

PENGELOMPOKAN CITRA BATIK MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN FITUR TEKSTUR DENGAN METODE GLCM

Sulistiawan

Jurusan Teknik Informatika FIK UDINUS, Jl. Nakula No. 5-11 Semarang-50131

sulistiawann1@gmail.com

Abstrak - Pengelompokan citra batik dilakukan untuk mengelompokkan batik yang memiliki kemiripan dengan batik lainnya kedalam satu *cluster* dengan menggunakan algoritma *k-means* berdasarkan fitur tekstur citra batik yang diperoleh. Fitur-fitur tekstur dicari menggunakan metode *GLCM* berdasarkan sudut 0° , 45° , 90° dan 135° . Metode evaluasi *cluster* menggunakan *DBI* dan *purity* untuk mengukur seberapa bagus cluster yang dihasilkan. Dari hasil penelitian yang menggunakan 116 citra batik dengan 5 *cluster* menghasilkan nilai *DBI* untuk setiap sudut adalah 0,47891, 0,42772, 0,49173, 0,47152, serta nilai *purity* 0,4655, 0,5431, 0,5431, 0,5172. Ini membuktikan bahwa pengelompokan dengan *k-means* menggunakan fitur-fitur *GLCM* menghasilkan cluster yang masih belum cukup maksimal.

Kata kunci : Batik, K-means, Clustering, *gray level co-occurrence matrix (GLCM)*, *devian-bouldin index (DBI)*.

I. PENDAHULUAN

Batik merupakan salah satu warisan budaya Indonesia yang telah diakui oleh UNESCO pada tanggal 2 oktober 2009. Setiap batik memiliki motif serta keunikan tersendiri yang membedakannya dengan motif batik lainnya. Motif batik sendiri terbagi kedalam dua kelompok, yakni motif batik geometri dan non geometri. Banyaknya batik dengan motif atau pola yang beranekaragam membuat batik menjadi susah dikenali, sehingga dilakukan pengelompokan untuk mengenali batik dengan batik lainnya yang memiliki kemiripan dari segi motif [1] [2] [16] [17]. Pengelompokan citra batik berdasarkan daerah asal dengan penggunaan nama label yang sama dirasakan masih kurang efektif karena nama dari suatu citra batik belum tentu mencerminkan isi dari batik. Oleh karena itu dilakukan pengelompokan citra batik berdasarkan karakteristik yang dimiliki citra batik tersebut.

Pengelompokan dilakukan dengan membandingkan fitur-fitur yang diperoleh dari citra yang digunakan untuk mengelompokkan batik yang memiliki kemiripan dengan batik lainnya [3] [4] [5] [6].

Content based image retrieval (CBIR) merupakan metode yang dapat digunakan untuk mendapatkan informasi dari suatu citra serta menentukan kemiripan suatu citra. Untuk itu *gray level co-occurrence matrix (GLCM)* digunakan untuk memperoleh fitur-fitur tekstur dari citra [6] [7] [8] [11] [12], kemudian fitur-fitur tersebut di kelompokkan menggunakan *algoritma k-means* ke dalam suatu *cluster* yang memiliki kemiripan [10] [18] [24].

II. TEORI PENUNJANG

2.1 *Content Based Image Retrieval*

Content based image retrieval (CBIR) merupakan metode yang digunakan dalam

pencarian suatu gambar atau citra digital pada suatu database. Pencarian dilakukan dengan menganalisa bagian-bagian dari suatu citra untuk mendapatkan informasi citra tersebut atau biasa disebut dengan "Content based". *Image retrieval* sendiri merupakan proses yang digunakan untuk mendapatkan sejumlah informasi dari citra berdasarkan masukan atau inputan citra. Dalam penelitian ini, CBIR digunakan untuk mendapatkan fitur-fitur dari citra batik, fitur-fitur tersebut dihitung untuk menentukan kemiripannya dengan batik lainnya [5] [6].

