

Identifikasi Ras Manusia Melalui Citra Wajah Menggunakan Metode Histogram of Oriented Gradient dan Klasifikasi Linear Discriminant Analysis

I Wayan Hanson Narandika ¹⁾, Bambang Hidayat ²⁾, Johan Arif ³⁾

^{1),2)}Fakultas Teknik Elektro, Universitas Telkom
Jl. Terusan Buah Batu No. 1 Bandung

³⁾Fakultas Ilmu dan Teknologi Kebumihan, Institut Teknologi Bandung
Jl. Ganesha No. 10 Bandung

Email : hansonnarandika@student.telkomuniversity.ac.id

Abstrak. Ras merupakan suatu sistem klasifikasi yang digunakan untuk mengelompokkan manusia melalui ciri fenotipe, asal usul geografis, dan ciri jasmani. Ras merupakan salah satu informasi demografis yang dimiliki suatu individu. Informasi demografis dapat berupa ras, usia dan jenis kelamin. Identifikasi karakteristik individu berdasarkan citra wajah sudah banyak diteliti sebelumnya. Namun penelitian terkait identifikasi informasi demografis seperti ras masih belum banyak dilakukan. Berdasarkan permasalahan ini penulis mengembangkan algoritma yang mampu mengidentifikasi ras berdasarkan ciri dari citra wajah. Ciri citra wajah tersebut diekstrak menggunakan metode Histogram of Oriented Gradient (HOG) dan diklasifikasikan menggunakan metode Linear Discriminant Analysis (LDA). Pada sistem yang diteliti ras dikelompokkan menjadi 3 kelas, yaitu Ras Mongoloid, Negroid, dan Kaukasoid. Hasil akhir dari penelitian ini adalah sistem yang mampu melakukan identifikasi dan klasifikasi ras dari suatu individu. Sistem memiliki akurasi 82 % dengan menggunakan 90 sampel citra latih dan 90 citra uji.

Kata kunci: ras, histogram of oriented gradient, linear discriminant analysis.

1. Pendahuluan

Ras merupakan suatu sistem klasifikasi yang digunakan untuk mengelompokkan manusia melalui ciri fenotipe, asal usul geografis, dan ciri jasmani [1]. Pada umumnya ras manusia dibagi menjadi 3, yaitu Mongoloid, Negroid, dan Kaukasoid. Masing-masing ras memiliki ciri-ciri fisik yang membedakan antar satu sama lain. Ciri-ciri tersebut dapat dilihat dari warna rambut, warna kulit, dan bentuk mata karena masing-masing ras memiliki ciri khas tersendiri. Ras merupakan salah satu informasi demografis yang dimiliki suatu individu. Informasi demografis dapat berupa ras, usia dan jenis kelamin. Identifikasi karakteristik individu berdasarkan citra wajah sudah banyak diteliti pada bidang *face recognition* [2,3]. Namun penelitian terkait identifikasi informasi demografis seperti ras masih belum banyak dilakukan [2,4]. Berdasarkan permasalahan ini penulis mengembangkan algoritma yang mampu mengidentifikasi ras berdasarkan ciri dari citra wajah menggunakan *image processing*. Dataset citra wajah yang akan digunakan pada penelitian ini berasal dari *The Chicago Face Database* [5]. Sebanyak 180 citra wajah yang terdiri dari Ras Mongoloid, Negroid, dan Kaukasoid digunakan sebagai citra latih dan citra uji pada penelitian ini.

Masing-masing citra wajah ras memiliki ciri yang berbeda. Ciri tersebut akan diekstrak menggunakan metode *Histogram of Oriented Gradient* (HOG). Ekstraksi ciri dilakukan untuk mendapatkan ciri karakteristik dari citra wajah masing-masing ras. Penulis memilih metode ini karena metode ini memiliki akurasi yang tinggi dan cocok digunakan untuk mendeteksi pola [6]. *Histogram of Oriented Gradient* (HOG) adalah sebuah metode yang digunakan dalam *image processing* sebagai deteksi objek. Teknik ini menghitung nilai *gradient* dalam daerah tertentu pada suatu citra. Tiap citra memiliki karakteristik yang ditunjukkan oleh distribusi *gradient*. Karakteristik ini diperoleh dengan membagi citra ke dalam daerah kecil yang disebut *cell*. Pada tiap *cell* disusun sebuah *histogram* dari sebuah *gradient* [7]. *Histogram of Oriented Gradient* (HOG) memiliki beberapa parameter, yaitu *cell size*, *block size*, dan *bin numbers*.

