

# IMAGE CLUSTER BERDASARKAN WARNA UNTUK IDENTIFIKASI KEMATANGAN BUAH TOMAT DENGAN METODE VALLEY TRACING

M. Helmy Noor<sup>1</sup>, Moch. Hariadi<sup>2</sup>

Program Pascasarjana, Jurusan Teknik Elektro, Program Studi Jaringan Cerdas Multimedia  
Fakultas Teknologi Industri, Institut Teknologi Sepuluh Nopember Surabaya  
Kampus ITS, Jl. Raya ITS, Sukolilo-Surabaya 60111, Indonesia, Telp. 031-5939214, Fax. 031-5939363  
Email :m\_miplbn@yahoo.com

## Abstrak

Ciri yang digunakan dalam identifikasi kematangan buah adalah ciri warna (fitur R, G, dan B). Selanjutnya dilakukan clustering dengan metode Single Linkage Hierarchical Method (SLHM) terhadap ciri warna yang diperoleh. Dalam clustering, umumnya harus dilakukan inisialisasi jumlah cluster yang diinginkan terlebih dahulu, padahal pada beberapa kasus clustering, user bahkan tidak tahu berapa banyak cluster yang bisa dibangun. Untuk itu, dalam penelitian ini diaplikasikan metode Valley Tracing. Metode ini merupakan constraint yang akan melakukan identifikasi terhadap pergerakan variance dari tiap tahap pembentukan cluster, dan menganalisa polanya untuk membentuk suatu cluster secara otomatis (automatic clustering). Jumlah cluster yang diperoleh menunjukkan jumlah buah yang diidentifikasi, kemudian nama buah dan jenis kematangan masing-masing buah diperoleh dengan membandingkan nilai centroid tiap cluster dengan nilai centroid data training yang sebelumnya telah disimpan dalam database dan mempunyai label nama buah.

**Kata kunci** : Pengenalan kematangan buah, SLHM, Image Clustering, Clustering Otomatis, Valley Tracing.

## 1. PENDAHULUAN

### 1.1. Latar Belakang

Dewasa ini perkembangan teknologi informasi berkembang sangat pesat. Hal ini diikuti pula dengan banyaknya penelitian-penelitian baru dalam bidang tersebut, diantaranya adalah yang berkaitan dengan *object recognition* (pengenalan objek). Aplikasi *object recognition* yang telah ada salah satunya adalah proses identifikasi / pengenalan buah berdasarkan ciri warna. Pada pengenalan buah berdasarkan ciri warna sebelumnya, sistem hanya mampu mengidentifikasi objek buah tunggal. Sistem yang dibangun belum mampu mengenali objek berupa sekumpulan buah. Sehingga dibutuhkan suatu solusi untuk permasalahan tersebut.

Dalam penelitian ini akan dibangun aplikasi *image clustering* untuk mengidentifikasi buah tomat berdasarkan warna. Dalam *clustering*, umumnya harus dilakukan inisialisasi jumlah cluster yang diinginkan terlebih dahulu, padahal pada beberapa kasus clustering, user bahkan tidak tahu berapa banyak cluster yang bisa dibangun. Untuk itu, dalam penelitian ini diaplikasikan metode *Valley Tracing* yang mampu menyelesaikan masalah tersebut. Metode ini bisa melakukan *clustering* secara otomatis (*Automatic Clustering*) terhadap fitur warna berupa R,G,B dari buah tomat dengan memanfaatkan *Single Linkage Hierarchical Method* (SLHM), yaitu mendeteksi pergerakan varian pada tiap tahap pembentukan clusternya untuk menemukan *global optimum* sehingga bisa dibangun cluster secara otomatis (*automatic Clustering*). Nilai centroid masing-masing hasil *clustering* selanjutnya akan digunakan untuk mengidentifikasi jenis buah dan kematangannya.

### 1.2. Perumusan Masalah

Permasalahan utama pada penelitian ini bagaimana membangun sistem identifikasi buah tomat, dari permasalahan ini akan dibahas 3 permasalahan yang penting yaitu :

- Bagaimana mengidentifikasi berapa jenis kematangan buah tomat dalam gambar input.
- Bagaimana mengidentifikasi jenis buah dan jenis kematangannya yang ada pada gambar input.
- Bagaimana menerapkan metode *Valley Tracing* untuk membangun *cluster* secara otomatis (*Automatic Clustering*).

### 1.3. Tujuan Penelitian

Dari rumusan masalah di atas maka penelitian ini bertujuan untuk memperbaiki teknik retrieval image menggunakan clustering untuk menentukan jumlah buah dan kematangannya yang ada dalam suatu image.

