METODE NILAI JARAK GUNA KESAMAAN ATAU KEMIRIPAN CIRI SUATU CITRA (KASUS DETEKSI AWAN *CUMULONIMBUS* MENGGUNAKAN *PRINCIPAL COMPONENT ANALYSIS*)

Dwi Nugraheny

Teknik Informatika
Sekolah Tinggi Teknologi Adisutjipto
Jalan Janti Blok R Lanud Adisutjipto, Yogyakarta
henynug@gmail.com

Abstract

One commonality or similarity matching phase characteristics of an image is by using the method of distance measurement. Distance is an important aspect in the development of methods of grouping and regression. Before the grouping of data or object to the detection process, first determined the size of the proximity distance between data elements. In this study, there will be a comparison of several methods including distance measurement using Euclidean distance, Manhattan/ City Block Distance, Mahalanobis which will be implemented in the case of Cumulonimbus image clouds detection using Principal Component Analysis (PCA). The average percentage of accuracy of image similarity value Cumulonimbus clouds using the Euclidean distance method was 93 percent and the distance Manhattan/ City Block Distance is 90 percent, while the Mahalanobis distance method was 50 percent.

Keywords: Similarity, Cumulonimbus, Euclidean, Manhattan, Mahalanobis, PCA

Abstrak

Salah satu tahap pencocokan kesamaan ataupun kemiripan ciri-ciri suatu citra adalah dengan menggunakan metode pengukuran jarak. Jarak merupakan aspek penting dalam pengembangan metode pengelompokan maupun regresi. Sebelum dilakukan pengelompokkan data atau objek untuk proses deteksi, terlebih dahulu ditentukan ukuran jarak kedekatan antar elemen data. Pada penelitian ini, akan dilakukan perbandingan dari beberapa metode pengukuran jarak diantaranya menggunakan jarak *Euclidean, Manhattan/City Block Distance*, *Mahalanobis* yang akan di implementasikan pada deteksi citra awan *Cumulonimbus* menggunakan *Principal Component Analysis* (PCA). Hasil rata-rata persentase keakuratan nilai kemiripan citra awan *Cumulonimbus* menggunakan metode jarak Euclidean adalah 93 persen dan jarak *Manhattan/City Block Distance* adalah 90 persen, sedangkan metode jarak *Mahalanobis* adalah 50 persen.

Kata kunci: Kemiripan, *Cumulonimbus, Euclidean, Manhattan, Mahalanobis*, PCA.

1. Pendahuluan

Perhitungan nilai kesamaan suatu citra merupakan salah satu komponen esensial dari hasil proses suatu algoritma ataupun metode untuk mencocokkan antara citra hasil *training*

dengan citra baru sebagai pembanding. Keakuratan perhitungan nilai kesamaan/ kemiripan sangat menentukan apakah citra hasil proses *training* dari suatu algoritma memiliki ciri-ciri kesamaan ataupun kemiripan dengan citra baru sebagai pembanding. Proses pencocokan kesamaan ataupun kemiripan ciri-ciri suatu citra tersebut salah satunya dengan menggunakan metode pengukuran jarak. Jarak merupakan aspek penting dalam pengembangan metode pengelompokan maupun regresi. Sebelum dilakukan pengelompokkan data atau objek untuk proses deteksi, terlebih dahulu ditentukan ukuran jarak kedekatan antar elemen data.

Beberapa metode pengukuran jarak dalam mengukur tingkat kesamaan (*similarity*) atau kemiripan suatu citra diantaranya menggunakan jarak *Euclidean*, *Manhattan/City Block Distance*, *Mahalanobis*, *Correlation*, *Angle-based*, *Squarred Euclidean*. Sehingga pada akhirnya akan dibandingkan jarak manakah yang memiliki tingkat kesamaan ataupun kemiripan paling tinggi untuk suatu citra.

