

KINERJA METODE MAHALANOBIS DISTANCE YANG DIBENTUK DARI DUA UKURAN PUSAT DAN DUA DISPERSI MULTIVARIAT (UNTUK UKURAN SIMILARITAS KLASIFIKASI IMAGE)

Tania Kantacarini¹⁾, Dyah Erny Herwindiaty²⁾, Janson Hendryli³⁾

¹⁾²⁾³⁾ Teknik Informatika Universitas Tarumanagara

Jl. Letjen S. Parman No.1, Jakarta

Tania.535160079@stu.untar.ac.id¹⁾, dyahh@untar.ac.id²⁾, jansonh@untar.ac.id³⁾

ABSTRACT

Distance is a space that connects two points or two locations which can be calculated by length and time. Distance is used to measure the similarity of two objects (for example, an image object). An image can be considered to be similar to another image if the similarity size value is small. On the contrary, if the value of the similarity distance between the training object and the object being tested is large, the object can be said to be different or not. In this design, the image classification of Lakes, Forests and Settlements will be carried out by taking the Color feature using the Color Moment extraction method and the Texture feature using the GLCM (Gray Level Co-occurrence Matrix) extraction method and taking the method of calculating the distance between one data and another data that has High similarity using the Mahalanobis Distance calculation method with two center sizes namely Mean and Median and three multivariate dispersions, namely the covariance matrix formed by the mean center value, the covariance matrix formed by the value of Median, and the covariance matrix formed by the value of Grand Median. From the research conducted, the performance results that can be considered for use are the Mahalanobis Distance with a median center size with a covariance matrix formed by the Median center value with an accuracy of 69.855% and a covariant matrix formed by the Grand Median center value with an accuracy of 69.565%. In this case the percentage is taken from testing images based on color characteristics using the Color Moment extraction method.

Kata Kunci: Citra, Klasifikasi, Mahalanobis Distance, Matriks Covarian, Median

1. Pendahuluan

Notion distance adalah sebuah konsep atau ide mengenai jarak. Salah satu kegunaan jarak adalah untuk mengukur similaritas dua objek (misal objek citra). Sebuah citra dikatakan mirip dengan gambar dalam database jika nilai ukuran similaritasnya kecil. [1] Maka sebaliknya, jika nilai jarak similaritas antar

objek latih dan objek yang diuji bernilai besar objek dapat dikatakan tidak mirip atau tidak sejenis.

Saat ini, terdapat beragam metode *Notion of Distance* untuk menghitung jarak diantaranya yang paling populer digunakan adalah metode *Euclidean Distance*. Tetapi jarak *Euclidean* sangat sensitif terhadap besarnya sampel dan besarnya sebaran varians, oleh karena itu jika dalam suatu kasus objek yang dibandingkan memiliki varian yang sangat berbeda, maka hasil perhitungan *Euclidean Distance* menjadi tidak bagus. [2]

Penghitungan jarak Mahalanobis ini didasarkan pada korelasi antara variabel dan dengan pola yang berbeda dapat diidentifikasi dan dianalisis berdasarkan titik referensinya.[3] Hal tersebut dapat mengatasi permasalahan dari metode penghitungan jarak *metode Euclidean Distance* yang rentan terhadap besaran varian. Dalam perancangan ini akan dilakukan klasifikasi *image* dengan menggunakan perhitungan jarak antara data satu dengan data lainnya yang memiliki similiaritas yang tinggi. Jenis gambar yang diambil sebagai objek penelitian adalah pemukiman dan ruang terbuka hijau. Adapun objek ruang terbuka hijau yang diambil adalah: danau, perkebunan. Ciri image yang digunakan untuk klasifikasi adalah ciri warna dan ciri tekstur.

Pada perancangan ini akan digunakan metode Mahalanobis Distance yang dibentuk berdasarkan 3 macam variabel multivariat. Yang pertama, Mahalanobis Distance yang dibentuk dari ukuran pusat *mean* dan dispersi multivariat Matriks Covariat dari ukuran pusat *mean* tersebut. Lalu kedua, Mahalanobis Distance yang dibentuk dari ukuran pusat median dan dispersi multivariat Matriks Covariat yang dibentuk dari median tersebut. Dan yang ketiga, Mahalanobis Distance yang dibentuk dari ukuran pusat median dan dispersi multivariat Matriks Covariat yang dibentuk dari Spool median atau gabungan seluruh median.

