

PENGELOMPOKAN JENIS KUPU-KUPU MENGGUNAKAN FITUR EKSTRAKSI GLCM DAN ALGORITMA K-MEANS

Tommy Kusuma
Jurusan Teknik Informatika, Fakultas Ilmu Komputer
Universitas Dian Nuswantoro
Email : tommykusuma1993@gmail.com

Abstrak— Kupu – kupu merupakan fauna yang hampir mengalami kepunahan. Kepunahan ini disebabkan oleh banyak faktor, mulai dari faktor alam sampai faktor manusia. Tujuan penelitian ini dilakukan agar dapat mengenali dan mengelompokkan jenis kupu – kupu melalui media gambar. Metode yang akan digunakan dalam penelitian ini adalah metode GLCM dan metode K-Means. Metode GLCM dapat digunakan untuk menganalisa dan mengenali suatu gambar dengan menggunakan tekstur dari sebuah gambar. Sedangkan untuk melakukan pengelompokan dari gambar, menggunakan metode K-Means. Pada penelitian ini penulis menggunakan 10 jenis gambar dataset spesies kupu – kupu. Sehingga akan terbentuk 10 kluster pada metode K-Means. Hasil yang diharapkan penulis adalah dapat mengenali dan mengelompokkan gambar kupu – kupu sesuai dengan jenisnya. Hasil yang didapat mempunyai tingkat akurasi paling tinggi sebesar 24,76%. Saran pada penelitian selanjutnya supaya peneliti menggunakan resolusi gambar yang sama dan sudut dari pengambilan gambar tidak berbeda – beda.

Kata Kunci: Image Processing GLCM K-Means

1. Pendahuluan

Image retrieval merupakan ilmu yang digunakan untuk melakukan proses pencarian gambar berdasarkan fitur gambar secara otomatis. Pencarian gambar ini dilakukan dengan cara melakukan pencarian informasi dan keterangan – keterangan dari setiap gambar. Proses pencarian informasi dan keterangan pada gambar, supaya menghasilkan akurasi yang tinggi maka diperlukan sumber gambar yang banyak [1]. Pada penelitian ini penulis menggunakan dataset gambar kupu – kupu. Dataset gambar kupu – kupu ini yang akan digunakan untuk mengelompokkan jenis kupu – kupu yang dilindungi atau yang memiliki jumlah spesies

kupu – kupu sedikit. Dari gambar kupu – kupu tersebut nantinya akan dapat diketahui jenis dan spesies dari kupu – kupu tersebut. Jadi dapat membantu mengidentifikasi peternak kupu – kupu dengan cara jenis kupu – kupu yang dikategorikan sebagai fauna yang dilindungi.

Fitur merupakan karakteristik dari suatu gambar. Setiap gambar atau objek pasti mempunyai ciri – ciri yang berbeda, sehingga antara satu gambar dengan gambar lain tidak mungkin mempunyai kesamaan fitur. Dikarenakan fitur merupakan ciri utama dari sebuah gambar atau objek yang sifatnya sebagai tanda [2].

Penelitian ini akan menggunakan algoritma Gray Level Co ocured Matrix atau disingkat GLCM sebagai fitur *ekstraksi* dari dataset gambar kupu - kupu. Diperlukan fitur *ekstraksi* dalam melakukan pengelompokan pada gambar. Hal ini sangat diperlukan agar hanya fitur yang penting saja yang akan tetap disimpan yang akan digunakan untuk diolah dan dilakukan proses, dan fitur yang tidak digunakan akan dibuang. Fitur yang tidak digunakan dibuang dikarenakan dapat memperlambat fitur melakukan proses. Kecepatan sangat berpengaruh terhadap fitur *ekstraksi*.

Setelah dataset gambar kupu – kupu tersebut diterapkan fitur *ekstraksi* GLCM, maka dari setiap gambar pada dataset gambar kupu – kupu akan dilakukan pengelompokan. Pada penelitian ini algoritma yang digunakan penulis untuk melakukan pengelompokan dataset gambar kupu – kupu adalah algoritma K-Means [3].