### 1.4. Batasan Masalah

Batasan pada penelitian ini adalah sebagai berikut :

1. Objek yang digunakan adalah buah tomat.
2. Buah tomat yang diidentifikasi adalah buah yang memiliki 1 warna dan dominan.
3. Fitur yang diambil dalam objek gambar adalah cirri warna dengan fitur R,G,B yang kemudian diolah dalam

metode clustering.

4. Metode clustering yang digunakan adalah Single Linkage Hirarchical Method.
5. Ukuran image atau gambar sama dengan format JPG
6. Warna dasar tempat meletakkan objek tetap (background putih).

### 1.5. Manfaat

Manfaat yang dapat diberikan dalam penelitian ini adalah mengembangkan teknik-teknik pengenalan obyek yang telah ada sebelumnya melalui image clustering dengan metode Valley Tracing sehingga kelemahan-kelemahan dalam penelitian-penelitian yang dibuat sebelumnya dapat disempurnakan terutama dalam pengenalan obyek yang memiliki karakteristik spesifik, yang sebelumnya belum dapat diidentifikasi.

## 2. TINJAUAN PUSTAKA

S. Ray, R.H. Turi, (1999), [1], melakukan penelitian tentang kekurangan utama algoritma K-Mean adalah bahwa jumlah cluster harus yang disediakan sebagai parameter.

Pada penelitian mereka memperlihatkan suatu validitas pengukuran sederhana berdasarkan pada pengukuran jarak intra-cluster dan inter-cluster yang membolehkan jumlah cluster untuk ditentukan secara otomatis.

Prosedur dasar melibatkan semua segmentasi image untuk lebih dari 2 cluster sampai ke Kmax cluster, di mana Kmax merepresentasikan batas atas pada jumlah dari cluster. Kemudian pengukuran validitas dihitung untuk menentukan cluster terbaik dengan mencari nilai minimum dari pengukuran

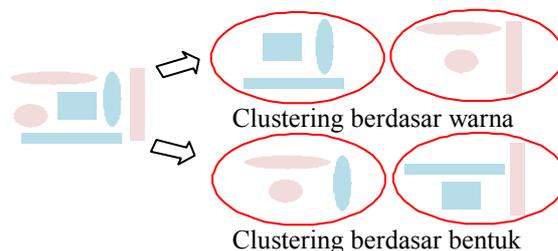
Ali Ridho Barakhbah, Kohei Arai, (2004), [2], penelitian ini mengusulkan pendekatan baru untuk mencari global optimum dari clustering dan menganalisa pergerakan variance dari cluster untuk setiap tingkat pembentukan cluster.

Kemudian menganalisa pola untuk mencari global optimum untuk menghindari local optima. Penelitian ini juga menganalisa kemungkinan untuk membuat cluster otomatis (automatic clustering).

### 2.1. Dasar Teori

#### 2.2.1 Pengertian Clustering

*Clustering* adalah suatu cara menganalisa data dengan cara mengelompokkan objek kedalam kelompok-kelompok berdasar suatu kesamaan tertentu. Bisa juga diartikan sebagai proses untuk mendefinisikan pemetaan/mapping  $f: D \rightarrow C$  dari beberapa data  $D = \{t_1, t_2, \dots, t_n\}$  kedalam beberapa cluster  $C = \{c_1, c_2, \dots, c_n\}$  berdasarkan kesamaan antar  $t_i$ . Sebuah *cluster* adalah sekumpulan obyek yang digabung bersama karena persamaan atau kedekatannya. Sedangkan *Image clustering* (pengelompokkan gambar) adalah proses untuk membagi atau mengelompokkan suatu *image* ke dalam beberapa bagian yang berbeda, dimana pada tiap bagian yang berbeda tersebut anggotanya mempunyai kesamaan khusus (homogen). *Clustering* biasa digunakan pada banyak bidang, seperti : *data mining*, *pattern recognition* (pengenalan pola), *image classification* (pengklasifikasian gambar), ilmu biologi, pemasaran, perencanaan kota, pencarian dokumen, dan lain sebagainya.



Gambar 2.1 Ilustrasi clustering

#### 2.2.2. Karakteristik Clustering

Karakteristik clustering dibagi menjadi 4, yaitu :

##### 1. Partitioning clustering

Partitioning clustering disebut juga *exclusive clustering*, dimana setiap data harus termasuk ke cluster tertentu. Karakteristik tipe ini juga memungkinkan bagi setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster yang lain.

Contoh : K-Means, residual analysis.