Principal Component Analysis (PCA) adalah suatu teknik handal untuk mengekstrasi struktur dari suatu set data dengan dimensi yang cukup banyak (Budi Santosa, 2007). PCA dapat digunakan untuk menyelesaikan proses-proses citra digital sebagai salah satu bentuk data. PCA mampu mendapatkan pola suatu data sehingga dapat digunakan untuk mengetahui kemiripan atau perbedaan dengan data yang lain (Smith I Lindsay, 2002)

Berdasarkan uaraian tersebut di atas, pada penelitian ini akan dilakukan perbandingan beberapa metode jarak antar vektor dan di implementasikan menggunakan *Principal Component Analysis* (PCA) untuk membangun model deteksi citra awan *Cumulonimbus.*

2. Tinjauan Pustaka dan Landasan Teori

2.1 Tinjauan Pustaka

Bajwa I.S, dkk, (2009), melakukan penelitian dalam jurnalnya tentang pengklasifikasian gambar berdasarkan fitur-fitur utamanya menggunakan algoritma *Principal Component Analysis* (PCA). Pada penelitian ini pengklasifikasian dilakukan pada gambar awan untuk mengidentifikasikan jenis-jenis awan dengan harapan diperoleh tingkat keakuratan yang lebih baik. Sistem didesain dengan membaca fitur-fitur berasal dari gambar *grayscale* untuk membuat *Imagespace*

Dwi Nugraheny (2010), dalam makalah Deteksi Awan *Cumulonimbus* menggunakan *Principal Component Analysis* (PCA), menjelaskan bahwa Model deteksi awan *Cumulonimbus (Cb)* menggunakan PCA melalui tahap ekstraksi fitur awan (*training*) dan tahap deteksi (*recognition*). Tahap ekstraksi fitur awan membangun vektor baris citra awan sampai diperoleh *eigenclouds*. Adapaun metode jarak antar vektor yang digunakan adalah jarak *Euclidean.*

Pada penelitian ini akan digunakan beberapa metode jarak selain *Euclidean* untuk melihat perbandingan nilai keakuratan kemiripan citra *Cumulonimbus* hasil *training* dengan citra awan lain ataupun sebagai citra awan *Cumulonimbus.*

2.2 Landasan Teori

2.2.2. *Similarity Measurement*

Similarity Measurement adalah proses pengukuran kemiripan suatu objek terhadap objek acuan. Dalam *Similarity Measurement* akan dilakukan pengukuran jarak (*distance*), di mana semakin meningkat jarak (*distance*) antara dua objek, maka semakin berbeda dua objek tersebut, *distance* biasanya adalah ukuran dari ketidakmiripan (Rencher, A. C., 2002)

2.2.3. Beberapa Metode Perhitungan Kesamaan atau Kemiripan Citra

Beberapa metode perhitungan kesamaan atau kemiripan citra yang biasa disebut dengan jarak merupakan aspek penting dalam pengembangan metode pengelompokan maupun regresi. Sebelum dilakukan pengelompokkan data atau objek untuk dideteksi, terlebih dahulu ditentukan ukuran jarak kedekatan antar elemen data.

- Jarak *Euclidean* (*Euclidean distance*)

Metode *Euclidean* membandingkan jarak minimum image pengujian (*testing*), dengan database image pelatihan (*training*). Jarak e*uclidean* dari dua vektor *x* dan *y* dihitung dengan Persamaan 1:

$$d(x,y) = \left(\sum_i (x_i - y_i)^2 \right)^{\frac{1}{2}} \quad (1)$$

Semakin kecil nilai $d(x,y)$, maka semakin mirip kedua vektor yang dicocokkan/dibandingkan. Sebaliknya semakin besar nilai $d(x,y)$ maka semakin berbeda kedua vektor yang dicocokkan (Budi Santosa, 2007).

- Jarak *Manhattan* (*Manhattan Distance*)

Merupakan salah satu pengukuran yang paling banyak digunakan meliputi penggantian perbedaan kuadrat dengan menjumlahkan perbedaan absolute dari variabel-variabel. Prosedur ini disebut blok absolute atau lebih dikenal dengan *city block distance* (lihat Persamaan 2).

$$d(x,y) = L_p = i(x\,y) = \sum_i^n \|x_i - y_i\| \quad (2)$$

- Jarak *Mahalanobis* (*Mahalanobis Distance*)

Penghitungan jarak *Mahalanobis* ini didasarkan pada korelasi antara variabel dan dengan pola yang berbeda dapat diidentifikasi dan dianalisis berdasarkan titik referensinya (lihat Persamaan 3).

$$d(x,y) = -\sum_i^n z_i x_i y_i \quad (3)$$

Di mana nilai $Z_i = \sqrt{\frac{\lambda_i}{\lambda_i + \alpha^2}}$ $\alpha = 0{,}25$

λ_i bergantung pada besarnya *eigenvalues*.