2.2 Gray Level Co-occurrence Matrix

Gray level co-occurrence matrix (GLCM) merupakan matrik yang menggambarkan frekuensi munculnya pasangan piksel pada jarak d dan orientasi arah dengan sudut θ tertentu dalam citra yang digunakan untuk menghitung fitur-fitur glcm. Jarak d yang digunakan adalah 1 yang dinyatakan dalam piksel, sementara orientasi sudut dinyatakan dalam derajat dengan sudut 0° , 45° , 90° , dan 135° . Dalam metode glcm, terdapat beberapa langkah-langkah yang digunakan untuk menghitung fitur-fitur glcm dari citra *grayscale* yang digunakan, antara lain [6] [8] [9] [16] [18] [19] [20] [21] [22]:

- ❖ Membuat area kerja matrik dari citra batik.
- ❖ Menentukan hubungan spasial antara piksel referensi dengan piksel tetangga, dengan sudut θ dan jarak d .
- ❖ Menghitung jumlah kookurensi dan mengisikannya pada area kerja matrik.
- ❖ Menjumlahkan matrik kookurensi dengan transposenya untuk menjadikannya simetris.
- ❖ Dilakukan normalisasi matrik untuk mengubahnya kebentuk probabilitasnya.

- ❖ Menghitung nilai fitur-fitur ekstraksi dari normalisasi yang diperoleh.

Terdapat 5 fitur-fitur glcm yang digunakan dalam penelitian ini, antara lain :

1. Angular Second Moment (ASM)

ASM merupakan ukuran homogenitas dari suatu citra.

$$ASM = \sum_{i=1}^L \sum_{j=1}^L (GLCM(i, j))^2$$

2. Kontras

Kontras merupakan ukuran keberadaan variasi aras keabuan piksel citra.

$$Kontras = \sum_i^L \sum_j^L |i - j|^2 GLCM(i, j)$$

3. Inverse Different Moment (IDM)

IDM digunakan untuk mengukur homogenitas.

$$IDM = \sum_{i=1}^L \sum_{j=1}^L \frac{(GLCM(i, j))^2}{1 + (i-j)^2}$$

4. Entropi

Entropi menyatakan ukuran ketidakteraturan aras keabuan didalam citra.

$$Entropi = -\sum_{i=1}^L \sum_{j=1}^L (GLCM(i, j)) \log(GLCM(i, j))$$

5. Korelasi

Korelasi merupakan ukuran ketergantunganlinier antar nilai aras keabuan dalam citra.

$$Korelasi = \sum_{i=1}^L \sum_{j=1}^L \frac{(i-\mu_i')(j-\mu_j')(GLCM(i, j))}{\sigma_i \sigma_j}$$

Persamaan tersebut didapat dari mean yang merupakan nilai intensitas dari citra keabuan dan standart deviasi terlebih dahulu. Standart deviasi didapat dari akar kuadrat varian yang menunjukkan sebaran nilai *piksel* dalam citra, dengan rumus sebagai berikut:

$$mean\ i = \mu_i' = \sum_{i=1}^L \sum_{j=1}^L i * GLCM(i, j)$$

$$\text{mean } j = \mu_j' = \sum_{i=1}^L \sum_{j=1}^L j * GLCM(i, j)$$

$$\text{varian } i = \sigma_i^2 = \sum_{i=1}^L \sum_{j=1}^L GLCM(i, j) (i - \mu_i')^2$$

$$\text{varian } j = \sigma_j^2 = \sum_{i=1}^L \sum_{j=1}^L GLCM(i, j) (j - \mu_j')^2$$

$$\text{standart deviasi } i = \sigma_i = \sqrt{\sigma_i^2}$$

$$\text{standart deviasi } j = \sigma_j = \sqrt{\sigma_j^2}$$

2.3 Algoritma K-means

Algoritma k-means merupakan algoritma yang digunakan untuk pengelompokan iteratif, langkah-langkah dari metode *k-means* adalah sebagai berikut [8] [9] [10] [11] [12] [13] [14] [15] :

