

Identifikasi Area Kanker Ovarium pada Citra CT Scan Abdomen Menggunakan Metode Expectation Maximization

Lestari Handayani

Jurusan Teknik Informatika
UIN Sultan Syarif Kasim Riau
lestari.handayani@uin-suska.ac.id

Abstrak

Kanker ovarium merupakan penyakit mematikan, karena pasien terlambat untuk menyadari penyakit ini dan terlambat datang untuk pengobatan. Untuk mengetahui kondisi penderita dilakukan pemeriksaan seperti USG dengan Doppler, CT Scan atau MRI. Semua pemeriksaan di atas belum bisa memastikan diagnosis kanker ovarium, akan tetapi hanya sebagai pegangan tindakan operasi. Untuk itu diperlukan suatu sistem yang mampu menganalisa kondisi kanker tersebut. Satu tahap dalam proses tersebut adalah mengidentifikasi area kanker. Pada penelitian ini digunakan data pemeriksaan kanker berupa citra hasil CT Scan Abdomen. Metode yang digunakan yaitu Expectation Maximization dengan model Gaussian Mixture Model (EM-GMM) untuk segmentasi area kanker ovarium. Dari hasil eksperimen diperoleh hasil bahwa metode EM-GMM dapat membagi citra menjadi beberapa kelas berdasarkan fitur piksel, namun masih belum bagus membedakan bagian kanker dengan bagian bukan kanker. Hal ini terlihat dari hasil perhitungan persentase piksel yang diperkirakan kanker atau bukan, diperoleh nilai TP 45 %, sedangkan FP 55%. Hal ini dikarenakan piksel di bagian yang diperkirakan kanker, juga berada di bagian lain yang bukan kanker. Untuk memperbaikinya disarankan ditambahkan satu fitur misalnya fitur bentuk.

Kata kunci: CT Scan Abdomen, Expectation Maximization, Gaussian Mixture Model, Kanker Ovarium.

Abstract

Ovarian Cancer is a deadly disease, because the patient is too late to be aware of this disease and come late to treatment. To detect the condition of patient, it's need examination such as USG with Doppler, CT scan abdomen or MRI. The examination cannot used to diagnose ovarian cancer, but only to do operation. Therefore, we need systems to analyze of this condition. One part of the systems is how to identify area of cancer. In this paper, we use image from ct scan examination result. The method Expectation Maximization with Gaussian Mixture Model (EM GMM) is used to segmentation of ovarian cancer areas. The experiment result is EM-GMM method can separate image into some classification based on pixel feature, even though not so good to distinguish area of cancer and not cancer. It's seen from the results of calculation of the percentage of pixels that estimated cancer or not, the value of TP(True Positive) is 45%, while FP(False Positive) is 55%. It caused both of them are same in pixel value. To improve the result, we need another feature to segmentation, for example is shape feature.

Keywords: CT scan Abdomen, Expectation Maximization, Gaussian Mixture Model, Ovarian Cancer.

1. Pendahuluan

Penyakit kanker adalah penyakit yang mematikan nomor 1 di dunia. Menurut Cancer Institute, jumlah penderita kanker di dunia terus bertambah 6,25 juta orang setiap tahunnya. Dua per tiga dari penderita kanker di dunia berada di Negara-negara berkembang, termasuk Indonesia. Dari data Departemen Kesehatan menunjukkan jumlah penderita kanker di Indonesia mencapai 6% dari populasi Indonesia.

Kanker ovarium merupakan penyakit mematikan kedua pada wanita, setelah kanker serviks. Tingkat kematian yang tinggi disebabkan karena penyakit ini pada awalnya bersifat asimtomatik dan baru menimbulkan keluhan apabila sudah terjadi metastasis. Dimana 60%-70% pasien datang berobat pada stadium lanjut, sehingga terlambat saat dilakukan pengobatan karena kanker ovarium telah mengganas dan menyebar ke organ lain.

Kanker ovarium sebagian besar berbentuk tumor kista dan sebagai kecil berbentuk tumor padat. Apabila ditemukan kista pada seorang wanita, harus dilakukan pemeriksaan lebih lanjut untuk menentukan apakah kista tersebut bersifat jinak atau ganas (kanker ovarium). Untuk memperkuat dugaan dilakukan pemeriksaan seperti tindakan USG dengan Doppler untuk menentukan arus darah dan bahkan diperlukan pemeriksaan CT-Scan/MRI. Pemeriksaan laboratorium yang bisa dilakukan untuk menunjang diagnosis adalah pemeriksaan tumor marker seperti Ca-125 dan Ca 72-4, beta-HCG dan alfafetoprotein.