Berdasarkan metode ekstraksi ciri warna yang sederhana yaitu metode *Color Moments* serta metode ekstraksi ciri tekstur yaitu metode *Gray Level Co-occurrence Matrics* (GLCM) diharapkan penelitian ini mampu menunjukkan bagaimana performa metode *Mahalanobis Distance* dalam melakukan klasifikasi antara citra pemukiman dan ruang terbuka hijau.

2. Dasar Teori

2.1 Pengolahan Citra Digital

Sebuah citra dapat didefinisikan sebagai sebuah fungsi dua dimensi, $f(x,y)$ dimana x dan y adalah spasial (*plane*) koordinat, dan *amplitude* dari f berada di poin itu. Ketika x,y dan nilai intensitas dari f semua terbatas, kuantitas diskrit, maka sebuah citra dikatakan citra digital. Pengolahan citra digital berarti pemrosesan citra digital menggunakan komputer digital.[4]

2.2 Klasifikasi

Klasifikasi gambar merupakan sebuah proses ekstraksi informasi kelas dari multiband. Tujuan dari kegiatan klasifikasi gambar ini adalah untuk mengidentifikasi lalu mengelompokkan gambar-gambar digital berdasarkan dari fitur-fitur yang terdapat dalam citra digital ke kelompok tertentu dengan yang memiliki similaritas tinggi.

- *Supervised Classification*
Supervised Classification adalah proses klasifikasi yang membutuhkan proses pelatihan data terlebih dahulu untuk mendapatkan data latih yang memiliki label. Label-label pada data latih tersebut menunjukkan bahwa data latih memiliki kelas tertentu sehingga ketika terdapat data uji dengan similaritas ciri yang sama dengan data latih pada kelas tertentu maka dapat ditentukan label dari data uji tersebut.
- *Unsupervised Classification*
Unsupervised Classification adalah proses klasifikasi yang tidak membutuhkan proses pelatihan data seperti yang dibutuhkan pada *Supervised Classification* melainkan hanya melakukan pengelompokan data uji berdasarkan dari similaritas antar data tersebut ke dalam beberapa kelompok yang disebut *cluster*.

2.3 HSV Color Space

1. Hue (H) yang merepresentasikan warna melingkar dengan rentang dari 0 derajat sampai 360 derajat dengan urutan warna merah, kuning, hijau, cyan, biru, magenta dan kembali lagi ke warna merah.
2. Saturation (S) yang memiliki rentang dari 0.0 sampai 1.0, mulai dari nilai 0.0 yang merepresentasikan unsaturated color atau warna abu-abu hingga nilai 1.0 yang merepresentasikan saturated color yang tidak memiliki komponen warna putih.
3. Value (V) yang memiliki rentang dari 0.0 sampai 1.0, mulai dari nilai 0.0 yang merepresentasikan warna yang gelap hingga nilai 1.0 yang merepresentasikan warna terang. Sehingga semakin tinggi nilai valuenya warna menjadi semakin terang.

Nilai Hue, Saturation dan Value didapatkan dari nilai Red, Green, dan Blue yang dihitung dengan rumus sedemikian rupa seperti berikut ini:

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B}$$

$$V = \max(r, g, b)$$

$$S = \begin{cases} 0, & \text{jika } V = 0 \\ 1 - \frac{\min(r,g,b)}{V}, & \text{jika } V > 0 \end{cases}$$

$$H = \begin{cases} 0, & \text{jika } S = 0 \\ \left[\frac{60 \cdot (g-b)}{S \cdot V} \right], & \text{jika } V = r \\ 60 \cdot \left[2 + \frac{b-r}{S \cdot V} \right], & \text{jika } V = g \\ 60 \cdot \left[4 + \frac{r-g}{S \cdot V} \right], & \text{jika } V = b \end{cases}$$

