertumpuk dari *input* citra sel yang diberikan.
2. Melakukan penghitungan jumlah sel yang bertumpuk dari *input* citra sel yang diberikan.
3. Menentukan garis perpotongan dari sel yang bertumpuk sehingga menjadi sel yang berbeda.
4. Merekonstruksi kontur sel dari sel yang bertumpuk sehingga menjadi sel yang berbeda.

# MANFAAT TUGAS AKHIR

Tugas akhir ini dikerjakan dengan harapan dapat memberikan manfaat pada bidang kesehatan dalam melakukan penghitungan jumlah sel secara otomatis dengan akurasi yang tinggi.

# TINJAUAN PUSTAKA

Pada bab ini akan dibahas tinjauan pustaka yang dipergunakan pada tugas akhir ini yaitu:segmentasi citra, *Gaussian mixture models*, *Expectation–Maximization algorithm*, *Bayesian classification*, dan *linear discriminant analysis*

* 1. Segmentasi citra

Segmentasi citra adalah suatu proses pra-pengolahan pada sistem pengenalan objek dalam citra. Dalam prosesnya dilakukan pembagian dari suatu citra menjadi wilayah-wilayah yang homogen berdasarkan kriteria keserupaan yang tertentu antara tingkat keabuan suatu piksel dengan tingkat keabuan piksel-piksel tetangganya, kemudian hasil dari proses segmentasi ini akan digunakan untuk proses tingkat tinggi lebih lanjut yang dapat dilakukan terhadap suatu citra, misalnya proses klasifikasi citra dan proses identifikasi objek [[3](#Gon02)].

* 1. *Gaussian mixture models*

*Gaussian mixture models* adalah sebuah metode penganalisaan data atau *data mining*. Metode ini mengelompokkan data-data di dalam suatu *dataset* menjadi kelompok-kelompok data yang sebelumnya tidak terdefinisi dan berdasarkan dari distribusi *Gauss* atau distribusi normal. Distribusi normal adalah kurva yang mempunyai dua parameter, parameter pertama adalah μ atau *mean* dan parameter kedua adalah σ atau standar deviasi. Kumpulan distribusi normal biasa disebut dengan *normal probability density function* atau *normal pdf* [[4](#Dou09)].

* 1. *Expectation–maximization algorithm*

*Expectation-maximization* (EM) *algorithm* adalah sebuah algoritma untuk mengestimasi suatu nilai di dalam model dari suatu variabel laten yang tidak diketahui, algoritma EM merupakan iterasi yang melakukan penggabungan dari dua proses yaitu proses *expectation* (E) fungsi untuk mengekspetasi nilai parameter yang dianggap memiliki nilai logaritma kemungkinan (*log-likelihood*) yang tinggi dan *maximization* (M) di mana parameter *log-likelihood* yang didapatkan dari proses E dapat dimaksimalkan. Kemudian parameter yang dihasilkan dari proses M dapat digunakan untuk menentukan distibusi dari variabel laten untuk proses E selanjutnya [[5](#Sin05)].

* 1. *Bayesian classification*

*Bayesian classification* adalah klasifikasi statistik yang dapat memprediksi kelas suatu anggota probabilitas. Untuk klasifikasi Bayes sederhana yang lebih dikenal sebagai *naïve Bayesian classifier* dapat diasumsikan bahwa efek dari suatu nilai atribut sebuah kelas yang diberikan adalah bebas dari atribut-atribut lain. Asumsi ini disebut *class conditional independence* yang dibuat untuk memudahkan perhitungan [[6](#Dud00)].

* 1. *Linear discriminant analysis*

*Linear discriminant analysis* adalah suatu teknik yang digunakan untuk memisahkan atau mengkategorikan suatu kumpulan data atau objek menjadi dua buah kelas atau lebih. Proses ini digunakan sebelum melakukan proses klasifikasi [[7](#Tab05)].

# RINGKASAN ISI TUGAS AKHIR

Tugas akhir ini mengenai identifikasi sel bertumpuk dengan menggunakan *unsupervised Bayesian classification.* Bagan proses utama dari proses identifikasi sel bertumpuk ditunjukan pada Gambar 1.

*Input* citra mikroskopik sel darah

*Preprocessing*(1)

Segmentasi citra dan identifikasi sel bertumpuk

(2)

Rekonstruksi kontur

(3)

*Output*

Gambar 1. Tahapan pengerjaan identifikasi sel bertumpuk

Pada tahap 1, *preprocessing* dilakukan tiga proses yaitu *sequential filtering*, *adaptive thresholding*, dan *distance transform*. Tiga proses tersebut dilakukan jika *input* citra memiliki sel yang bertumpuk, maka dapat dipastikan hasil dari proses pada *distance transform* akan berupa *Gaussian mixtures models*. Kemudian dilanjutkan pada proses selanjutnya.