Semua pemeriksaan di atas belum bisa memastikan diagnosis kanker ovarium, akan tetapi hanya sebagai pegangan tindakan operasi. Untuk itu diperlukan suatu sistem yang mampu menganalisa kondisi kanker tersebut. Satu tahap dalam proses tersebut adalah mengidentifikasi area kanker. Beberapa penelitian yang telah dilakukan untuk membantu diagnosa penyakit kanker yaitu Campadelli [2],[3], dan Voiculescu [10]. Pada penelitian ini, kali ini diusulkan metode Expectation Maximization untuk membantu menganalisa area kanker ovarium. Karena menurut Carson[4], Kokkinos[5] dengan menggunakan metode ini didapatkan hasil yang maksimum.

Pada laporan penelitian ini akan dijelaskan secara singkat mengenai metode Expectation Maximization pada bab 2. Kemudian pada bab 3 diuraikan proses analisa area kanker ovarium menggunakan Expectation Maximization dengan Gaussian Mixture Model(EM-GMM), selanjutnya dilakukan eksperimen dengan hasil eksperimen dipaparkan pada sub bab 3.2., terakhir kesimpulan dari penelitian ini di bab 4.

2. Metode Expectation Maximization

Expectation Maximization Segmentation (EM) adalah salah satu algoritma yang digunakan dalam *machine learning*. Algoritma EM secara luas digunakan untuk memperkirakan parameter-parameter model menggunakan data yang tidak berlabel. EM termasuk algoritma clustering yang masuk dalam kategori *partitional clustering*. EM berdasarkan model yang menggunakan perhitungan probabilitas. Metode iteratif tersebut akan menghasilkan *Maximum Likelihood* (ML), yang menghasilkan parameter baru, yaitu bobot *mixture*, *mean*, dan kovarian atau standard deviasi. EM adalah formula umum dari problem seperti "Chicken+Egg"(Mix.Gaussians, Mix.Experts, Neural Nets,HMMs, Bayes-Nets,...). Isodata adalah contoh spesifik dari EM. General EM for mix.Gaussian disebut Soft-Clustering yang dapat konvergen menjadi *Maximum Likelihood*. EM terdiri dari dua tahap yaitu *Expectation* dan *Maximization*.

E-step untuk mencari parameter yang sama, pada tahap ini juga dilakukan perhitungan *expected values* dari parameter menggunakan hipotesis, kemudian **M-step** untuk menghitung nilai *mean* (rata-rata) dan perbedaan yang digunakan untuk mengestimasi ulang parameter dan dilakukan secara berulang-ulang hingga mencapai local maksimum. **M-step** juga digunakan untuk menghitung nilai hipotesis maximum *likelihood* dengan mengasumsikan parameter yang sama dengan *expected value* dari tahap estimasi. Kedua tahap tersebut dilakukan berulang-ulang sampai hipotesa dari converge(nilai yang terpusat) mencapai nilai yang stationer. Untuk model algoritma EM, setiap cluster memiliki *distribution probability* (kemungkinan penyebaran) yang sama dan untuk setiap kejadian data digunakan parameter nilai estimate pada setiap distribution.

Secara singkat algoritma EM dapat dilihat sebagai berikut:

- Step1** : Menentukan jumlah k cluster terhadap objek.
- Step2** : Sebelum data dikelompokkan dalam k cluster, terlebih dahulu ditentukan nilai *mean*, standard deviasi dan probabilitas terhadap objek.
- Step3** : Dilakukan tahap ekspektasi, yaitu berdasarkan nilai mean dan standard deviasi akan dicari *probability* untuk setiap objek terhadap k cluster dengan menggunakan *probability density function* (pdf).

$$y = f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

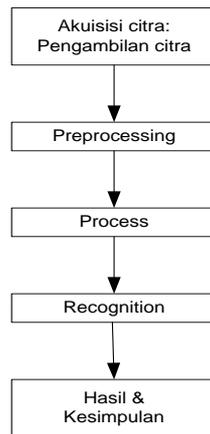
- Step4** : Dilakukan tahap *maximization*, yaitu berdasarkan nilai *probability* setiap objek pada step 2, akan di hitung kembali *mean*, standard deviasi dan probabilitas baru.
- Step5** : Jika selisih antara probabilitas lama dengan probabilitas baru yang didapat lebih besar dari nilai yang ditoleransi maka dilakukan kembali step 2, namun jika tidak maka iterasi berhenti dan akan didapat hasil cluster.

