

Implementasi Metode SVM untuk Klasifikasi Bunga dengan Ekstraksi Fitur Histogram of Gradient (HOG)

Henny Leidiyana ^{1,*}, Joni Warta ²

¹ Fakultas Teknik Informatika, Prodi Sistem Informasi; Universitas Bina Sarana Informatika; Jl. Kramat Raya No.98, Kwitang, Kec. Senen, Kota Jakarta Pusat 10450, Telp. 021-8000063; e-mail: henny.hnl@bsi.ac.id

² Informatika; Universitas Bhayangkara Jakarta Raya; Jl. Perjuangan No.81 RT.003/RW.002, Marga Mulya, ec. Bekasi Utara, Kota Bks, Jawa Barat 17143; e-mail: joniwarta@dsn.ubharajaya.ac.id

* Korespondensi: e-mail: henny.hnl@bsi.ac.id

Diterima: 20 Juli 2022 ; Review: 30 Juli 2022; Disetujui: 31 Juli 2022; Diterbitkan: 31 Juli 2022

Abstract

Feature extraction techniques are applied to obtain features that will be useful in classifying and recognizing images. Feature extraction techniques are very helpful in various image processing applications. Many studies show the effectiveness of feature extraction before classification. There is also research showing that a feature extraction method may be good for certain classification approaches but may also be preferable. The impact of the sample and its size can also determine the results of the application of feature extraction techniques in the classification process. In this study, the author aims to prove the effectiveness of the application of the HOG feature extraction technique on the classification with the SVM method on flower images. The experiment was carried out on two groups of images, where the first group was image classes with relatively uniform colors and shapes, both in shape and color. and the second group is image classes with relatively different colors and shapes in the same class. The results showed that image datasets with relatively uniform colors and shapes do not require the application of any feature extraction to produce high accuracy. For classification by performing feature extraction in this study it gives different results for the two interest groups.

Keywords: Feature extraction, HOG, SVM

Abstrak

Teknik ekstraksi fitur diterapkan untuk mendapatkan fitur yang akan berguna dalam mengklasifikasikan dan mengenali citra. Teknik ekstraksi fitur sangat membantu dalam berbagai aplikasi pengolahan citra. Banyak penelitian menunjukkan efektifitas ekstraksi fitur sebelum klasifikasi. Terdapat juga penelitian yang menunjukkan bahwa suatu metode ekstraksi fitur mungkin baik untuk pendekatan klasifikasi tertentu tetapi mungkin juga sebaiknya. Dampak sampel dan ukurannya juga bisa menentukan hasil penerapan teknik ekstraksi fitur dalam proses klasifikasi. Dalam penelitian ini peneliti bertujuan untuk membuktikan efektifitas penerapan Teknik ekstraksi fitur HOG pada klasifikasi dengan metode SVM terhadap citra bunga. Percobaan dilakukan pada dua kelompok citra, di mana kelompok pertama adalah kelas-kelas citra dengan warna dan bentuk yang relative seragam baik bentuk maupun warnanya. dan kelompok kedua adalah kelas-kelas citra dengan warna dan bentuk yang relative berbeda dalam kelas yang sama. Hasil penelitian diketahui bahwa *dataset* citra dengan warna dan bentuk yang *relative* seragam tidak perlu penerapan ekstraksi fitur apapun dapat menghasilkan akurasi yang tinggi. Untuk

klasifikasi dengan melakukan ekstraksi fitur dalam penelitian ini memberikan hasil yang berbeda untuk dua kelompok bunga.

Kata kunci: *Feature extraction, HOG, SVM*

1. Pendahuluan

Banyak tanaman bunga yang merupakan tanaman hias yang cocok sebagai dekorasi internal maupun eksternal ruangan. Beberapa tanaman hias juga memiliki nilai ekonomis tinggi yang dapat menjadi sumber investasi dan menambah penghasilan. Namun penentuan jenis bunga memerlukan keahlian. Identifikasi spesies bunga merupakan hal yang tidak mudah bagi masyarakat umum bahkan mungkin untuk para ahli yang menangani masalah tanaman setiap hari. Bagian tanaman yang paling menonjol dan dapat di lihat secara visual adalah bunganya, subjek penelitian intensif oleh ahli botani dan menjadi kunci untuk identifikasi spesies. Bunga memiliki keragaman dalam warna, bentuk, dan tekstur sehingga memungkinkan untuk menggunakan serangkaian metode yang dikembangkan untuk tugas klasifikasi objek (Wäldchen et al., 2018).