##### 2. Hierarchical clustering

Pada hierarchical clustering, Setiap data harus termasuk ke cluster tertentu. Dan suatu data yang termasuk ke cluster tertentu pada suatu tahapan proses, tidak dapat berpindah ke cluster lain pada tahapan berikutnya.

- Contoh: Single Linkage, Centroid Linkage, Complete Linkage, Average Linkage.
3. Overlapping clustering  
Dalam overlapping clustering, setiap data memungkinkan termasuk ke beberapa cluster. Data mempunyai nilai keanggotaan (membership) pada beberapa cluster.  
Contoh: Fuzzy C-means, Gaussian Mixture.
  4. Hybrid  
Karakteristik hybrid adalah Mengawinkan karakteristik dari partitioning, overlapping dan hierarchical.

### 2.2.3. ALgoritma Clustering

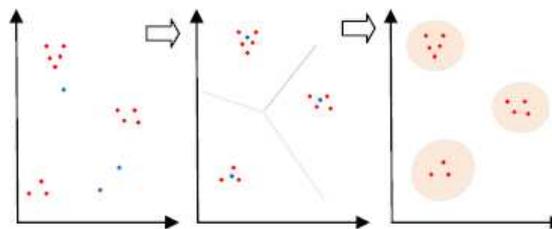
Ada beberapa algoritma yang sering digunakan dalam clustering, yaitu :

#### 1. K-Means

Termasuk partitioning clustering yang memisahkan data ke  $k$  daerah bagian yang terpisah. K-means algorithm sangat terkenal karena kemudahannya dan kemampuannya untuk mengklaster data besar dan data outlier dengan sangat cepat. Sesuai dengan karakteristik partitioning clustering, Setiap data harus termasuk ke cluster tertentu, dan Memungkinkan bagi setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster yang lain.

Algoritma K-Means :

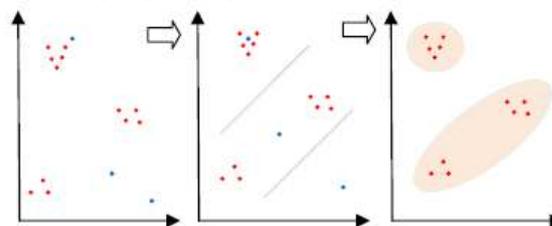
1. Menentukan  $k$  sebagai jumlah cluster yang ingin dibentuk
2. Membangkitkan  $k$  centroids (titik pusat cluster) awal secara random
3. Menghitung jarak setiap data ke masing-masing centroids
4. Setiap data memilih centroids yang terdekat
5. Menentukan posisi centroids baru dengan cara menghitung nilai rata-rata dari data-data yang memilih pada centroid yang sama.
6. Kembali ke langkah 3 jika posisi centroids baru dengan centroids lama tidak sama.



Gambar 2.2 Ilustrasi Algoritma K-means

Karakteristik K-Means :

- K-means sangat cepat dalam proses clustering
- K-means sangat sensitif pada pembangkitan centroids awal secara random
- Memungkinkan suatu cluster tidak mempunyai anggota
- Hasil clustering dengan K-means bersifat tidak unik ( selalu berubah-ubah) - terkadang baik, terkadang jelek.
- K-means sangat sulit untuk mencapai global optimum



Gambar 2.3 Ilustrasi kelemahan K-means

#### 2. Hierarchical Clustering

Dengan metode ini, data tidak langsung dikelompokkan kedalam beberapa cluster dalam 1 tahap, tetapi dimulai dari 1 cluster yang mempunyai kesamaan, dan berjalan seterusnya selama beberapa iterasi, hingga terbentuk beberapa cluster tertentu .

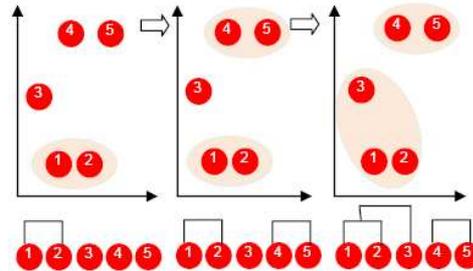
Arah hierarchical clustering dibagi 2, yaitu :

- a) Divisive
  - Dari 1 cluster ke  $k$  cluster
  - Pembagian dari atas ke bawah (*top to down division*)
- b) Agglomerative
  - Dari  $N$  cluster ke  $k$  cluster

- Penggabungan dari bawah ke atas (*down to top merge*).