3. Analisa dan Hasil

Pada bab ini diuraikan proses analisa area kanker ovarium menggunakan metode segmentasi Expectation Maximization dan dipaparkan hasil eksperimen.

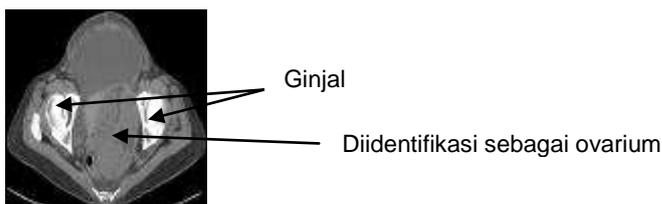
3.1. Analisa Area Kanker Ovarium menggunakan EM-GMM

Adapun tahap-tahap yang akan dilakukan untuk segmentasi citra pada citra CT Scan Abdomen dengan metode algoritma *Expectation Maximization(EM)* untuk identifikasi area kanker ovarium dapat dilihat dari diagram alir dibawah ini:



Gambar 1. Flowchart metode EM

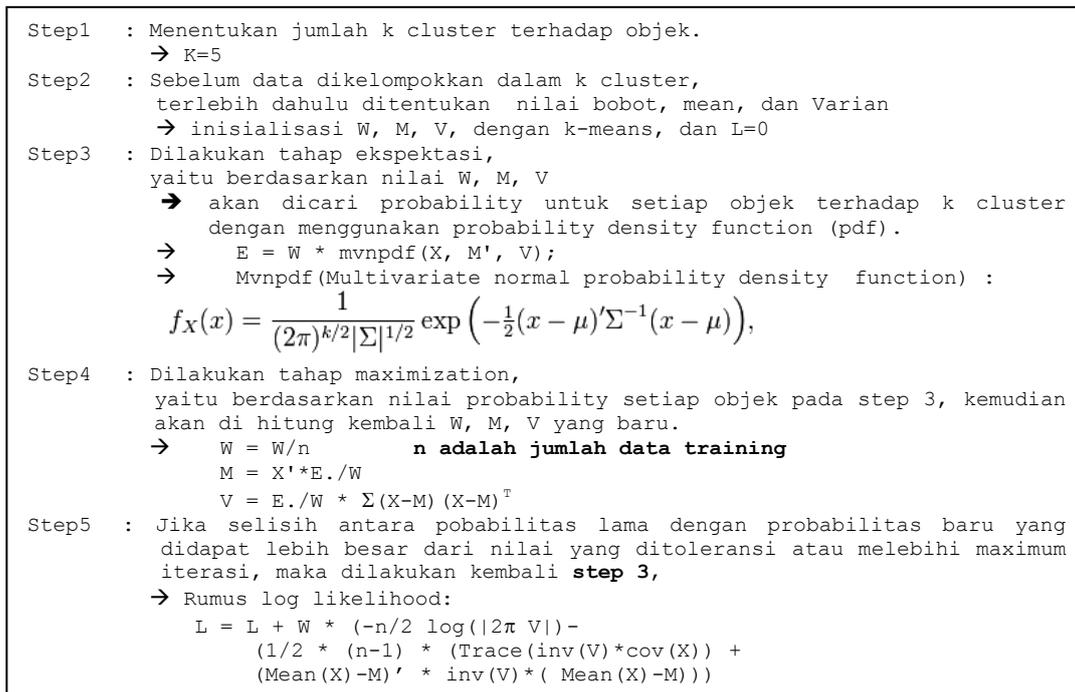
Pada tahap akuisisi citra dilakukan pengambilan data dari alat CT Scan. Data yang digunakan untuk analisa kanker ovarium yaitu citra CT scan abdomen. Dari hasil CT-Scan Abdomen diperoleh beberapa *slice* citra yang jumlahnya tergantung pada tinggi pasien yaitu antara 35-46 *slice*. Gambar yang bisa digunakan untuk melihat adanya kanker ovarium yang berada di daerah perut bagian bawah, yaitu citra ke 37-46. Contoh gambar hasil citra CT Scan Abdomen yang terdapat ovarium dapat dilihat pada gambar 2. Pada gambar 2 ini ovarium ditunjuk dengan panah, terlihat pada gambar bahwa ovarium sudah membesar dan menekan ginjal kiri, kanan disebelahnya.