2.2.4. *Principal Component Analysis* (PCA)

Principal Component Analysis (PCA) adalah suatu teknik handal untuk mengekstrasi struktur dari suatu set data dengan dimensi yang cukup banyak (Budi Santosa, 2007). PCA dapat digunakan untuk menyelesaikan proses-proses citra digital sebagai salah satu bentuk data. PCA mampu mendapatkan pola suatu data sehingga dapat digunakan untuk mengetahui kemiripan atau perbedaan dengan data yang lain (Smith I Lindsay, 2002) Pada dasarnya algoritma *Principal Component Analysis* (PCA) digunakan untuk menghitung perbedaan antara dua vektor citra dalam *eigenspace* (adalah suatu ekstraksi fitur dan teknik reduksi dimensi yang optimal jika dipandang dari sudut pandang teori informasi. Ide dasarnya adalah menentukan komponen utama dari serangkaian citra yang mengumpul membentuk suatu arah

menurut varian maksimumnya, maka dimensi-dimensi yang berkontribusi tersebutlah yang dipertahankan dan sisanya disingkirkan untuk tahap pemrosesan berikutnya (Mudrova M, dkk., 2002).

2.2.5. Jenis-Jenis Awan Udara Naik

Jenis awan ini terletak pada ketinggian antara 500-1500 m. Macam-macam jenis awan udara naik (*http://www.artikelsiana.com*) adalah awan *Cumulus* (Cu) dan awan *Cumulonimbus* (Cb).

Awan *Cumulus* (Cu) adalah awan tebal dengan puncak-puncak yang tinggi, terbentuk di siang hari karena udara naik. Jika berhadapan dengan matahari terlihat terang dan jika memperoleh sinar hanya sebelah saja akan menimbulkan bayangan yang berwarna kelabu. Gambar 1 merupakan salah satu jenis awan *Cumulus*.

Gambar 1. Jenis awan *Cumulus*

Awan *Cumulonimbus* (Cb) adalah awan yang menimbulkan hujan dengan kilat guntur. Awan ini mmiliki volume yang besar posisi yang rendah dengan puncak yang tinggi sebagai menara atau gunung dan puncaknya melebar, sehingga merupakan awan tebal. *Cumulonimbus* (Gambar 2) merupakan jenis awan yang menghasilkan kilat/halilintar, guntur, mengandung hujan air dan hujan batu es, berangin kencang, serta berangin taufan (*tornadoes*). Paling tinggi dari semua jenis awan, yang dapat menjangkau semua layer dengan luas 60.000 kaki. Jenis awan ini biasanya mempunyai ciri berbentuk luas ke atas, karenanya berangin kencang pada tingkatan atmosfer tertinggi.

Gambar 2. Jenis awan *Cumulonimbus*

3. Metodologi

Metodologi yang dilakukan pada penelitian ini, melalui tahap-tahap seperti pada Gambar 3.

a. Identifikasi Masalah

Pada tahap identifikasi masalah ini, dilakukan tentang pemahaman masalah dari penelitian terdahulu dari penelitian penulis. Kemudian melakukan pengumpulan materi dan data pendukung guna memperkuat isi penelitian ini.

b. Analisis Kebutuhan Sistem

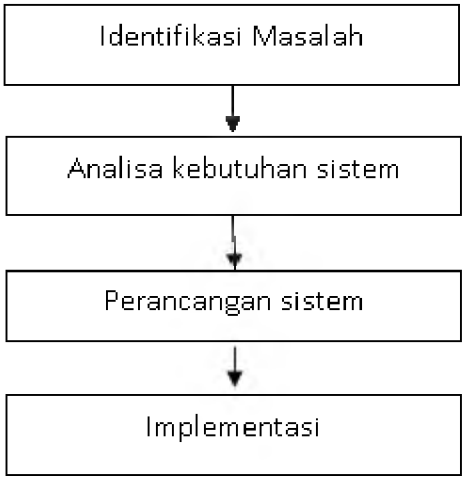
Analisis kebutuhan fungsional ini yaitu melakukan analisis fungsional dari sistem terhadap masalah perbandingan keakuratan nilai kesamaan atau kemiripan ciri suatu citra (pada kasus deteksi citra awan *Cumulonimbus* menggunakan Algoritma PCA) dengan beberapa metode jarak. Sedangkan analisis non fungsional yaitu menganalisis kebutuhan perangkat keras dan perangkat lunak yang digunakan pada penelitian ini.

c. Perancangan

Pada tahap perancangan ini dilakukan perancangan sistem menggunakan alat bantu Data Flow Diagram, Perancangan antarmuka.

d. Implementasi

Pada tahap implementasi ini dilakukan coding dan pengujian sistem, apakah telah sesuai dengan perencanaan, analisa dan perancangan yang telah diuraikan sebelumnya.