Pada tahap 2, segmentasi dan identifikasi sel bertumpuk yaitu proses dilakukannya segmentasi terhadap citra kemudian dilakukan identifikasi sel bertumpuk. Terdapat tiga proses yaitu: algoritma EM, *cluster validation*, dan *separation line decision*. Proses algoritma EM dilakukan untuk mengestimasi *cluster* yang terbentuk dari sel yang bertumpuk.

Proses *cluster validation* dilakukan untuk menghitung jumlah sel yang bertumpuk dari *cluster* yang telah terbentuk. Pada *cluster validation* terdapat *cluster validity index* [[8](#Bou06)].

(1)

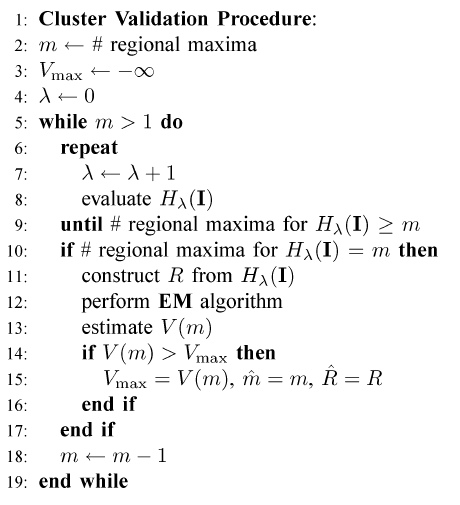
Pada Persamaan 1, adalah *cluster validity index* yang didapat dari penghitungan dari yaitu *separation* dibagi dengan yaitu *compactness*. Kemudian, terdapat *H-maxima transform* yang digunakan untuk menghilangkan *regional maxima* yang kurang signifikan, namun tetap menyimpan *regional maxima* yang signifikan [[9](#Soi99)].

(2)

Persamaan 2 adalah *H-maxima transform*, di mana I adalah *cluster validity index*, λ adalah *depth*, dan Φ adalah *reconstruction* *dilation operators*.

Nilai dari *regional maxima* didapat dari *distance image* sel yang bertumpuk. *Cluster validation* dilakukan terus menerus hingga nilai dari sama dengan dua. Variabel adalah nilai dari *cluster validity index* yang telah dilakukan maksimasi menggunakan proses M seperti pada Persamaan 3.

(3)



Gambar 2. *Pseudocode* dari *cluster validation*

Pengerjaan *cluster validation* melibatkan algoritma EM dan *H-maxima transform* yang telah dijelaskan sebelumnya. Gambar 2 menunjukan *pseudocode* untuk *cluster validation.*

Tahap selanjutnya adalah proses *separation line decision*, yaitu proses pembentukan perpotongan garis antara sel yang bertumpuk. Tujuan dari pembentukan perpotongan garis adalah untuk membuat batasan dari *cluster* yang terbentuk pada proses sebelumnya, kemudian untuk membuat titik perpotongan yang akan menjadi acuan untuk rekonstruksi kontur dari sel yang bertumpuk tersebut.

Pada proses *separation line decision*, digunakan metode *linear discriminant analysis* seperti pada Persamaan 4.

(4)

Tahap 3 adalah rekonstruksi kontur, yaitu proses pembentukan kontur dari sel yang bertumpuk tersebut berdasarkan titik yang ditemukan pada tahap *separation line decision*. Bentuk dari kontur dari titik yang ditemukan nanti akan berbentuk elips sehingga digunakan pembentukan elips. Persamaan 5 menunjukan persamaan dari pembentukan elips.

(5)

Di mana *a, b, c, d, e,* dan *f* merepresentasikan koefisien dari elips. *F(p, q)* dapat ditampilkan dalam bentuk vektor seperti pada Persamaan (6) dengan mendefinisikan dan .

{6}

Setelah melalui berbagai tahapan diharapkan *output* yang dihasilkan berupa citra sel darah yang baru dengan kontur yang direkonstruksi dari sel yang bertumpuk dan jumlah dari sel darah yang terdapat pada citra hasil *output*.

# METODOLOGI

## Penyusunan proposal tugas akhir

Tahap awal untuk memulai pengerjaan tugas akhir adalah penyusunan proposal tugas akhir. Pada proposal ini, penulis melakukan implementasi identifikasi sel bertumpuk menggunakan *unsupervised Bayesian classification* pada citra mikroskopik sel darah.