Gambar 2. Citra 1 CT Scan Abdomen

Data yang diperoleh dari alat CT scan sudah sangat bagus, tidak perlu banyak preprocessing. Namun, untuk masuk ke proses segmentasi, maka data di potong pada bagian yang menggambarkan area kanker dan ukurannya dikecilkan menjadi 100x100 piksel karena proses cukup memakan waktu yang lama.

Agar program dapat memperlihatkan hasil segmentasi pada citra, maka dilakukan *recoding* dimana tahapan-tahapan yang dilakukan terlihat pada gambar 3. Pada algoritma ini saya menggunakan pendekatan Gaussian Mixture Model sehingga diberi nama EM-GMM. Sedangkan pada tahap ekspektasi, digunakan Multivariate normal probability density function, agar hasil segmentasi lebih bagus.

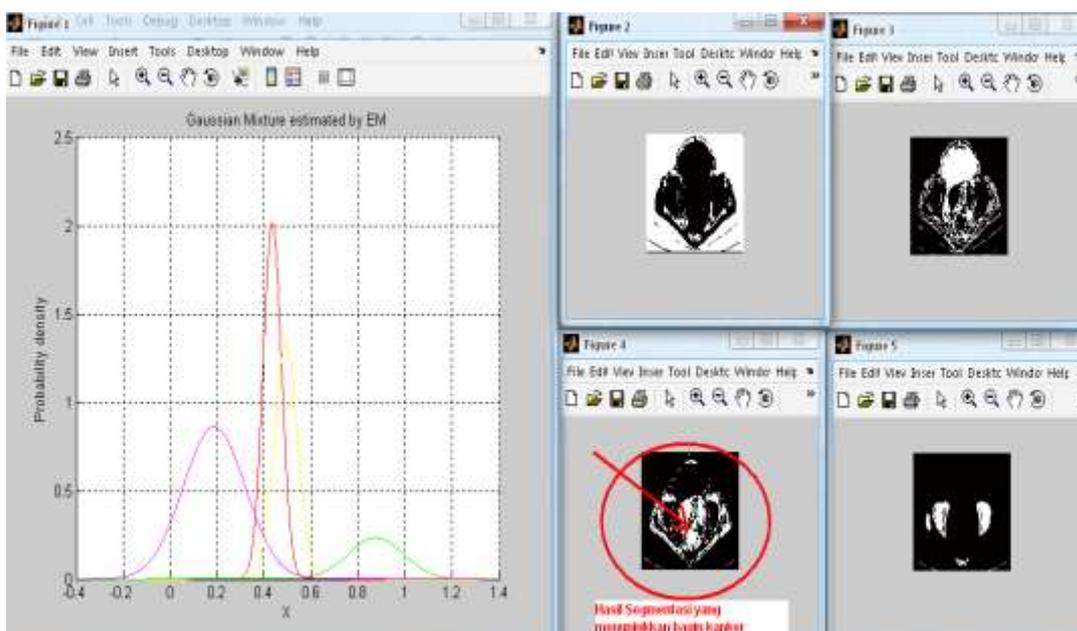


Gambar 3. Algoritma EM-GMM untuk identifikasi area kanker ovarium

Terakhir dilakukan tahap recognition atau pengenalan area kanker ovarium. Pada tahap ini dibutuhkan bantuan dari radiologist yang ahli membaca hasil citra ct scan. Dari hasil diskusi dengan radiologist, diperoleh bahwa ovarium pada normalnya berukuran sangat kecil dan berada di area tengah bagian perut bawah. Untuk kasus penderita kanker, area kanker ovarium ditandai dengan area perut bawah yang membengkak, dan terkadang menekan organ lain di sekitarnya. Dari hasil program, dicari bagian hasil segmentasi yang menggambarkan area kanker ovarium dengan melihat grafik probability density tiap kelas.