Gambar 3 Tahap-tahap penelitian

3.1 Analisis Kebutuhan Fungsional

Sistem akan melakukan ukuran jarak kedekatan antar elemen data piksel untuk mendapatkan nilai kesamaan atau kemiripan ciri citra apakah terdeteksi sebagai awan *Cumulonimbus* ataukah bukan *Cumulonimbus*. Pendeteksian kemiripan ciri citra awan dilakukan dengan beberapa metode jarak. Adapun beberapa metode jarak yang akan digunakan pada sistem ini diantaranya Jarak *Euclidean* (*Euclidean distance*), Jarak *Manhattan* (*Manhattan Distance*), Jarak *Mahalanobis* (*Mahalanobis Distance*), dan Jarak *Squared Euclidean* (*Squared Euclidean Distance*).

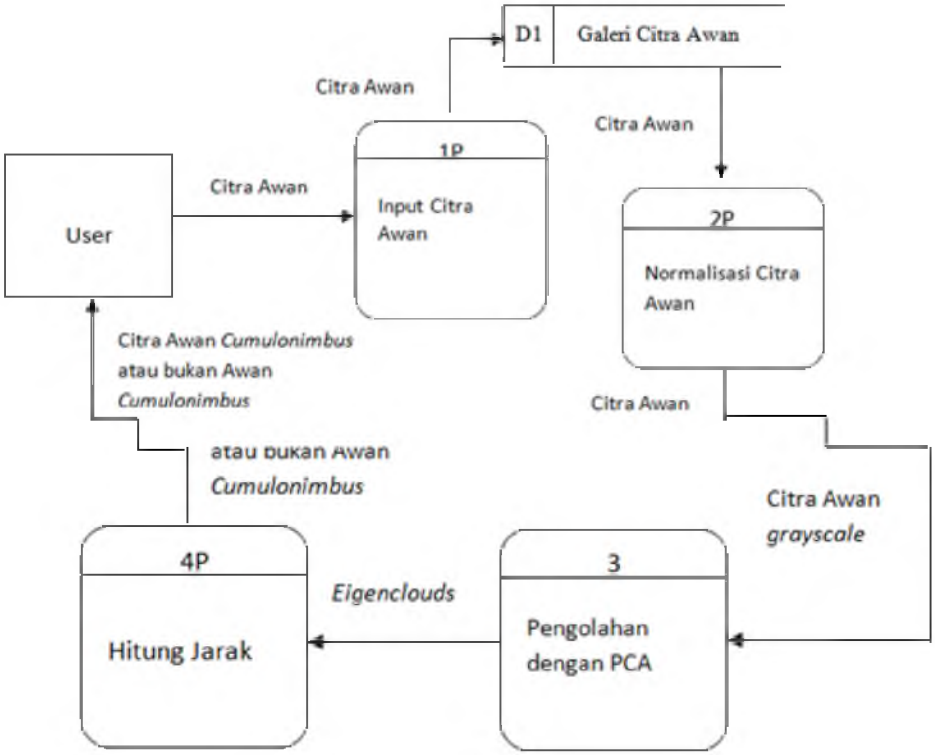
3.2 Perancangan Sistem

Perancangan sistem pada penelitian Metode Nilai Jarak Guna Kesamaan Atau Kemiripan Ciri Suatu Citra (Kasus Pada Deteksi Awan *Comulonimbus*) menggunakan

Diagram Alir Data yang terdiri dari beberapa level yaitu mulai dari level 0 sampai dengan level 2.

Diagram konteks/ diagram 0/ *High level* pada Gambar 4 menunjukkan bahwa Citra awan dalam bentuk digital diolah ke dalam sistem menggunakan satu persatu metode jarak guna memperoleh nilai kesamaan atau kemiripan ciri suatu citra awan. Hasil pemrosesan sistem akan diperoleh output hasil deteksi citra awan dalam bentuk teks, serta dari ouput tersebut dapat diihat perbandingan tingkat keberhasilan citra awan yang terdeteksi dari berbagai metode jarak yang digunakan.