1. Menentukan jumlah cluster dan ambang batas perubahan fungsi objektif.
2. Menentukan centroid awal yang digunakan
3. Menghitung jarak setiap data ke masing-masing centroid menggunakan jarak *euclidean* untuk mendapatkan jarak terdekat data dengan *centroidnya*.
4. Menentukan centroid baru dengan menghitung nilai rata-rata dari data yang ada pada centroid yang sama.
5. Ulangi langkah 3 dan 4 hingga kondisi konvergen tercapai, yaitu perubahan fungsi objektif sudah dibawah ambang batas yang diinginkan, atau tidak ada data yang berpindah cluster, atau perubahan posisi centroid sudah dibawah ambang batas yang sudah ditentukan.

Perhitungan jarak antara data dan centroid dilakukan dengan menggunakan persamaan *euclidean distance*, persamaannya sebagai berikut [13] [14] [23] :

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

Dimana :

d_{ij} = Jarak objek antara objek i dan j

P = Dimensi data

x_{ik} = Koordinat dari obyek i pada dimensi k

x_{jk} = Koordinat dari obyek j pada dimensi k

2.4 Validasi Cluster

Validasi cluster dilakukan untuk mengetahui seberapa bagus cluster yang telah diperoleh dari proses *clustering* [23]. Metode yang digunakan untuk menentukan validitas cluster dalam penelitian ini menggunakan *davies-bouldin index (DBI)* dan *purity*.

1. Davies-Bouldin Index (DBI)

Davies-bouldin index merupakan salah satu metode yang digunakan untuk mengukur validitas cluster pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi. Langkah-langkah dalam menghitung DBI adalah sebagai berikut :

- ❖ Menghitung Sum of square within cluster (SSW) merupakan persamaan yang digunakan untuk mengetahui matrik kohesi dalam sebuah cluster ke-i.

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$$

- ❖ Menghitung Sum of square between cluster (SSB), merupakan persamaan yang digunakan untuk mengetahui separasi antar cluster.

$$SSB_{i,j} = d(c_i, c_j)$$

- ❖ Setelah nilai kohesi dan separasi diperoleh, kemudian dilakukan pengukuran rasio (R_{ij}) untuk mengetahui nilai perbandingan antara cluster ke-i dan cluster ke-j.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *davies-bouldin index* (DBI) dari persamaan berikut :

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$$

2. Purity

Purity digunakan untuk menghitung kemurnian dari suatu *cluster* yang direpresentasikan sebagai anggota *cluster* yang paling banyak sesuai (cocok) disuatu kelas [11] [24]. Untuk menghitung nilai purity setiap *cluster* dapat menggunakan rumus berikut :

$$Purity(j) = \frac{1}{n_j} \max(n_{ij})$$

Sementara untuk menghitung purity keseluruhan jumlah K *cluster*, digunakan persamaan sebagai berikut :

$$Purity = \sum_{i=1}^k \frac{n_j}{n} Purity(j)$$

III. HASIL & IMPLEMENTASI

Hasil pengujian dalam pengelompokan 116 data citra batik yang digunakan menunjukkan hasil yang berbeda-beda untuk ke-empat sudut θ yang digunakan, namun data dari *cluster* yang diikuti sebagian besar memiliki data yang sama meskipun terdapat beberapa data yang

berbeda. Jumlah data dari cluster yang diikuti untuk ke-empat sudut yang berbeda tersebut antara lain sebagai berikut :

Data Cluster	Sudut 0	Sudut 45	Sudut 90	Sudut 135
Cluster1	32	31	42	33
Cluster2	14	35	27	35
Cluster3	36	15	18	12
Cluster4	30	32	26	32
Cluster5	4	3	3	4

Data yang diperoleh tersebut merupakan hasil dari pengelompokan menggunakan algoritma *k-means* dari fitur-fitur tekstur citra batik yang diperoleh menggunakan metode *gray level co-occurrence matrix*.