3.2. Hasil Eksperimen

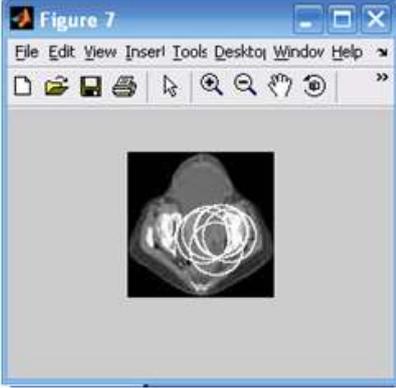
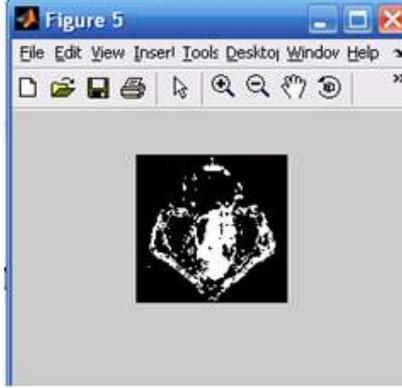
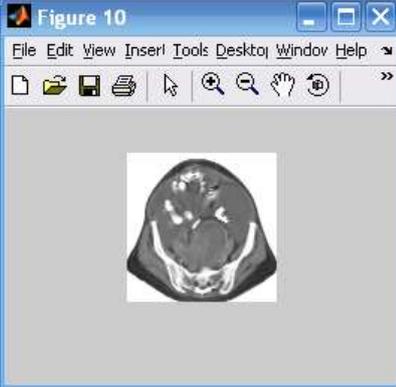
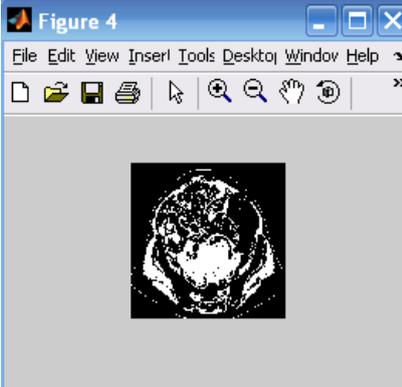
Eksperimen dilakukan menggunakan data dari CT Scan penderita kanker ovarium yang diperoleh dari RS. Kanker Dharmais. Jumlah data tidak banyak, hanya 5 citra. Hasil eksperimen algoritma EM-GMM untuk identifikasi area kanker ovarium terlihat pada gambar 4 di bawah ini.

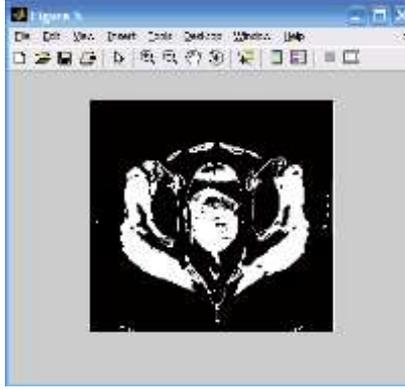
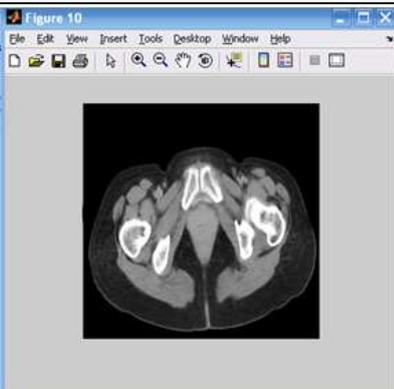
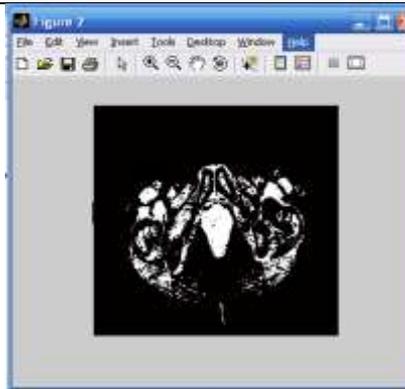


Gambar 4. Hasil segmentasi citra1

Pada gambar 4 di atas dimunculkan grafik Gaussian Mixture yang diperoleh dari metode EM. Diperoleh ada 4 garis warna yang melambangkan kelas-kelas hasil segmentasi EM. Dari hasil pengamatan beberapa eksperimen didapat bahwa kelas yang diduga area kanker ovarium adalah kelas ke tiga yang berwarna hijau, kelas ketiga digambarkan pada figure 4 seperti terlihat pada gambar 4 di atas. Pada kelas ketiga tersebut ternyata hasil segmentasi belum bagus, Untuk lebih rinci, dilakukan perhitungan persentase piksel yang benar kanker (TP/ *True Positive*) dan perhitungan persentase piksel yang bukan kanker (FP/*False Positive*). Hasil eksperimen dan perhitungan terpapar di tabel 1 hasil eksperimen.