## Studi literatur

Pada tahap ini dilakukan pencarian informasi dan studi literatur yang diperlukan untuk pengumpulan data dan desain sistem yang akan dibuat. Referensi tersebut berisikan tentang:

1. Konsep mengenai *Expectation–Maximization algorithm.*
2. Konsep mengenai *Gaussian mixture models.*
3. Konsep mengenai *Bayessian classifier.*

Referensi ini dapat dicari dari buku, jurnal, artikel, laporan penelitian, dan situs-situs di internet. Studi literatur ini adalah terkoleksinya referensi yang relevan dengan perumusan masalah. Tujuannya untuk memperkuat permasalahan serta sebagai dasar teori dalam melakukan studi dan juga menjadi dasar untuk melakukan identifikasi sel bertumpuk menggunakan *unsupervised Bayesian classification* pada citra mikroskopik sel darah.

## Analisis dan desain perangkat lunak

Implementasi merupakan tahap untuk membangun algoritma tersebut. Untuk membangun algoritma yang telah dirancang sebelumnya, diimplementasikan dengan menggunakan Matlab.

## Pengujian dan evaluasi

Pada tahap ini dilakukan uji coba dengan menggunakan 10 gambar untuk mencoba jalannya aplikasi apakah telah sesuai dengan rancangan dan desain implementasi yang dibuat. Pengujian hasil dilakukan dengan membandingkan hasil dari sistem secara otomatis dengan hasil penghitungan sel secara manual. Penghitungan akurasi dilakukan dengan 2 cara yaitu:

1. Menghitung akurasi dengan *precision* dan *recall* dari hasil yang didapat.
2. Menghitung akurasi dengan mempertimbangkan *undersegmented* dan *oversegmented* dari hasil yang didapat.

## Penyusunan Buku Tugas Akhir

Pada tahap ini dilakukan penyusunan laporan yang menjelaskan dasar teori dan metode yang digunakan dalam tugas akhir ini serta hasil dari implementasi aplikasi perangkat lunak yang telah dibuat. Sistematika penulisan buku Tugas Akhir secara garis besar antara lain:

1. Pendahuluan
   1. Latar Belakang
   2. Rumusan Masalah
   3. Batasan Tugas Akhir
   4. Tujuan
   5. Metodologi
   6. Sistematika Penulisan
2. Tinjauan Pustaka
3. Desain dan Implementasi
4. Pengujian dan Evaluasi
5. Kesimpulan dan Saran
6. Daftar Pustaka

# JADWAL KEGIATAN

Alur pengerjaan tugas akhir ditunjukan pada Tabel 1.

Tabel 1. Alur Pengerjaan Tugas Akhir

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tahapan | 2014 | | | | | | | | | | | | | | | | |
| Maret | | | | April | | | | Mei | | | | | Juni | | | |
| Penyusunan Proposal |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Studi Literatur |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Perancangan sistem |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Implementasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Pengujian dan evaluasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Penyusunan buku |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# DAFTAR PUSTAKA

x

|  |  |
| --- | --- |
| [1] | M Jung, J Shim, B Ko, and J Nam, "Automatic Cell Segmentation and Classification using Morphological Features and Bayesian Network," Image Processing : Machine Vision Applications, Proc. Of SPIE-IS&T Electronic Imaging, vol. 6813, San Jose, 2008. |
| [2] | C Jung, Kim C, S W Chae, and S Oh, "Unsupervised Segmentation of Overlapped Nuclei Using Bayesian Classification," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2825-2832, December 2010. |
| [3] | Rafael C Gonzales and Richard E Woods, *Digital Image Processing*, 2nd ed.: Prentice Hall, 2002. |
| [4] | A Reynolds Douglas, "Gaussian Mixture Models," in *Encyclopedia of Biometrics*.: Springer, 2009, pp. 659-663. |
| [5] | Ajit P Singh. "The EM Algorithm," November 2005. [Online]. Available: http://www.cs.cmu.edu/~awm/15781/assignments/EM.pdf . [Diakses 8 Maret 2014]. |
| [6] | Richard O Duda, Peter E Hart, and David G Stork, *Pattern Classification*, 2nd ed.: Wiley-Intersceince, 2000. |
| [7] | Barbara G Tabachnick and Linda S Fidell, *Using Multivariate Statistics*, 5th ed.: Pearson, 2005. |
| [8] | M Bouguessa, S Wang, and H Sun, "An objective approach to cluster validation," *Pattern Recognition Letters*, vol. 27, no. 13, pp. 1419–1430, 2006. |
| [9] | P Soile, *Morphological Image Analysis: Principles and Applications*, 2nd ed. Berlin, Germany: Springer, 2004. |

x