Sedangkan hasil dari validasi *cluster* menggunakan *Davies-bouldin index* (DBI) dan *purity* dari data yang telah diperoleh dalam proses *clustering* didapatkan hasil sebagai berikut :

No	Sudut	DBI	Purity
1	0	0,4789	0,4655
2	45	0,4277	0,5431
3	90	0,4917	0,5431
4	135	0,4715	0,5172

IV. PENUTUP

Berdasarkan penelitian yang telah dilakukan oleh peneliti, dalam pengelompokan citra batik menggunakan algoritma *k-means* berdasarkan fitur-fitur tekstur *GLCM* menghasilkan nilai evaluasi *DBI* untuk setiap sudut adalah 0,47891, 0,42772, 0,49173 dan 0,47152, sedangkan nilai *purity* untuk setiap sudut adalah 0,4655, 0,5431, 0,5431 dan 0,5172. *Cluster* yang baik adalah *cluster* yang memiliki nilai *DBI* mendekati 0 dan nilai *purity* mendekati 1, sedangkan hasil dari penelitian yang diperoleh menunjukkan

nilai *DBI* dan *purity* yang masih cukup rendah untuk dapat dikatakan mendekati nilai 0 atau 1. Sehingga dapat dikatakan bahwa dalam proses *clustering* pada pengelompokan citra batik ini menghasilkan *cluster* yang masih belum cukup bagus. Hasil ini dapat dipengaruhi oleh beberapa faktor yakni, perbedaan kualitas gambar yang digunakan dalam penelitian ini serta jumlah *cluster* yang digunakan yang dapat mempengaruhi hasil akhir yang diperoleh.

DAFTAR PUSTAKA

- [1] A. Kurniawardhani, N. Suciati dan I. Arieshanti, "Klasifikasi Citra Batik Menggunakan Metode Ekstraksi Ciri Yang Invarian Terhadap Rotasi," *JUTI*, vol. 12, no. 2, pp. 48-60, Juli 2014.
- [2] B. Arisandi, N. Suciati dan A. Y. Wijaya, "Pengenalan Motif Batik Menggunakan Rotated Wavelet Filter dan Neural Network," *JUTI*, vol. 9, no. 2, pp. 13-19, Juli 2011.
- [3] A. A. Kasim dan A. Harjoko, "Klasifikasi Citra Batik Menggunakan Jaringan Syaraf Tiruan Berdasarkan Gray Level Co-Occurrence Matrices (GLCM)," Seminar Nasional Aplikasi Teknologi Informasi (SNATI), Yogyakarta, 21 Juni 2014.
- [4] A. Halim, Hardy, C. Dewi dan S. Angkasa, "Aplikasi Image Retrieval Menggunakan Kombinasi Metode Color Moment Dan Gabor Texture," *JSM STIMIK Mikrosil*, vol. 14, no. 2, Oktober 2013.
- [5] A. H. Rangkuti, "Content Based Batik Image Classification Using Wavelet Transform and Fuzzy Neural Network," *Journal of Computer Science*, vol. 10, no. 4, pp. 604-613, 2014.
- [6] A. Kadir dan A. Susanto, *Pengolahan Citra Teori dan Aplikasi*, Yogyakarta: Andi, Juni 2012.
- [7] A. A. Pratama, N. Suciati dan D. Purwitasari, "implementasi Fuzzy C-Means Untuk Pengelompokan Citra Batik Berdasarkan Motif Dengan Fitur Tekstur," *Jurnal TEKNIK POMITS*, vol. 1, no. 1, pp. 1-4, 2012.
- [8] A. E. Minarno dan N. Suciati, "Batik Image Retrieval Based on Color Difference Histogram and Gray Level Co-Occurrence Matrix," *TELKOMNIKA*, vol. 12, no. 3, pp. 597-604, September 2014.
- [9] H. Rangkuti, A. Harjoko dan A. E. Putro, "Content Based Batik Image Retrieval," *Journal of Computer Science*, vol. 10, no. 6, pp. 925-934, 2014.
- [10] D. H.B.Kekre, S. D. Thepade, T. K. Sarode dan V. Suryawanshi, "Image Retrieval using Texture Features extracted from GLCM, LBG and KPE," *International Journal of Computer Theory and Engineering*, vol. 2, no. 5, pp. 1793-8201, Oktober 2010.
- [11] R. Listia dan A. Harjoko, "Klasifikasi Massa pada Citra Mammogram Berdasarkan Gray Level Cooccurrence matrix (GLCM)," *IJCCS*, vol. 8, no. 1, pp. 59-68, Januari 2014.
- [12] S. K dan M. L, "An Efficient Image Retrieval Based on Color, Texture (GLCM & CCM) feature, and Genetic-Algorithm," *International Journal Of Merging Technology And Advanced Research In Computing*, pp. 1-9.
- [13] B. Sebastian, A. Unnikrishnan dan K. Balakrishnan, "Gray Level Co-occurrence Matrices : Generalitation and Some New Features," *International Journal of Computer Science, Engineering and Information Tecnology*, vol. 2, no. 2, pp. 151-157, April 2012.
- [14] P. Maheshwary dan N. Sricastava, "Prototype System for Retrieval of Remote Sensing Images Based on Color Moment and Gray Level Co-occurrence Matrix," *International Journal of Computer Science Issues*, vol. 3, pp. 20-23, 2009.
- [15] B. K. C. Ramamurthy, "Content Based Medical Image Retrieval with Texture Content Using Gray Level Co-occurrence