Gambar 4. Diagram Konteks Metode Jarak pada suatu Citra (Kasus pada Deteksi Awan *Comulonimbus* Menggunakan Algoritma PCA)



Gambar 5. Diagram Level 0 tentang Metode Jarak pada suatuCitra (Kasus pada Deteksi Awan *Comulonimbus* Mengguna-kan Algoritma PCA)

Pada Gambar 5 terdapat 4 proses utama tersebut terdiri dari proses menginput citra awan (proses 1) yang dientrikan oleh user. Selanjutnya citra awan disimpan ke *database*.

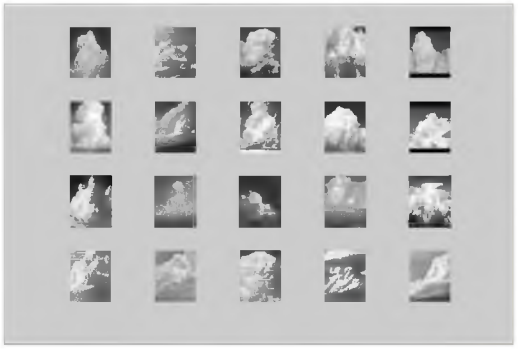
Sebelum citra awan di *training* dalam algoritma *Principal Componen Analysis* (PCA) diproses 3, terlebih dahulu citra awan dilakukan normalisasi. Pada proses normalisasi (proses 2) tersebut citra awan berwarna dikonversi dari citra awan berwara menjadi citra awan *grayscale* akan meringankan sistem dalam hal pembacaan citra. Selain citra diubah ke warna *grayscale* resolusi citra diturunkan menjadi ukuran yang sama yaitu 200 x 180 piksel. Setelah citra diolah oleh algoritma PCA, kedekatan jumlah nilai piksel citra yang disebut *eigenclouds* dilakukan perhitungan jarak yang diuji coba dengan beberapa metode perhitungan jarak. Hasil nilai perhitungan jarak ini dapat digunakan sebagai acuan apakah citra awan hasil *training* dengan citra input baru terdapat kemiripan fitur berdasarkan komponen-komponen utama yang dimiliki citra awan *Cumulonimbus* atau bukan *Cumulonimbus* dari kumpulan galery citra awan.

4. Implementasi

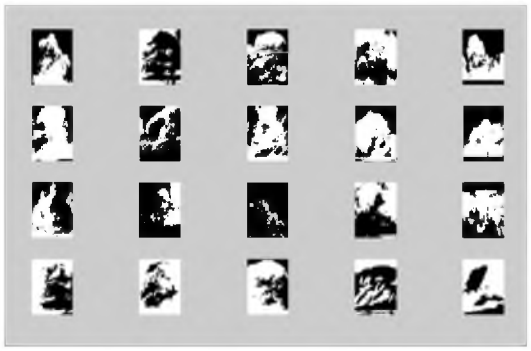
Implementasi pada aplikasi ini, dilakukan *training* sejumlah 40 buah citra awan *cumulonimbus* (Cb) dan kemudian akan dilakukan perhitungan jarak kemiripan fitur dengan metode jarak *Euclidean* (*Euclidean distance*), metode jarak *Manhattan* (*Manhattan Distance*), serta metode jarak *Mahalanobis* (*Mahalanobis Distance*).

4.1 Proses *Training* Citra Awan

Proses *Training* citra awan dilakukan dilakukan sebelum dilakukan proses deteksi. Proses *training* ini dilakukan beberapa tahapan diantaranya proses normalisasi, proses mengekstrak citra awan menggunakan PCA (*Principal Component Analysis*) sampai diperoleh fitur-fitur utama awan. Pada Gambar 6 merupakan *form* hasil proses normalisasi sebanyak 40 citra awan *Cumulonimbus*. Gambar 7 merupakan form hasil proses ekstraksi citra awan dari proses *training* dan citra yang dihasilkan dalam bentuk citra biner.