Tekstur dapat digunakan untuk memisahkan wilayah yang diinginkan dalam sebuah gambar. Tekstur juga merupakan salah satu fitur visual tingkat rendah yang paling banyak digunakan yang mengacu pada sifat permukaan bawaan suatu objek dan hubungannya dengan sekitarnya dan juga berisi informasi penting tentang susunan struktural permukaan dan hubungannya dengan sekitarnya (Gandhani & Singhal, 2015).

Klasifikasi dengan metode SVM terhadap objek bunga sudah pernah dilakukan (M.-E. Nilsback & Zisserman, 2008; M. E. Nilsback & Zisserman, 2006; Wäldchen et al., 2018) dengan mengkombinasikan teknik ekstraksi fitur. Penelitian terkait dengan ekstraksi menggunakan fitur *HOG* terhadap tanaman juga sudah pernah dilakukan dalam beberapa penelitian seperti klasifikasi varietas kacang tanah (Li et al., 2018), identifikasi spesies tanaman melalui daun (Bao et al., 2020; Liu et al., 2020; Sharma et al., 2019).

Penelitian bidang lain yang membuktikan bahwa feature *HOG* memberikan pengaruh yang cukup besar yaitu deteksi pejalan kaki (Dalal & Triggs, 2005), dan kombinasi *HOG+SVM* yang memberikan hasil yang cepat dan efektif pada penelitian deteksi tangan real-time (Guo et al., 2013).

Tujuan dari percobaan ini untuk membuktikan efektifitas ekstraksi fitur *HOG* dalam meningkatkan akurasi pada citra bunga. Percobaan dilakukan pada dua kelompok citra, dimana kelompok pertama adalah kelas-kelas citra dengan warna dan bentuk yang berbeda dalam kelas yang sama dan kelompok kedua adalah kelas-kelas citra dengan warna dan bentuk yang relative seragam baik bentuk maupun warnanya. Untuk tiap kelompok dilakukan klasifikasi dengan dan tanpa penerapan ekstraksi fitur *HOG*.

2. Metode Penelitian

Untuk menjelaskan metode penelitian yang memadukan teori, fakta, dan kajian pustaka untuk menjadi dasar dalam penelitian ini maka peneliti menuliskan tahapan penelitian yang di gambarkan dalam bentuk diagram pada Gambar 1.

2.1. Tahapan penelitian

Seperti ditunjukkan pada Gambar 1, tahapan penelitian ini dibagi menjadi empat yaitu:

a. *Preprosesing*

Langkah awal yaitu pelabelan terhadap *dataset* citra bunga dengan memisahkan ke dalam folder-folder kelas bunga. Untuk percobaan pertama yaitu mengklasifikasi citra dengan SVM, citra di ubah ke dalam dimensi untuk masukan proses klasifikasi. Untuk percobaan kedua, untuk penerapan ekstraksi fitur *HOG*, citra di ubah ukurannya menjadi 64x128 (mengikuti arahan penemu fitur *HOG*).

b. Ekstraksi Fitur

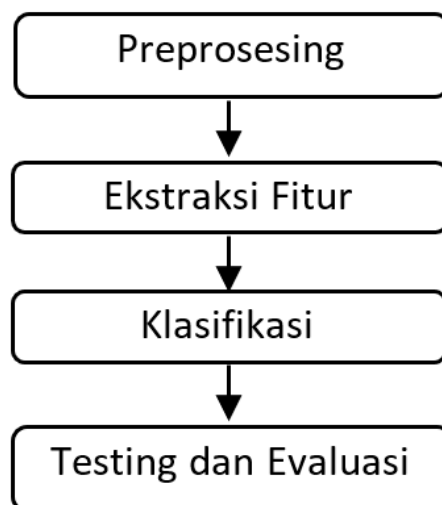
Pada percobaan kedua dilakukan ekstraksi fitur dengan Teknik *Histogram of Gradient (HOG)* dari citra yang telah diubah ukurannya.