Keterangan:

R, G, B = Nilai Merah, Hijau dan Biru piksel
 r, g, b = Nilai Merah, Hijau, dan Biru Rumus
 H, S, V = Nilai Hue, Saturation, dan Value

2.4 Color Moments

Color Moments adalah pengukur yang bisa dipakai untuk membedakan citra-citra berdasarkan fitur warna. Setelah dihitung, momen-momen ini menghasilkan nilai untuk kesamaan warna pada citra-citra. Nilai kesamaan ini dapat dibandingkan dengan citra lain untuk proses *image retrieval*. [5]

Tiga momen sentral dari distribusi warna pada citra adalah *Mean*, *Standard Deviation*, dan *Skewness*. Sebuah warna dapat didefinisikan oleh 3 warna atau lebih. Momen-momen dihitung pada tiap-tiap *channel* dari sebuah citra dimana *channel* tersebut terdiri atas *channel Hue*, *Saturation* dan *Value* yang disebut juga sebagai *HSV Color Space*

Berikut adalah 3 rumus *Color Moments* :

$$E_i = \sum_{j=1}^N \frac{1}{N} P_{ij}$$

$$\sigma_i = \sqrt{\left(\frac{1}{N} \sum_{j=1}^N (P_{ij} - E_i)^2 \right)}$$

$$S_i = \sqrt[3]{\left(\frac{1}{N} \sum_{j=1}^N (P_{ij} - E_i)^3 \right)}$$

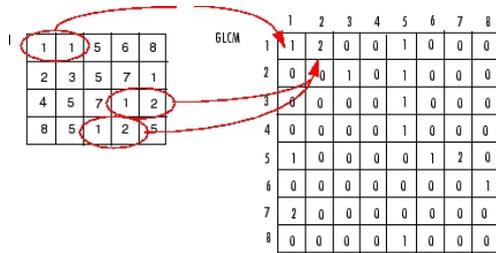
Keterangan:

$E_i, \sigma_i,$ dan S_i = Nilai Mean, Standar Deviasi, dan Skewness
 N = total banyaknya piksel dalam citra
 P_{ij} = posisi piksel pada baris i kolom j

2.5 GLCM (Gray Level Co-Occurrence Matrix)

GLCM (Gray Level Co-Occurrence Matrix) merupakan sebuah metode yang digunakan untuk

melakukan ekstraksi fitur tekstur. Input yang dibutuhkan untuk melakukan ekstraksi fitur tekstur ini yaitu berupa citra *grayscale*. Jika citra yang diinput pada umumnya merupakan citra RGB maka citra tersebut harus dikonversi terlebih dahulu agar menjadi citra *grayscale*. Dimana citra *grayscale* akan digunakan untuk membentuk matriks GLCM.



Gambar 1 Ilustrasi Matriks GLCM

GLCM merupakan sebuah matriks yang jumlah baris dan jumlah kolomnya ditentukan berdasarkan dengan jumlah gray levels yang terdapat pada citra. GLCM menghitung seberapa sering piksel abu dengan nilai I muncul kearah tertentu terhadap piksel tetangga dengan jarak radius d dan nilai j. Arah tertentu yang digunakan dalam melakukan analisis piksel abu pada GLCM adalah horizontal (0 derajat), vertikal (90 derajat), dan diagonal (45 derajat dan 135 derajat) seperti pada gambar diatas. Nilai GLCM untuk 4 sudut diatas dengan jarak d = 1 akan dinyatakan dalam matriks P(d, θ) dengan d adalah jarak, dan θ adalah sudut / arah. Berikut adalah vektor ciri yang bisa didapatkan dari metode GLCM beserta penjelasannya[6] :

1. Contrast

Perhitungan nilai kontras atau variasi intensitas lokal antara piksel referensi dengan piksel tetangga pada keseruan citra yang akan memberikan nilai matriks P(i,j) yang non diagonal.

$$Contrast = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - j)^2 P(i, j)$$