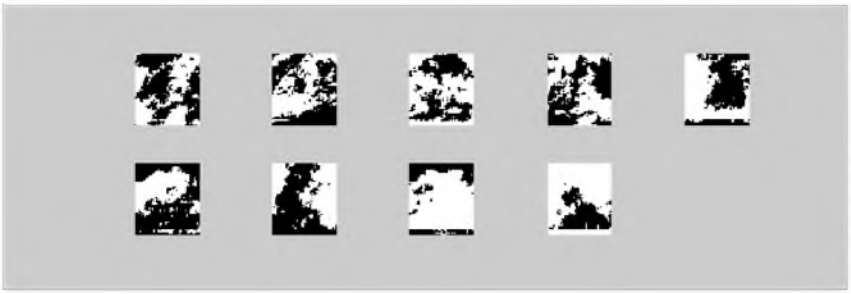


Gambar 6. Citra awan hasil proses normalisasi



Gambar 7. Citra awan yang dihasilkan dalam bentuk citra biner

Proses perhitungan nilai *eigen* dengan metode *Principal Component Analysis* (PCA) diperoleh nilai *eigenclouds,* yang dihasilkan dari e*igenvector* terpilih dan bersesuaian dengan *eigenvalue* tertinggi. Gambar 8 merupakan gambar citra awan *Cumulonimbus* hasil proses *training* dengan metode PCA untuk mengenalkan sistem atas ciri atau fitur utama citra awan *Cumulonimbus.*



Gambar 8. *Eigenclouds* citra awan *Cumulonimbus* yang dihasilkan setelah proses *training*

4.2 Pengujian dengan Metode Jarak *Euclidean, Manhattan* dan *Mahalanobis*

Sejumlah 40 buah citra awan *cumulonimbus* (Cb) dan citra awan bukan *cumulonimbus* (Cb) diujikan. Adapun hasil deteksi kemiripan citra awan dapat dilihat pada Tabel 1.

Tabel 1. Hasil deteksi kemiripan citra awan yang diuji dengan metode jarak *Euclidean, Manhattan, Mahalanobis*

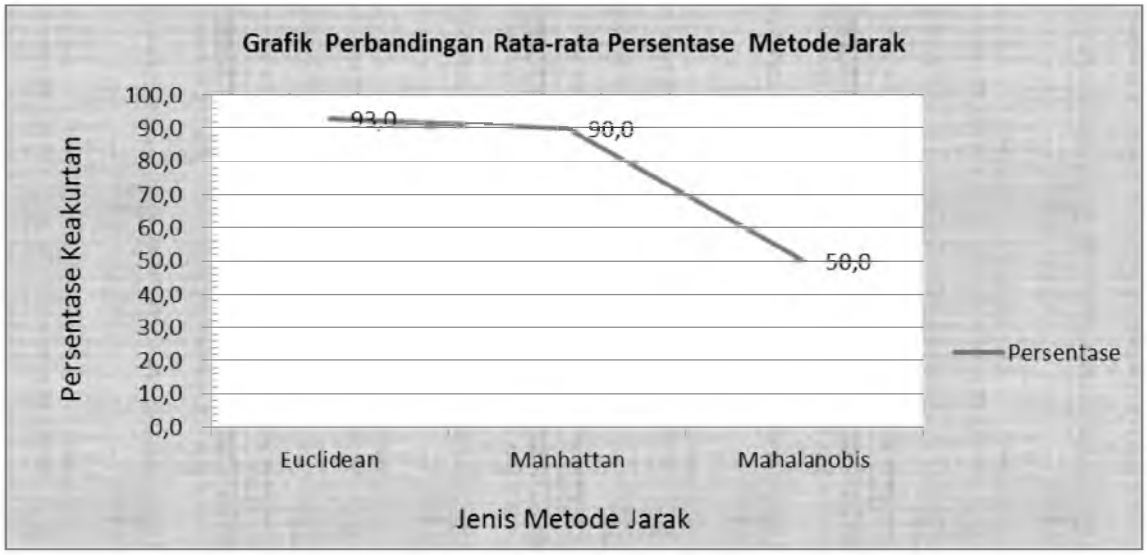
Metode Jarak	Jarak *Euclidean*		Jarak *Manhattan*		Jarak *Mahalanobis*	
Jenis awan	Cb	Bukan Cb	Cb	Bukan Cb	Cb	Bukan Cb
20 Cb	19	1	18	2	20	0
20 Bukan Cb	2	18	2	18	20	0

Hasil pengukuran akurasi dalam satuan persen (%) berdasarkan hasil pengujian di Tabel 1 bahwa kemiripan objek dengan metode Jarak *Euclidean* diperoleh awan *Cumulonimbus* 90% dan awan bukan *Cumulonimbus* 95%, serta rata-rata akurasi sebesar 93%. Semakin kecil nilai jarak, maka semakin mirip kedua vektor yang dicocokkan/dibandingkan. Sebaliknya semakin besar nilai jarak maka semakin berbeda kedua vektor yang dicocokkan (Budi Santosa., 2007).