c. Klasifikasi

Untuk kedua percobaan, dilakukan proses klasifikasi dengan SVM, sebelumnya data dibagi menjadi data training dan data testing. Untuk percobaan pertama, klasifikasi dilakukan dengan data yang telah diubah ke dalam format dimensi satu sedangkan percobaan kedua klasifikasi dilakukan dengan data.

d. Testing dan evaluasi

Testing dan Evaluasi dilakukan untuk mengukur model klasifikasi yang diperoleh. Evaluasi pada penelitian ini menggunakan confusion matrix (Sokolova & Lapalme, 2009). Confusion matrix adalah informasi mengenai hasil klasifikasi aktual yang dapat diprediksi oleh sebuah sistem klasifikasi.



Sumber: Hasil Penelitian (2022)

Gambar 1. Tahapan Penelitian

2.2. Dataset

Dataset diperoleh dari www.kaggle.com yaitu salah satu *dataset* yang bersifat *public* yang merupakan hasil foto oleh Olga Belitskaya di mana citra dipilih untuk mencerminkan fitur berbunga dari spesies tanaman ini. Bunga di kelompokkan menjadi dua, di mana kelompok pertama adalah kelas-kelas citra dengan warna dan bentuk yang relatif seragam baik bentuk maupun warnanya. dan kelompok kedua adalah kelas-kelas citra dengan warna dan bentuk yang relatif berbeda dalam kelas yang sama.

Aquilegia



Iris



Rhododendron



Viola



Sumber: Hasil Penelitian (2022)

Gambar 2. Bunga Kelompok Pertama

Calendula



Goldquelle



Shastadaisy



Veronica Chamaedrys



Sumber: Hasil Penelitian (2022)

Gambar 3. Bunga Kelompok Kedua

2.3. Ekstraksi Fitur

HOG bekerja dimana intensitas cahaya dari gradien yang didistribusikan secara lokal cocok untuk mengkarakterisasi garis dan bentuk objek. Citra di bagi menjadi sel yang mewakili daerah yang dibandingkan dan di bagi secara merata. Daerah-daerah ini dapat diklasifikasikan ke dalam blok-blok yang kemudian di normalisasi agar tetap konsisten terhadap iluminasi atau

efek fotometrik. Fitur yang di ekstraksi di atas blok di hitung untuk normalisasi. Blok-blok ini disebut *Histogram of Oriented Gradient (HOG)* (Sharma et al., 2019).

HOG memberikan kinerja yang sangat baik dibandingkan dengan fitur lain yang ada. Hipotesis dasarnya adalah bahwa penampilan dan bentuk objek lokal seringkali dapat dicirikan dengan baik oleh distribusi gradien intensitas lokal atau arah tepi, bahkan tanpa pengetahuan yang tepat tentang gradien atau posisi tepi yang sesuai (Cao et al., 2011; Dalal & Triggs, 2005). Tujuan *HOG* adalah untuk menggambarkan gambar dengan histogram gradien berorientasi lokal. Histogram ini mewakili kemunculan orientasi gradien tertentu di bagian lokal gambar (Cheon et al., 2012). Ekstraksi fitur *HOG* terdiri dari beberapa langkah penting. Langkah dari *HOG Feature Descriptor* adalah (Dalal & Triggs, 2005):

a. *Preprocessing* (Normalisasi dan Pengubahan Ukuran Gamma/Warna)

Pada langkah ini mengubah ukuran citra. Untuk deskriptor fitur *HOG*, ukuran gambar yang paling umum adalah 64x128 (lebar x tinggi) piksel atau dapat menggunakan ukuran citra dengan rasio 1:2 seperti 128x256 atau 256x512.

b. Menghitung Gradien

Langkah selanjutnya adalah menghitung gradien gambar dengan menghitung gradien sehingga tampak lokasi-lokasi di mana intensitas gradien piksel berubah.

c. *Spatial / Orientation Binning* (Membagi gambar menjadi sel).