- Matrix and K-Means Clustering Algorithm,” *Journal of Computer Science*, vol. 8, no. 7, pp. 1070-1076, 2012.
- [16] M.-W. Lin, J.-R. Tapamo dan B. Ndovie, “A Texture-based Method For Document Segmentation and Classification,” *ARIMA/SACJ*, no. 36, pp. 49-56, 2006.
- [17] Widiarina dan R. S. Wahono, “Algoritma Cluster Dinamik untuk Optimasi Cluster pada Algoritma K-Means dalam Pemetaan Nasabah Potensial,” *Journal of Intelligent Systems*, vol. 1, no. 1, pp. 32-35, Februari 2015.
- [18] R. Handoyo, R. R. M dan S. M. Nasution, “Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K-Means pada Pengelompokan dokumen,” *JSM STMIK Mikrosil*, vol. 15, no. 2, pp. 73-82, Oktober 2015.
- [19] Sundar.C, M. ChitraDevi dan G. Geetharamani, “An Analisis on The Perfomance of K-Means Clustering Algorithm For Cardiotocogram Data Cluster,” *International Journal on Computational Sciences & Aplications*, vol. 2, no. 5, pp. 11-20, Oktober 2012.
- [20] T. Khotimah, “Pengelompokan Surat Dalam Al Qur'an Menggunakan Algoritma K-Means,” *Jurnal SIMETRIS*, vol. 5, no. 1, pp. 83-88, April 2014.
- [21] J. O. Ong, “Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University,” *Jurnal Ilmiah Teknik Industri*, vol. 12, no. 1, pp. 10-20, Juni 2013.
- [22] M. H. Noor dan M. Hariadi, “Image Cluster Berdasarkan Warna Untuk Identifikasi Kematangan Buah Tomat Dengan Metode Valley Tracing,” *Seminar Nasional Informatika*, pp. 15-24, 23 Mei 2009.
- [23] E. Prasetyo, *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*, Yogyakarta: Andi, September 2014.
- [24] K. R. Prilianti dan H. Wijaya, “Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering,” *Jurnal Cybermatika*, vol. 2, no. 1, Juni 2014.