Pengujian citra awan *Cumulonimbus* dan awan bukan *Cumulonimbus* dengan metode jarak *Manhattan*, ternyata memperoleh hasil yang hampir sama dengan menggunakan metode jarak *Euclidean.* Hasil pengukuran akurasi dalam satuan persen (%) untuk pengujian dengan metode Jarak *Manhattan* diperoleh awan *Cumulonimbus* 90% dan awan bukan *Cumulonimbus* 90%, serta rata-rata akurasi sebesar 90%.

Hasil pengukuran akurasi dalam satuan persen (%) untuk pengujian dengan metode Jarak *Mahalanobis* diperoleh awan *Cumulonimbus* 100% dan awan bukan *Cumulonimbus* 0%, serta rata-rata akurasi sebesar 50%. Pada jarak *Mahalanobis* bergantung pada besarnya *eigenvalues*, suatu citra dikatakan mirip jika *eigenvecto*r A > 0 (bernilai semakin besar).

Gambar 9 merupakan Grafik rata-rata hasil perbandingan persentase keakuratan citra awan *Cumulonimbus* dengan beberapa metode jarak.



Gambar 9. Grafik Perbandingan Persentase Keakuratan Metode Jarak pada Citra Awan *Cumulonimbus*

5. Kesimpulan

1. Rata-rata nilai persentase keakuratan kesamaan ataupun kemiripan citra awan *Cumulonimbus* menggunakan Algoritma PCA dengan metode jarak *Euclidean* adalah 93% dan metode jarak *Manhattan* menghasilkan rata-rata nilai persentase keakuratan 90% serta dengan metode jarak *Mahalanobis* yaitu 50%.

2. Pembacaan citra awan oleh sistem tidak terdeteksi sebagai citra awan *Cumulonimbus* jika citra awan *Cumulonimbus* berupa citra awan yang *blu*r (*noise*), hal ini disebabkan sistem belum diberi teknik *filtering*.

Daftar Pustaka

Artikelsiana, pengertian awan dan jenis-jenis-awan, http://www.artikelsiana.com, diakses tanggal 10 Juni 2015

Budi Santosa, 2007, Data Mining (Teori dan Aplikasi), Graha Ilmu, Yogyakarta.

Bajwa I.S., Naweed M.S., Asif M.N., Hyder.S.I., 2009, *Feature Based Image Classification by using Principal Component Analysis*, ICGST-GVIP Journal, ISSN 1687-398X, Vol. 9, Issue II.

Dwi Nugraheny, 2010, *Deteksi Awan Comulonimbus Menggunakan Principal Component Analysis (PCA)*, Prosiding Seminar Nasional Aplikasi Sains & Teknologi, hal A-77 ISSN No. 1979-911X.

Hair Jr., Joseph F., Black, William C., Babin, Barry C., dan Rolph E. Anderson, 2010, Multivariate Data Analysis 7/e, Pearson Prentice Hall, New Jersey.

Jenness Enterprises, 2008, *Mahalanobis Distance*, http://www.jennessent.com/arcview/mahalanobis_description.htm, diakses tanggal 03 Juli 2015.

Krueger J, et.al., 2004, Thresholds for Eigenface Recognition, http://cnx.org/content/m12533/1.2/, diakses tanggal 15 Juni 2015.

Mudrova M, et.al., 2002, Principal Component Analysis (PCA) in Image Processing, Institute of Chemical Technology, Prague Department of Computing and Control Engineering.

McAndrew, A, 2004, An Introduction to Digital Image Processing with Matlab, School of Computer Science and Mathematics Victoria University of Technology

Rancher, AC, 2004, Methods of Multivariate Analysis Second Edition, John Wiley & Sons, Canada.

Smith I Lindsay, 2002, A tutorial on Principal Component Analysis. Cornell University, USA. February.

http://www.free-pictures-photos.com/clouds, diakses tanggal 10 Juni 2015.

http://vortex.plymouth.edu/clouds.html, diakses tanggal 11 Juni 2015.