Langkah selanjutnya membagi citra menjadi 8x8 sel dan hitung gradien untuk semua sel 8x8. Jadi, jika untuk citra 128x256 akan ada 8 sel pada arah horizontal pada tiap baris dan 32 sel pada arah vertikal untuk setiap kolom. Setiap sel memiliki 8x8x3 = 192 piksel, dan gradien setiap sel memiliki besar dan arah (2 nilai) sehingga setiap sel memiliki nilai 8x8x2 = 128 sebagai informasi gradien. Gradien dan arah masing-masing 8x8 blok dinyatakan dalam 9 balok histogram

d. Normalisasi Blok.

Kelompokkan 4 sel menjadi satu blok (2x2) normalisasi kemudian bandingkan blok sel tetangga untuk menekan efek iluminasi di seluruh gambar.

e. Dapatkan Vektor Fitur *HOG*.

Setelah menghitung semua normalisasi blok, gabungkan menjadi satu vektor untuk mendapatkan vektor fitur satu dimensi dari informasi di setiap sel.

2.4. Klasifikasi dengan SVM

Dalam pembelajaran mesin, SVM adalah model termasuk dalam supervised learning yang dapat menganalisis data dan mengenali pola untuk proses klasifikasi (Bao et al., 2020). Setiap item data di *plot* sebagai titik dalam ruang n-dimensi (di mana n adalah jumlah fitur) dengan nilai setiap fitur menjadi nilai koordinat tertentu. Kemudian dilakukan klasifikasi dengan mencari *hyperplane* yang membedakan kelas. Prosedur klasifikasi:

a. Ambil *dataset* berlabel dengan objek citra yang diinginkan.

b. Partisi *dataset* menjadi data training dan data testing.

- c. Membuat model dengan melakukan proses pembelajaran dengan SVM menggunakan fitur yang diekstrak dari data training.
- d. Uji model menggunakan fitur yang diekstraksi dari data testing

3. Hasil dan Pembahasan

Langkah awal yaitu pengambilan *dataset* kemudian pelabelan terhadap *dataset* citra bunga dengan memisahkan ke dalam folder-folder kelas bunga. Dalam penelitian ini citra dibagi menjadi dua kelompok, kelompok pertama yaitu citra bunga *Aquilegia*, *Iris*, *Viola*, dan *Rhododendron* total 114 citra. Seperti terlihat pada Gambar 1, kelompok bunga ini adalah kelas-kelas yang berisi citra dengan warna dan bentuk yang relatif berbeda dalam kelas yang sama. Kelompok kedua yaitu citra bunga *Calendula*, *Goldquelle*, *Shastadaisy*, dan *Veronica Chamaedrys* total 125 citra. Seperti terlihat pada Gambar 2, kelompok bunga ini adalah kelas-kelas yang berisi citra dengan warna dan bentuk yang relatif sama pada tiap kelasnya.

Untuk percobaan pertama yaitu citra dari kedua kelompok di ubah terlebih dahulu ke dalam dimensi satu (*vector*) kemudian disimpan dalam *DataFrame*, yaitu struktur data berlabel dua dimensi yang berisi *vector* dan label tiap kelas. Sebelum klasifikasi di bagi menjadi data training 80% dan data testing 20%. Selanjutnya adalah klasifikasi dengan metode SVM, Untuk kelompok bunga pertama yaitu *Aquilegia*, *Iris*, *Rhododendron*, dan *Viola* menghasilkan akurasi sebesar 100% pada Gambar 4, dan untuk kelompok bunga kedua yaitu *Calendula*, *Goldquelle*, *Shastadaisy*, dan *Veronica Chamaedrys* menghasilkan akurasi sebesar 52.2% pada Gambar 5.