Algoritma Hierarchical clustering :

1. Menentukan  $k$  sebagai jumlah cluster yang ingin dibentuk
2. Setiap data dianggap sebagai cluster. Kalau  $N =$  jumlah data dan  $c =$  jumlah cluster, berarti ada  $c = N$ .
3. Menghitung jarak antar cluster
4. Cari 2 cluster yang mempunyai jarak antar cluster yang paling minimal dan gabungkan (berarti  $c = c - 1$ ).
5. Jika  $c > k$ , kembali ke langkah 3.



**Gambar 2.4** Ilustrasi Algoritma Hierarchical Clustering

Penghitungan jarak antar obyek, maupun antar *clusternya* dilakukan dengan *Euclidian distance*, khususnya untuk data numerik [2]. Untuk data 2 dimensi, digunakan persamaan sebagai berikut :

$$d(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \quad (2.1)$$

Algoritma hierarchical clustering banyak diaplikasikan pada metode peng-clusteran berikut :

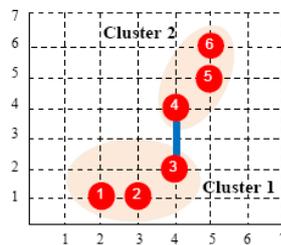
1. Single Linkage Hierarchical Method (SLHM)

Single Linkage adalah proses pengclusteran yang didasarkan pada jarak terdekat antar obyeknya (*minimum distance*).

Metode SLHM sangat bagus untuk melakukan analisa pada tiap tahap pembentukan cluster. Metode ini juga sangat cocok untuk dipakai pada kasus *shape independent clustering*, karena kemampuannya untuk membentuk *pattern* / pola tertentu dari cluster. Sedangkan untuk kasus *condensed clustering*, metode ini tidak bagus.

Algoritma Single Linkage Hierarchical Method :

1. Diasumsikan setiap data dianggap sebagai cluster. Kalau  $n =$  jumlah data dan  $c =$  jumlah cluster, berarti ada  $c = n$ .
2. Menghitung jarak antar cluster dengan *Euclidian distance*.
3. Mencari 2 cluster yang mempunyai jarak antar cluster yang paling minimal dan digabungkan (*merge*) kedalam cluster baru (sehingga  $c = c - 1$ )
4. Kembali ke langkah 3, dan diulangi sampai dicapai cluster yang diinginkan.



Jarak cluster 1 ke cluster 2  
 =  
 Jarak data 3 ke data 4

**Gambar 2.5** Ilustrasi Single Linkage

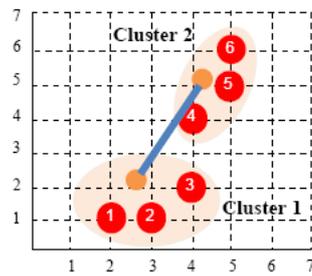
2. Centroid Linkage Hierarchical Method

Centroid Linkage adalah proses pengclusteran yang didasarkan pada jarak antar centroidnya. Metode ini bagus untuk memperkecil *variance within cluster* karena melibatkan centroid pada saat penggabungan antar cluster. Metode ini juga baik untuk data yang mengandung outlier.

Algoritma Centroid Linkage Hierarchical Method :

1. Diasumsikan setiap data dianggap sebagai cluster.  
 Kalau  $n =$  jumlah data dan  $c =$  jumlah cluster, berarti ada  $c = n$ .
2. Menghitung jarak antar cluster dengan *Euclidian distance*.
3. Mencari 2 cluster yang mempunyai jarak centroid antar cluster yang paling minimal dan digabungkan (*merge*) kedalam cluster baru (sehingga  $c = c - 1$ )

4. Kembali ke langkah 3, dan diulangi sampai dicapai cluster yang diinginkan.



Jarak cluster 1 ke cluster 2  
=  
Jarak centroid cluster 1 ke  
centroid cluster 2

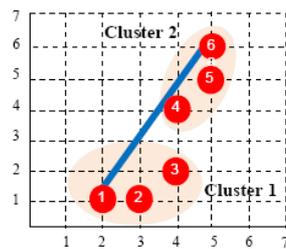
Gambar 2.6 Ilustrasi Centroid Linkage

3. Complete Linkage Hierarchical Method

Complete Linkage adalah proses pengclusteran yang didasarkan pada jarak terjauh antar obyeknya (*maksimum distance*). Metode ini baik untuk kasus clustering dengan normal data set distribution. Akan tetapi, metode ini tidak cocok untuk data yang mengandung outlier.