2. Metode Penelitian

2.1 GLCM

GLCM (Gray Level Cooccurrence Matrices), GLCM pertama kali diperkenalkan oleh Haralick. GLCM merupakan sebuah fitur ekstraksi yang digunakan untuk melakukan ekstraksi pada gambar. Ekstraksi ini dilakukan dengan cara merubah gambar menjadi grayscale. Grayscale sendiri merupakan sebuah pengolahan gambar berwarna yang kemudian dikonversi menjadi gambar keabu – abuan. Dalam metode GLCM derajat keabu – abuan sebuah citra dengan ukuran tertentu direpresentasikan sebagai sebuah matrik co-occurrence. Matrik co-occurrence merupakan matrik yang digunakan untuk melakukan proses perhitungan pada metode GLCM [3].

Berikut ini merupakan algoritma dasar dari GLCM [12]:

1. Hitung semua pasangan piksel dimana piksel pertama memiliki nilai i , dan pasangan piksel yang cocok dari piksel pertama dengan d memiliki nilai j .
2. Hitung dalam baris i dan kolom j dari matriks $P_d [i, j]$.
3. Perhatikan bahwa $P_d [i, j]$ tidak simetris, karena jumlah pasangan piksel yang memiliki tingkat abu-abu $[i, j]$ tidak selalu sama jumlah pasangan piksel yang memiliki tingkat abu-abu $[j, i]$.
4. Elemen $P_d [i, j]$ dapat dinormalisasi dengan membagi setiap entri dengan jumlah total pasangan piksel.
5. Normalisasi GLCM $N [i, j]$, didefinisikan oleh:

$$N[i, j] = \frac{P[i, j]}{\sum_i \sum_j P[i, j]} \quad (2.1)$$

Didalam GLCM terdapat beberapa fitur yang dapat digunakan untuk melakukan proses ekstraksi, seperti contoh fitur *contras*, *correlation*, *energy*, *homogeneity* dan lain sebagainya. Pada setiap fitur tersebut mempunyai rumus atau perhitungan sesuai dengan fitur yang akan dipergunakan. Dibawah ini akan diberikan rumus – rumus yang akan digunakan oleh penulis untuk melakukan penelitian [13].

$$CONTRAS = \sum \sum (i - j)^2 P(i, j) \quad (2.2)$$

$$ENERGY = \sum \sum P^2(i, j) \quad (2.3)$$

$$COR = \frac{\sum_i \sum_j (ij) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (2.4)$$

$$HOM = \sum_i \sum_j \frac{1}{1 + (i - j)^2} P_{i, j} \quad (2.5)$$

2.2 Kmeans

Clustering merupakan pengelompokan data, hasil dari observasi dan kasus ke dalam kelas yang hampir menyerupai. *Cluster* sendiri merupakan koleksi data yang mirip antara satu data dengan data yang lain, dan *cluster* memiliki perbedaan bila dibandingkan dengan data dari *cluster* yang lain [14].

K-Means merupakan salah satu contoh dari kelompok algoritma *clustering*. K-Means pertama kali dipublikasikan oleh Stuart Lloyd pada tahun 1984 dan merupakan algoritma *clustering* yang banyak digunakan. Algoritma K-Means ini dilakukan dengan cara berulang – ulang untuk memisahkan data dengan optimal dan memaksimalkan hasil dari partisi sehingga tidak ada perubahan data didalam setiap segmentasi. K-Means merupakan algoritma yang melakukan pendekatan *top-down*, karena sudah memulai dengan melakukan segmentasi yang sudah ditentukan terlebih dahulu [15].

Berikut ini merupakan alur algoritma K-Means:

1. Pemilihan acak K , K merupakan banyaknya *cluster* yang ingin dibentuk.
2. Tetapkan nilai K secara acak atau *random* sering disebut dengan nama centroid.
3. Hitung setiap jarak data yang ada terhadap masing – masing centroid, menggunakan rumus Euclidian dan City Block hingga ditemukan jarak dari setiap data dengan centroid. Berikut ini merupakan rumus Euclidian Distance dan City Block:

$$Euclidian = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (2.6)$$

$$City\ Block = \sum_{i=1}^n |p_i - q_i| \quad (2.7)$$

4. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid.
5. Lakukan langkah tersebut hingga nilai centroid tidak berubah atau stabil.