	precision	recall	f1-score	support
calendula	1.00	1.00	1.00	7
goldquelle	1.00	1.00	1.00	7
shastadaisy	1.00	1.00	1.00	6
veronicachamaedrys	1.00	1.00	1.00	3
accuracy			1.00	23
macro avg	1.00	1.00	1.00	23
weighted avg	1.00	1.00	1.00	23

Sumber: Hasil Penelitian (2022)

Gambar 4. Hasil Klasifikasi dengan SVM untuk Bunga Kelompok Pertama

	precision	recall	f1-score	support
calendula	0.27	0.75	0.40	4
goldquelle	0.67	0.40	0.50	10
shastadaisy	1.00	0.43	0.60	7
veronicachamaedrys	0.67	1.00	0.80	2
accuracy			0.52	23
macro avg	0.65	0.64	0.57	23
weighted avg	0.70	0.52	0.54	23

Sumber: Hasil Penelitian (2022)

Gambar 5. Hasil Klasifikasi dengan SVM untuk Bunga Kelompok Kedua

Pada percobaan kedua, untuk penerapan ekstraksi fitur *HOG*, citra diubah ukurannya menjadi 64x128 (mengikuti arahan penemu fitur *HOG*) kemudian dilakukan langkah-langkah menghitung Gradien, membagi gambar menjadi sel, normalisasi Blok, membuat Vektor Fitur *HOG*. Hasil penerapan ekstraksi fitur *HOG* di simpan dalam *DataFrame*. Sebelum klasifikasi data di bagi menjadi data training 80% dan data testing 20%. Selanjutnya adalah klasifikasi dengan metode SVM. Langkah terakhir adalah evaluasi, untuk kelompok bunga pertama yaitu *Aquilegia*, *Iris*, *Viola*, dan *Rhododendron* menghasilkan akurasi sebesar 48% pada Gambar 6 dan untuk kelompok bunga kedua yaitu *Calendula*, *Golquelle*, *Shastadaisy*, dan *Veronica Chamaedrys* menghasilkan akurasi sebesar 64% pada Gambar 7.

	precision	recall	f1-score	support
aquilegia	0.17	0.20	0.18	5
iris	0.62	0.89	0.73	9
rhododendron	1.00	0.40	0.57	5
viola	0.25	0.17	0.20	6
accuracy			0.48	25
macro avg	0.51	0.41	0.42	25
weighted avg	0.51	0.48	0.46	25

Sumber: Hasil Penelitian (2022)

Gambar 6. Hasil Klasifikasi dengan SVM dan *HOG* untuk Bunga Kelompok Pertama

	precision	recall	f1-score	support
aquilegia	0.56	0.83	0.67	6
iris	0.78	0.88	0.82	8
rhododendron	0.33	0.20	0.25	5
viola	0.75	0.50	0.60	6
accuracy			0.64	25
macro avg	0.60	0.60	0.59	25
weighted avg	0.63	0.64	0.62	25

Sumber: Hasil Penelitian (2022)

Gambar 7. Hasil Klasifikasi dengan SVM dan *HOG* untuk Bunga Kelompok Kedua

Proses klasifikasi pada penelitian terdiri dari langkah-langkah melakukan, preprosesing, pembagian data training dan data testing, ekstraksi fitur *HOG*, klasifikasi dengan SVM, dan evaluasi mendapatkan hasil berupa akurasi model menggunakan pemrograman *Python*.

4. Kesimpulan

Klasifikasi dengan SVM terhadap kelompok citra dengan data citra yang memiliki warna dan bentuk yang relative sama dalam penelitian ini menghasilkan akurasi yang sempurna yaitu 100% meskipun tidak menggunakan Teknik ekstraksi fitur apapun. Ketika diterapkan ekstraksi fitur *HOG*, akurasi berkurang menjadi 52.2%. Kondisi kebalikan terjadi pada citra kelompok kedua yaitu kelas bunga dengan data citra yang relative berbeda baik dari segi warna maupun bentuknya, klasifikasi tanpa menggunakan teknik ekstraksi fitur apapun hasilnya adalah 48%