Algoritma Complete Linkage Hierarchical Method :

1. Diasumsikan setiap data dianggap sebagai cluster. Kalau  $n$ =jumlah data dan  $c$ =jumlah cluster, berarti ada  $c=n$ .
2. Menghitung jarak antar cluster dengan *Euclidian distance*.
3. Mencari 2 cluster yang mempunyai jarak antar cluster yang paling maksimal / terjauh dan digabungkan (*merge*) kedalam cluster baru (sehingga  $c=c-1$ )
3. Kembali ke langkah 3, dan diulangi sampai dicapai cluster yang diinginkan.



Jarak cluster 1 ke cluster 2  
=  
Jarak data 1 ke data 6

Gambar 2.7 Ilustrasi Complete Linkage

4. Average Linkage Hierarchical Method

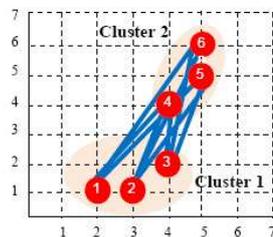
Average Linkage adalah proses pengclusteran yang didasarkan pada jarak rata-rata antar obyeknya (*average distance*). Metode ini relatif yang terbaik dari metode-metode hierarchial. Namun, ini harus dibayar dengan waktu komputasi yang paling tinggi dibandingkan dengan metode-metode hierarchial yang lain.

Jarak cluster 1 ke cluster 2  
=

$$\frac{\sum \text{Jarak antar data}}{n \times m}$$

Dimana :

$n$  = Jumlah data cluster 1  
 $m$  = Jumlah data cluster 2



Gambar 2.8 Ilustrasi Average Linkage

#### 2.2.4. Analisa Cluster

Analisa cluster adalah suatu teknik analisa *multivariate* (banyak variabel) untuk mencari dan mengorganisir informasi tentang variabel tersebut sehingga secara relatif dapat dikelompokkan dalam bentuk yang homogen dalam sebuah *cluster*. Secara umum, bisa dikatakan sebagai proses menganalisa baik tidaknya suatu proses pembentukan *cluster*. Analisa cluster bisa diperoleh dari kepadatan cluster yang dibentuk (*cluster density*). Kepadatan suatu cluster bisa ditentukan dengan *variance within cluster* ( $V_w$ ) dan *variance between cluster* ( $V_b$ ).

Varian tiap tahap pembentukan cluster bisa dihitung dengan rumus :

$$V_c^2 = \frac{1}{n_c - 1} \sum_{i=1}^n (y_i - \bar{y}_c)^2 \quad (2.2)$$

Dimana :

- $V_c^2$  = varian pada cluster  $c$
- $c$  = 1..k, dimana k = jumlah cluster
- $n_c$  = jumlah data pada cluster  $c$
- $y_i$  = data ke- $i$  pada suatu cluster
- $y_c$  = rata-rata dari data pada suatu cluster

Selanjutnya dari nilai varian diatas, kita bisa menghitung nilai *variance within cluster* ( $V_w$ ) dengan rumus :

$$V_w^2 = \frac{1}{N - c} \sum_{i=1}^c (n_i - 1) V_i^2 \quad (2.3)$$

Dimana :

- $N$  = Jumlah semua data
- $n_i$  = Jumlah data cluster  $i$
- $V_i$  = Varian pada cluster  $i$

Dan nilai *variance between cluster* ( $V_b$ ) dengan rumus :

$$V_b^2 = \frac{1}{c - 1} \sum_{i=1}^c n_i (\bar{y}_i - \bar{y})^2 \quad (2.4)$$

Dimana :

- $\bar{y}$  = rata-rata dari  $\bar{y}_i$

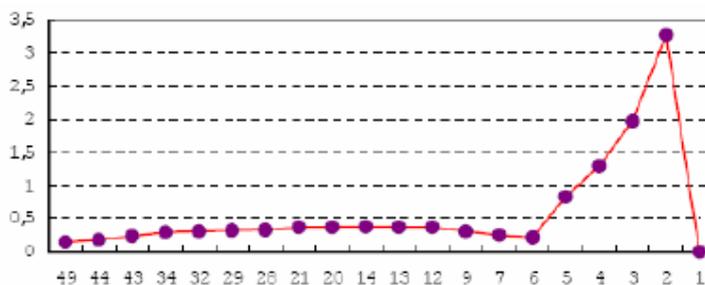
Salah satu metode yang digunakan untuk menentukan cluster yang ideal adalah batasan *variance*, yaitu dengan menghitung kepadatan *cluster* berupa *variance within cluster* ( $V_w$ ) dan *variance between cluster* ( $V_b$ ). Cluster yang ideal mempunyai  $V_w$  minimum yang merepresentasikan *internal homogeneity* dan maksimum  $V_b$  yang menyatakan *external homogeneity*.

$$V = \frac{V_w^2}{V_b^2} \times 100\% \quad (2.5)$$

Meskipun minimum  $V_w$  menunjukkan nilai cluster yang ideal, tetapi pada beberapa kasus kita tidak bisa menggunakannya secara langsung untuk mencapai *global optimum*. Jika kita paksakan, maka solusi yang dihasilkan akan jatuh pada *local optima*.