2.3 Pengukuran Kinerja Clustering

Metode *Sum Squared Error (SSE)* merupakan salah satu metode yang terdapat di dalam statistik. Metode SSE ini dipergunakan untuk mengukur selisih total dari nilai yang sebenarnya terhadap nilai yang telah dicapai. Istilah *SSE* sering disebut juga sebagai *Summed Square of Residuals*. Berikut ini merupakan rumus dari SSE:

$$SSE = \sum_{x \in \omega_k} |x - \mu(\omega_k)|^2 \quad (2.8)$$

Dimana nilai x adalah nilai data yang berada didalam kluster, sedangkan μ adalah nilai dari centroid, sedangkan ω adalah nilai dari cluster. Nilai *SSE* yang mendekati angka 0 menandakan bahwa model tersebut mempunyai komponen kesalahan acak terkecil dan nilai tersebut akan lebih berguna untuk peramalan terhadap suatu model yang diamati. Sebagai catatan bahwa sebelumnya *SSE* didefinisikan dalam metode kelayakan kuadrat minimum [16].

Metode Purity digunakan sebagai evaluasi kinerja algoritma clustering. Purity akan menghasilkan persentase akurasi kinerja clustering. Berikut ini merupakan rumus dari purity:

$$Purity = \frac{1}{N} \sum_k \max_j |\omega_k \cap c_j| \quad (2.9)$$

Dimana cara kerja purity ini adalah mengambil nilai maksimal dari elemen cluster yang terbentuk dan kemudian nilai perwakilan cluster tersebut akan dijumlahkan dengan elemen cluster yang terbentuk dari cluster yang lain sebagai perwakilan setiap cluster, Setelah itu dibagi dengan jumlah seluruh elemen data [17].

3. Hasil dan Pembahasan

Dibawah ini merupakan tabel hasil SSE dan Purity dari penelitian

Distance Degree	Normal			
	Euclidian		City Block	
	SSE	Purity	SSE	Purity
D=1 $\theta=0$	6,416	19,231%	96,637	19,351%
D=1 $\theta=45$	11,527	18,269%	123,506	19,351%
D=1 $\theta=90$	7,709	18,510%	100,652	18,750%
D=1 $\theta=135$	9,873	18,510%	128,020	18,990%
D=2 $\theta=0$	14,015	19,591%	151,795	19,351%
D=2 $\theta=45$	20,085	19,111%	182,121	18,750%

D=2 $\theta=90$	15,948	19,231%	152,702	19,351%
D=2 $\theta=135$	19,548	18,149%	179,359	18,029%

Distance Degree	Masking			
	Euclidian		City Block	
	SSE	Purity	SSE	Purity
D=1 $\theta=0$	4,692	21,875%	88,070	21,394%
D=1 $\theta=45$	8,069	20,673%	111,962	22,716%
D=1 $\theta=90$	5,321	22,115%	88,480	22,476%
D=1 $\theta=135$	7,606	20,670%	107,717	21,514%
D=2 $\theta=0$	10,402	22,837%	130,136	22,115%
D=2 $\theta=45$	15,631	23,438%	166,341	22,356%
D=2 $\theta=90$	12,395	22,837%	139,942	24,760%
D=2 $\theta=135$	15,857	22,957%	164,073	22,957%

Dari tabel penelitian diatas dapat diambil kesimpulan bahwa metode masking dapat meningkatkan nilai dari SSE dan Purity.