Tabel 1. Hasil Eksperimen

Nama	Citra Awal	Citra Hasil Segmentasi	%TP	%FP
Citra1			68	32
Citra2			34	66
Citra3			46	54

Citra4			35	65
Citra5			42	58
Rata-Rata			45	55

Terdapat beberapa fenomena yang cukup menarik untuk diamati dalam proses eksperimen yang dilakukan. Yaitu bahwa ternyata pada hasil segmentasi yang dihasilkan dilihat bahwa bagian kanker ovarium masih bercampur dengan bagian lain yang berupa bukan kanker (Lihat Gambar 4) dengan hasil rata-rata TP 45%. Hal ini dapat terjadi karena nilai piksel yang diduga kanker sama dengan nilai di sekitar yang bukan kanker, sehingga perlakuan rumus apapun di bagian ini akan tetap mengelompok menjadi satu kelas. Terkecuali jika di dalam rumus ada bagian yang membedakan posisi piksel. Meskipun program di recoding kembali untuk mempertimbangkan posisi piksel dengan menggunakan algoritma **K-means** (default MatLab) dan **Fuzzy C-means**, namun ternyata hasil yang didapatkan masih belum sempurna, bahkan tidak lebih baik.

Terdapat satu fenomena lagi, bahwa piksel yang diduga kanker memiliki nilai rentang antara 83-140. Dan dari hasil percobaan, beberapa piksel yang diluar rentang tersebut masih dikelompokkan ke dalam satu kelas yang diduga kanker. Nilai piksel yang diluar rentang tersebut dapat dikurangi (tidak memasukkannya ke dalam kelas yang diduga kanker), dengan cara :

1. Melihat Invarian dari tiap kelas
2. Melihat probabilitas piksel terhadap tiap kelas
3. Nilai Probabilitas antara 83-140 diatas 1, sementara pixel di bawah rentang 83 nilainya dibawah 1
4. Memperbaiki nilai Mean dan Varian kelas yang diduga kanker

Hasil pengamatan menarik lainnya yang didapatkan adalah bahwa daerah yang diduga kanker ovarium cenderung berbentuk oval dan berada digaris tengah horizontal. Untuk memperbaiki hasil segmentasi, bisa dilakukan dengan membuat satu fitur yang menunjukkan posisi piksel di tengah, kiri atau kanan. Jika piksel tersebut berada di area tengah, maka bisa diduga kanker, demikian sebaliknya.

4. Kesimpulan

Dari hasil eksperimen diperoleh hasil bahwa metode EM-GMM dapat membagi citra menjadi beberapa kelas berdasarkan fitur piksel, namun masih belum bagus membedakan bagian kanker dengan bagian bukan kanker. Hal ini terlihat dari hasil perhitungan persentase piksel yang diperkirakan kanker atau bukan, diperoleh nilai TP 45 %, sedangkan FP 55%. Hal ini dikarenakan piksel di bagian yang diperkirakan kanker juga berada di bagian lain yang bukan kanker. Untuk memperbaikinya disarankan ditambahkan satu fitur misalnya fitur shape.

Referensi

- [1] Borman, Sean. The Expectation Maximization Algorithm A short tutorial. <http://seanborman.com>.(10-2009)
- [2] Campadelli P, Casiraghi E, Lombardi G. Automatic Liver Segmentation From Abdominal CT Scans. 14th International Conference on Image Analysis and Processing, IEEE. 2007.
- [3] Campadelli P, Casiraghi E, Lombardi G. Automatic Segmentation Of Abdominal Organs From CT Scans. 19th IEEE International Conference on Tools with Artificial Intelligence. 2007; 513-516.
- [4] Carson C, Belongie S, Greenspan H, Malik J. Blobworld: Image segmentation using Expectation-Maximization and its application to image querying. IEEE Transactions On Pattern Analysis And Machine Intelligence. 2002; 24(8): 1026-1038.
- [5] Kokkinos I, Maragos P. An Expectation Maximization Approach to the Synergy Between Image Segmentation and Object Categorization. CiteSeerX, ICCV. 2005.
- [6] Movellan, Javier R., Tutorial on Generalized Expectation Maximization
- [7] Piater, Justus H., *Mixture Models and Expectation-Maximization*, Lecture at ENSIMAG, May 2002 revised 21 November 2005
- [8] Priyanto, Agus., pendugaan parameter model faktor dengan menggunakan metode maksimum *likelihood*, 2008.
- [9] Rennie, Jason D. M., A Short Tutorial on Using Expectation Maximization with Mixture Models, March 3, 2004
- [10] Voiculescu I, Golodetz S, Cameron S. Progress on a Decision-Support System for Abdominal CT Scans. IEEE. 2009; 116-119.