namun ketika dilakukan percobaan dengan implementasi fitur *HOG* kemudian klasifikasi dengan metode *SVM*, akurasi naik menjadi 64%. Dalam penelitian ini dapat disimpulkan bahwa *dataset* citra dengan warna dan bentuk yang relative seragam tidak perlu penerapan ekstraksi fitur apapun dapat menghasilkan akurasi yang tinggi. Untuk klasifikasi dengan melakukan ekstraksi fitur dalam penelitian ini memberikan hasil yang berbeda untuk dua kelompok bunga. Dengan demikian perlu penelitian lebih lanjut karena penerapan teknik ekstraksi fitur lain bisa jadi berbeda hasilnya. Jumlah kelas dan jumlah *dataset* juga bisa jadi mempengaruhi hasil mengingat dalam penelitian ini data yang digunakan untuk tiap kelas tidaklah banyak.

Daftar Pustaka

- Bao, T. Q., Kiet, N. T. T., Dinh, T. Q., & Hiep, H. X. (2020). Plant Species Identification from Leaf Patterns Using Histogram of Oriented Gradients Feature Space and Convolution Neural Networks. *Journal of Information and Telecommunication*, 4(2), 140–150. <https://doi.org/10.1080/24751839.2019.1666625>
- Cao, X., Wu, C., Yan, P., & Li, X. (2011). Linear SVM Classification using Boosting HOG Features for Vehicle Detection in Low-Altitude Airborne Videos. *2011 18th IEEE International Conference on Image Processing*, 2421–2424.
- Cheon, M., Lee, W., Yoon, C., & Park, M. (2012). Vision-Based Vehicle Detection System with Consideration of the Detecting Location. *IEEE Transactions on Intelligent Transportation Systems*, 13(3), 1243–1252. <https://doi.org/10.1109/TITS.2012.2188630>
- Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 886–893 vol. 1. <https://doi.org/10.1109/CVPR.2005.177>
- Gandhani, S., & Singhal, N. (2015). Content Based Image Retrieval: Survey and Comparison of CBIR System based on Combined Features. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 8(11), 417–422. <https://doi.org/10.14257/ijcip.2015.8.11.37>
- Guo, J., Cheng, J., Pang, J., & Guo, Y. (2013). Real-Time Hand Detection Based on Multi-Stage HOG-SVM Classifier. *2013 IEEE International Conference on Image Processing*, 4108–4111. <https://doi.org/10.1109/ICIP.2013.6738846>
- Li, Z., Niu, B., Peng, F., Li, G., Yang, Z., & Wu, J. (2018). Classification of Peanut Images Based on Multi-features and SVM. *IFAC-PapersOnLine*, 51(17), 726–731. <https://doi.org/https://doi.org/10.1016/j.ifacol.2018.08.110>
- Liu, W., Zhang, Y., Fan, H., Zou, Y., & Qin, Y. (2020). Detection of Late Blight in Potato Leaves Based on Multi-Feature and SVM Classifier. *Journal of Physics: Conference Series*, 1518(1). <https://doi.org/10.1088/1742-6596/1518/1/012045>
- Nilsback, M.-E., & Zisserman, A. (2008). Automated Flower Classification Over a Large Number of Classes. *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, 722–729. <https://doi.org/10.1109/ICVGIP.2008.47>

- Nilsback, M. E., & Zisserman, A. (2006). A Visual Vocabulary for Flower Classification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2, 1447–1454. <https://doi.org/10.1109/CVPR.2006.42>
- Sharma, P., Aggarwal, A., Gupta, A., & Garg, A. (2019). Leaf Identification Using HOG, KNN, and Neural Networks. In S. Bhattacharyya, A. E. Hassanien, D. Gupta, A. Khanna, & I. Pan (Eds.), *International Conference on Innovative Computing and Communications* (pp. 83–91). Springer Singapore.
- Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/https://doi.org/10.1016/j.ipm.2009.03.002>
- Wäldchen, J., Rzanny, M., Seeland, M., & Mäder, P. (2018). Automated Plant Species Identification—Trends and Future Directions. *PLoS Computational Biology*, 14(4), 1–19. <https://doi.org/10.1371/journal.pcbi.1005993>