4. Kesimpulan

Pada penelitian ini dapat diambil kesimpulan bahwa pengelompokan jenis kupu – kupu dengan menggunakan fitur ekstraksi algoritma GLCM dan algoritma K-Means agar dapat dikenali oleh komputer menghasilkan nilai akurasi pada penelitian ini masih belum sesuai dengan apa yang diharapkan oleh penulis, dikarenakan nilai akurasi yang masih rendah. Adapun peningkatan nilai hasil akurasi dengan melakukan penghilangan background kupu kupu. Akan tetapi walaupun Pada penelitian ini dihasilkan nilai akurasi rendah kemungkinan yang disebabkan oleh 2 faktor:

1. Resolusi gambar kupu - kupu yang berbeda beda.
2. Bentuk dan sudut pengambilan gambar kupu – kupu berbeda beda.

5. Daftar Pustaka

- [1] V. G.Tonge, "Content Based Image Retrieval by K-Means Clustering Algorithm," p. 4, 2011.
- [2] A. Z. Arifin, B. Bagus dan D. A. Navastara, "Klasifikasi Online Citra Daun," p. 6, 2010.
- [3] Z. Budiarmo, "Identifikasi Macan Tutul Dengan Metode Grey Level Coocurent Matrix (GLCM)," p. 6, 2010.

- [4] N. B. S. R dan B. R, "Wood Species Recognition Using GLCM and Correlation," *Advances in Recent Technologies in Communication and Computing*, p. 5, 2009.
- [5] A. A. Kusuma, R. R. Isnanto dan I. Santoso, "Pengenalan Iris Mata Menggunakan Pencirian Matriks Ko-Okurensi Aras Keabuan," Skripsi Teknik Elektro Universitas Diponegoro, Semarang, 2009.
- [6] N. Zulpe dan V. Pawar, "GLCM Textural Features for Brain Tumor Classification," *IJCSI International Journal of Computer Science*, vol. 9, no. 3, p. 6, 2012.
- [7] U. G. S. (USGS), "Butterflies and Moths of North America," 2014. [Online]. Available: <http://www.butterfliesandmoths.org>.
- [8] S.Arivazhagan, R. Shebiah, S. Nidhyandhan dan L.Ganesan, "Fruit Recognition using Color and Texture Features," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 1, no. 2, p. 5, 2010.
- [9] P. Mohanaiah, P. Sathyanarayana dan L. GuruKumar, "Image Texture Feature Extraction Using GLCM Approach," *International Journal of Scientific and Research Publications*, vol. 3, no. 5, p. 5, 2013.
- [10] M. Singha dan K.Hemachandran, "Content Based Image Retrieval using Color and Texture," *Signal & Image Processing : An International Journal (SIPIJ)* , vol. 3, no. 1, p. 19, 2012.
- [11] R. Kusumaningrum dan A. M. Arymurthy, "Color and Texture Feature for Remote Sensing – Image Retrieval System: A Comparative Study," *IJCSI International Journal of Computer Science Issues*, vol. 8, no. 5, p. 11, 2011.
- [12] R. M. Hamza dan D. T. A. Al-Assadi, "Genetic Algorithm to find optimal," p. 16, 2010.
- [13] D. H.B.Kekre, S. D. Thepade, T. K. S. Sarode dan V. S. Suryawanshi, "Image Retrieval using Texture Features extracted," *Computer Theory and Engineering*, vol. 2, p. 6, 2010.
- [14] R. Shrivastava, K. Upadhyay dan R. Bhati, "Comparison between K-Mean and C-Mean Clustering for CBIR," dalam *Second International Conference on Computational Intelligence, Modelling and Simulation*, India, 2010.
- [15] T. Rismawan dan S. Kusumadewi, "Aplikasi K-Means Untuk Pengelompokan Mahasiswa," dalam *Aplikasi Teknologi Informasi*, Yogyakarta, 2008.
- [16] S. Oktavia, M. N. Mara dan N. Satyahadewi, "Pengelompokan Kinerja Dosen Jurusan Matematika FMIPA UNTAN," *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*, vol. 2, no. 2, p. 7, 2013.
- [17] M. Deepa dan P. Revathy, "Validation of Document Clustering based on Purity and Entropy," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 1, no. 3, p. 6, 2012.