### 2.2.5. Identifikasi Pola Berdasarkan Pergerakan Varian

Identifikasi pola pergerakan varian merupakan metode untuk memperoleh cluster yang mencapai *global optimum*, yang mampu mengatasi masalah dari minimum  $V$ . Gambar 2.9 menunjukkan pergerakan varian pada tiap tahap pembentukan cluster, dimana dari gambar tersebut terlihat bahwa *global optimum* berada pada tahap ke 15, dengan 6 total cluster.



Gambar 2.9 Pergerakan Variance pada tiap Tahap Pembentukan cluster

Berikut tahap untuk menemukan *global optimum* dari tahap pembentukan cluster :

- a) Mendeskripsikan semua pola dari pergerakan varian, seperti gambar 2.9 diatas.
- b) Menganalisa kemungkinan *global optimum* yang berada pada tempat yang tepat.
- c) Melihat posisi dari *global optimum* yang mungkin.

Posisi yang mungkin untuk menemukan *global optimum* pada pergerakan varian, dikelompokkan menjadi 2, yaitu :

#### 1. Hill-climbing

Pada *Hill-climbing* didefinisikan bahwa kemungkinan mencapai *global optimum* terletak pada tahap ke- $i$ , jika memenuhi persamaan berikut :

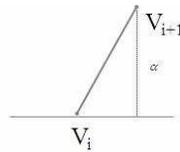
$$v_{i+1} > \alpha \cdot v_i \quad (2.6)$$

Dimana,  $\alpha$  adalah nilai tinggi.

Nilai tinggi digunakan untuk menentukan seberapa mungkin metode ini mencapai *global optimum*.

Nilai  $\alpha$  yang biasa digunakan adalah 2,3, dan 4.

Persamaan diatas, diperoleh berdasar analisa pergerakan varian pola *Hill-climbing* berikut :



**Gambar 2.10** Pola nilai beda *Hill-climbing*

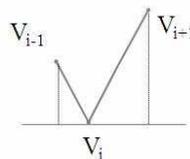
2. *Valley-tracing*

Pada *Valley-tracing* didefinisikan bahwa kemungkinan mencapai *global optimum* terletak pada tahap ke- $i$ , jika memenuhi persamaan berikut :

$$(v_{i-1} \geq v_i) \cap (v_{i+1} > v_i) \quad (2.7)$$

Dimana  $i = 1 \dots n$ , dan  $n$  tahap terakhir pembentukan cluster.

Persamaan diatas, diperoleh berdasar analisa pergerakan varian pola *Valley-tracing* berikut :



**Gambar 2.11** Pola nilai beda *Valley-tracing*

Berikut tabel yang menunjukkan pola-pola *Valley-tracing* yang mungkin mencapai global optimum. Pola yang mungkin ditandai dengan simbol  $\surd$ .

**Tabel 1.** Tabel kemungkinan pola *Valley-tracing* mencapai global optimum

Pola	Mungkin	Pola	Mungkin
	$\surd$		X
	$\surd$		X
	$\surd$		X
	X		$\surd$
	X		X
	X		X
	X		X
	X		X

Selanjutnya, baik dengan pendekatan metode *valley-tracing* maupun *hill-climbing* dilakukan identifikasi perbedaan nilai tinggi ( $\partial$ ) pada tiap tahap, yang didefinisikan dengan :

$$\begin{aligned} \partial &= (v_{i+1} - v_i) + (v_{i-1} - v_i) \\ \partial &= (v_{i+1} + v_{i-1}) - (2xv_i) \end{aligned} \quad (2.8)$$

Nilai  $\partial$  digunakan untuk menghindari *local optima*, dimana persamaan ini diperoleh dari maksimum  $\partial$  yang dipenuhi pada persamaan 7. Untuk membentuk cluster secara otomatis, yaitu cluster yang mencapai *global optima*, digunakan nilai  $\lambda$  sebagai threshold, sehingga cluster secara otomatis

terbentuk ketika memenuhi :

$$\max(\partial) \geq \lambda \quad (2.9)$$

Untuk mengetahui keakuratan dari suatu metode pembentukan cluster pada hierarchical method, baik menggunakan valley-tracing maupun hill-climbing, digunakan persamaan sebagai berikut :

$$\varphi = \frac{\max(\partial)}{\text{nilai terdekat ke max } (\partial)} \quad (2.10)$$

Dimana :

nilai terdekat ke max  $(\partial)$  adalah nilai kandidat max  $(\partial)$  sebelumnya.