2. Correlation

Korelasi adalah perhitungan dependensi linier dari gray level antara piksel dengan posisi ditentukan relatif dengan piksel disekitarnya. Nilai fitur selanjutnya merupakan turunan dari nilai – nilai yang sudah disebutkan diatas.

$$Correlation = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{(ixj) x P(i,j) - (\mu x + \mu y)}{stdx + stdy}$$

3. Energy

Citra yang homogen akan memiliki sedikit gray level, sehingga memberikan sedikit GLCM dengan nilai matriks P(i,j) yang relatif tinggi, serta akan menghasilkan sum of squares yang tinggi. Nilai Energy dapat dibentuk dari persamaan:

$$Energy = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (P(i, j))^2$$

4. Homogeneity

Menghasilkan nilai perhitungan yang mengukur tingkat kehomogenan suatu citra dengan rentang nilai [0 1]

$$Homogeneity = \sum_{ij} \frac{P(i, j)}{1 + |i - j|}$$

2.6 Mahalanobis Distance

Mahalanobis Distance merupakan sebuah metode yang digunakan untuk melakukan perhitungan jarak. Pada dasarnya metode Mahalanobis Distance melakukan perhitungan jarak dengan cara mengukur setiap objek dengan ukuran pusat dari himpunan objek tersebut.

Pada perancangan ini, ukuran pusat yang digunakan untuk membentuk Mahalanobis Distance adalah nilai pusat Mean dan Median.

Dan nilai Median dinyatakan dengan persamaan sebagai berikut :

Jika N ganjil maka, $med = x_{(\frac{N}{2}+1)}$

Jika N genap maka, $med = \frac{x_{\frac{N}{2}} + x_{(\frac{N}{2}+1)}}{2}$

Keterangan:

X = data ke-i pada himpunan data

n = banyaknya sampel dalam suatu himpunan data

Perhitungan jarak Mahalanobis dinyatakan dalam bentuk matriks dan vector serta dapat dihitung dengan persamaan umum sebagai berikut :

$$D^2 = (x_i - me) C^{-1} (x_i - me)^T$$

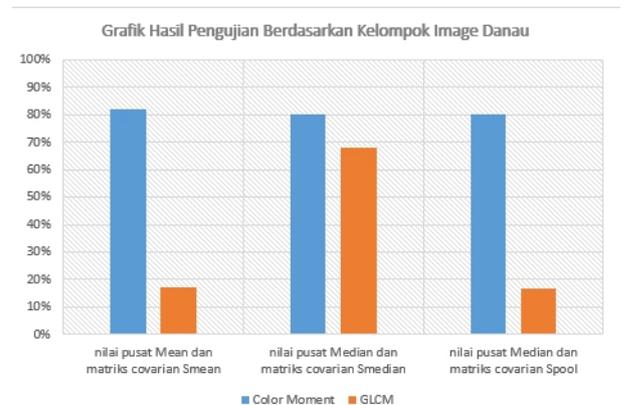
Keterangan:

x_i = variabel setiap data data ke- i dimulai dari 1 hingga N

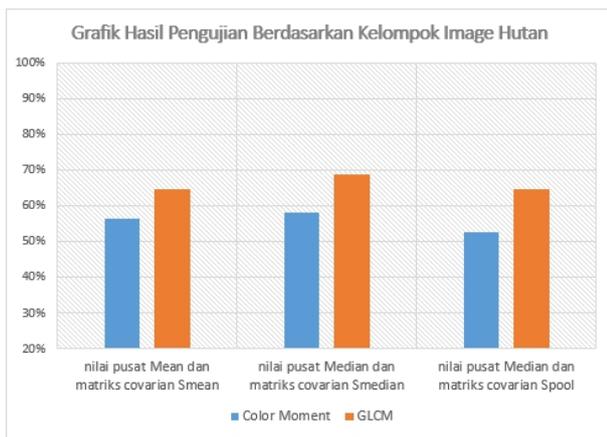
me = Nilai pusat dari kelompok variabel

C^{-1} = Invers matriks covarian

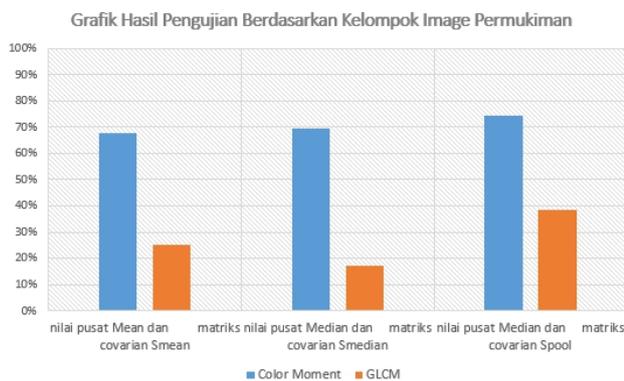
3. Hasil Pengujian



Gambar 2 hasil pengujian kelompok image danau



Gambar 3 hasil pengujian kelompok image hutan



Gambar 4 hasil pengujian kelompok image permukiman



Gambar 5 hasil pengujian keseluruhan

4. Kesimpulan

Kesimpulan yang dapat diperoleh dari hasil pengujian yang dilakukan terhadap program aplikasi Klasifikasi Image dengan menggunakan metode Mahalanobis Distance adalah sebagai berikut :

- Hasil klasifikasi gambar yang dihasilkan oleh program aplikasi dari masing masing metode memiliki kinerja dengan rata rata tingkat akurasi hasil klasifikasi berdasarkan ciri warna yaitu

69,371% dan rata rata keakuratan hasil klasifikasi berdasarkan ciri tekstur yaitu 42.36% dimana Metode perhitungan jarak Mahalanobis Distance lebih dipertimbangkan untuk mengklasifikasi gambar berdasarkan ciri warna.

- Dari hasil pengujian, Mahalanobis Distance yang dapat dipertimbangkan untuk digunakan sebagai ukuran similaritas adalah Mahalanobis Distance yang dibentuk dari ukuran pusat Median. Sebagai skala disperse multivariate adalah matriks kovarian yang dibentuk dari Median dan Grand Median, dalam hal ini peneliti menyebutnya sebagai Smedian dan SpoolMedian. Dengan tingkat akurasi 69,855% dan 69,565%. Dalam hal ini, persentase kinerja Mahalanobis Distance diambil dari hasil pengujian berdasarkan ciri warna.

5. Referensi

- [1] DYAH E. HERWINDIATI, SANI M. ISA, RAHMAT SAGARA, "THE NEW NOTION DISTANCE OF CONTENT BASED IMAGE RETRIEVAL (CBIR)". *Journal of the Indonesian Mathematical Society (JIMS)*, Vol. 16 No. 1 (2010), hal. 52.
- [2] Uzlifatul Jannah, "PERBANDINGAN JARAK EUCLID DENGAN JARAK MAHALANOBIS PADA ANALISIS CLUSTER HIRARKI", <http://etheses.uin-malang.ac.id/6739/1/06510057.pdf>
- [3] Dwi Nugraheny, "METODE NILAI JARAK GUNA KESAMAAN ATAU KEMIRIPAN CIRI SUATU CITRA (KASUS DETEKSI AWAN CUMULONIMBUS MENGGUNAKAN PRINCIPAL COMPONENT ANALYSIS)", <http://ejournals.stta.ac.id/index.php/angkasa/article/view/145>.
- [4] Gonzales, C. R dan Woods, R. E. *Digital Image Processing*. (Upper Saddle River: Prentice-Hall Inc, 2002), http://sdeuoc.ac.in/sites/default/files/sde_vid_eos/Digital%20Image%20Processing%203rd%20ed.%20-%20R.%20Gonzalez%2C%20R.%20Woods-ilovepdf-compressed.pdf, 2 April 2020
- [5] Noah Keen, *Color Moments*, http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0405/KEEN/av_as2_nkeen.pdf, 06 Maret 2020
- [6] Fritz Albrechtsen, *Statistical Texture Measures Computed from Gray Level Cooccurrence Matrices*, <http://citeseerx.ist.pdu.edu/viewdoc/download?doi=10.1.1.332.463&rep=rep1&type=pdf>, 17 Maret 2020.