Nilai  $\varphi$  yang lebih besar dari 2, menunjukkan cluster yang terbentuk merupakan cluster yang well-separated (terpisah dengan baik).

### 2.2.6. Konsep Warna

Warna pokok [3] dalam pengelolaan citra terdiri dari 3 (tiga) unsur, yaitu merah (R), hijau (H), dan biru (B). Jika warna-warna pokok tersebut digabungkan, maka akan menghasilkan warna lain. Penggabungan tersebut bergantung pada warna pokok yang tiap-tiapwarna memiliki nilai 256 (8 bit).

	(255, 0, 0)
	(255, 255, 0)
	(255, 255, 255)
	(0, 255, 0)
	(0, 0, 255)
	(128, 128, 128)
	(255, 0, 255)
	(0, 0, 0)

Warna yang dideskripsikan dengan RGB adalah pemetaan yang mengacu pada panjang gelombang dari RGB. Pemetaan menghasilkan nuansa warna untuk masing-masing R, G, dan B. Masing-masing R, G, dan B didiskritkan dalam skala 256, sehingga RGB akan memiliki indeks antara 0 sampai 255.

## 3. METODE PENELITIAN

Untuk mencapai tujuan dalam penelitian ini, diperlukan suatu metode penelitian yang tepat sehingga dapat digunakan sebagai pedoman dalam pelaksanaannya secara sistematis. Metode penelitian yang akan digunakan dalam penelitian ini nantinya meliputi tahapan penelitian yang terdiri dari :

### 3.1. Studi Kepustakaan

Adapun langkah-langkah yang dilakukan dalam penelitian ini diantaranya adalah melakukan pengumpulan serta mempelajari literatur-literatur yang berhubungan dengan image clustering dari penelitian-penelitian sebelumnya, dasar-dasar image processing, literature pendukung lainnya yang berhubungan dengan penelitian ini.

### 3.2. Rancang Bangun Sistem

Dalam penelitian ini memiliki 2 tahap yaitu tahap Learning (learning process) dan tahap identifikasi (identification proses).

#### 1. Tahap Learning (Learning Process)

Learning process adalah tahap pembelajaran kepada komputer untuk mengenali nilai RGB dari suatu image buah tomat, dimana nilai RGB tersebut adalah nilai centroid dari suatu proses image clustering.

Yang pertama kali di lakukan dalam learning process adalah pengambilan nilai RGB dari suatu image buah. Kemudian nilai RGB di setiap pixel di image tersebut di klusterisasi dengan menggunakan algoritma K-Means, sehingga di dapatkan nilai-nilai centroid dari suatu image tersebut.

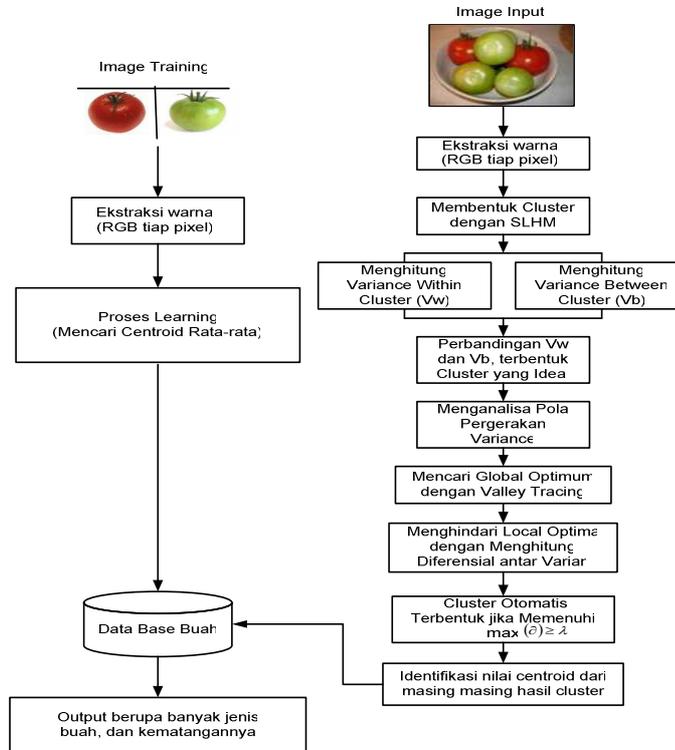
#### 2. Tahap Identifikasi (Identifikasi Process)

Pertama kali gambar buah tomat yang akan di identifikasi akan diambil nilai RGBnya untuk nantinya akan dilakukan proses automatic clustering. Pada proses automatic clustering ini akan menggunakan metode valley tracing untuk mencari berapa cluster yang terbentuk secara otomatis sehingga akan di dapatkan nilai centroidnya dimana nilai centroid tersebut akan digunakan untuk melakukan proses identifikasi.

Proses identifikasi ini akan merujuk pada database yang ada, dimana program akan menganalisa nilai-nilai

RGB dari image yang telah di klusterisasi dengan nilai-nilai yang terdapat dalam database yang berisi nilai-nilai RGB dari hasil cluster pada proses learning sehingga jenis buah tomat dan kematangannya yang terdapat dalam image input dapat di ketahui.

### 3.2.1. Blok Diagram Sistem



Gambar 3.1. Blok Diagram Sistem

### 3.3. Pengujian Unjuk Kerja Sistem

Pada pengujian sistem dilakukan implementasi dari rancangan sistem yang telah dijelaskan dari blok diagram sistem yang ada. Gambar 3.1 menunjukkan blok diagram sistem dari proses training, ekstraksi warna, learning, cluster sampai pada identifikasi kematangan buah pada sebuah image yang berisi kumpulan buah.

### 3.4. Analisa Sistem

Penganalisaan yang dilakukan yaitu dengan mengambil data training dari masing masing image buah tomat dalam posisi yang berbeda, lalu diberi nama dengan nama buah serta jenis kematangannya kemudian disimpan dalam database yang dijadikan sebagai data training. Gambar 3.1 diatas menggambarkan blok diagram proses identifikasi, dimana setelah gambar diekstraksi warnanya, sehingga didapat nilai R,G,dan B, selanjutnya nilai tersebut di-cluster oleh system dengan menggunakan algoritma Single Linkage Hierarchical Method yang juga mengaplikasikan metode valley tracing untuk membentuk cluster secara otomatis. Pada tiap tahap pembentukan cluster dilakukan penghitungan nilai beda tinggi  $\delta$ . Nilai tersebut dihitung dengan cara terlebih dahulu melakukan pengecekan dengan metode valley tracing, jika memenuhi, maka nilai  $\delta$  dihitung. Selanjutnya pada nilai  $\delta_{max}$  itulah cluster yang optimal terbentuk. Dari cluster yang terbentuk diperoleh nilai centroid. Masuk ke proses identifikasi, nilai centroid hasil dihitung kedekatannya dengan nilai centroid masing-masing data learning menggunakan euclidean distance. Nilai jarak yang paling minimal dipilih dengan representasi label pada data training.

## 4. KESIMPULAN DAN SARAN

### 4.1. Kesimpulan

1. Tidak semua dari hasil percobaan yang dilakukan menghasilkan suatu identifikasi yang baik dan benar, karena proses capture yang kurang sempurna, sehingga pada saat klusterisasi terdapat kluster-kluster lain yang terbentuk.
2. Untuk mendapatkan cluster yang tepat pada saat pengclusteran otomatis terhadap data fitur R,G,B. Kualitas gambar juga menentukan tepat tidaknya hasil cluster.

#### 4.2. Saran

Hasil yang dicapai dari penelitian ini belum sempurna. Untuk meningkatkan hasil yang dicapai, maka diperlukan :

1. *Pre-processing* terhadap image yang akan diproses dengan clustering, hal ini untuk mereduksi pixel-pixel yang mengganggu yang mengakibatkan terbentuknya cluster yang tidak diinginkan.
2. Penggunaan algoritma clustering yang lebih ampuh untuk membentuk cluster, yang mampu menangani data yang mengandung *outlier* (dalam hal ini pixel yang mengganggu).
3. Pengambilan gambar dengan pencahayaan yang bagus, agar diperoleh kualitas gambar yang bagus.

#### 5. DAFTAR PUSTAKA

- Ali Ridho Barakbah, Kohei Arai, 2004, "Determining Constraints of Moving Variance to Find Global Optimum and Make Automatic Clustering" *IES Proc.* Pp 409-413.
- C.J. Veenman, M.J.T. Reinders, E. Backer, 2002, "A Maximum Variance Cluster Algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 9, pp. 1273-1280.
- <http://www.elet.polimi.it/upload/matteucc/Clustering/tutorialhtml/index.html>
- S. Ray, R.H. Turi, 1999, "Determination of Number of Clusters in K-means Clustering and Application in Colour Image Segmentation", *4th ICAPRDT Proc.*, pp.137-143.