Setelah dilakukan ekstraksi ciri maka akan didapatkan *database* latih berupa ciri karakteristik dari masing-masing ras. Untuk dapat mengidentifikasi ras dari suatu individu maka diperlukan metode klasifikasi yang mampu mengklasifikasikan ras berdasarkan ciri yang sudah didapatkan. Penulis

menggunakan metode *Linear Discriminant Analysis* (LDA) sebagai metode klasifikasi karena metode ini merupakan metode yang memiliki akurasi yang tinggi dan metode klasifikasi ini sering digunakan untuk identifikasi citra wajah [8]. Pada sistem yang diteliti ras dikelompokkan menjadi 3 kelas, yaitu Mongoloid, Negroid, dan Kaukasoid.

Linear Discriminant Analysis merupakan salah satu metode klasifikasi yang lebih dikenal dengan sebutan *Fisher's Linear Discriminant*. Metode klasifikasi ini merupakan perpaduan antara perhitungan operasi matematika dan statistika [9]. Untuk membuat *classifiers* pada LDA, dilakukan perhitungan rata-rata sampel dari masing-masing kelas. Langkah selanjutnya adalah menghitung sampel kovarian dengan cara mengurangi rata-rata sampel dari setiap kelas pengamatan dan mengambil nilai matriks kovarian pengamatan sebagai hasilnya [10]. Berikut merupakan skema pembuatan *classifiers* pada LDA. Untuk mengestimasi rata-rata kelas data yang belum diberi pembobotan maka digunakan rumus sebagai berikut.

$$\hat{\mu}_k = \frac{\sum_{n=1}^N M_{nk} x_n}{\sum_{n=1}^N M_{nk}} \dots\dots\dots (1)$$

Dimana M merupakan matriks *N-by-K*. M_{nk} akan bernilai 1 jika data pengamatan n berasal dari kelas k . M_{nk} akan bernilai 0 jika sebaliknya. Selanjutnya akan dihitung data yang sudah memiliki bobot positif w_n . Dapat dihitung menggunakan persamaan berikut.

$$\hat{\mu}_k = \frac{\sum_{n=1}^N M_{nk} w_n x_n}{\sum_{n=1}^N M_{nk} w_n} \dots\dots\dots (2)$$

Dengan demikian, estimasi dari matriks kovarian untuk data yang belum diberi pembobotan dapat dirumuskan sebagai berikut.

$$\hat{\Sigma} = \frac{\sum_{n=1}^N \sum_{k=1}^K M_{nk} (x_n - \hat{\mu}_k)(x_n - \hat{\mu}_k)^T}{N - K} \dots\dots\dots (3)$$

Setelah pembuatan *classifiers*, maka langkah selanjutnya adalah melakukan prediksi kelas dari data uji. Berikut merupakan persamaannya.

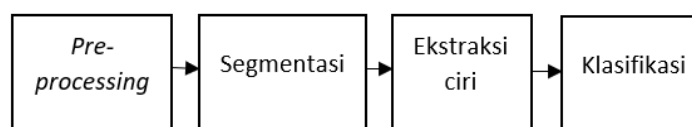
$$\hat{y} = \arg \min_{y=1 \dots\dots\dots k} \sum_{k=1}^K \hat{P}(k | x) C(y | k) \dots\dots\dots (4)$$

Dimana $\hat{y}$ merupakan prediksi kelas, K merupakan jumlah kelas, $\hat{P}(k | x)$ merupakan probabilitas kelas k untuk pengamatan data x dan $C(y | k)$ merupakan nilai klasifikasi pengamatan y terhadap k yang merupakan kelas sebenarnya.

Hasil yang diharapkan dari sistem yang diteliti adalah sistem mampu memiliki akurasi yang tinggi dan dapat dimanfaatkan pada bidang penegakan hukum, forensik, *social media*, dan *human-computer interaction*.

2. Pembahasan

Pada penelitian ini, gambaran umum sistem yang diteliti dan diimplementasikan adalah sebagai berikut.

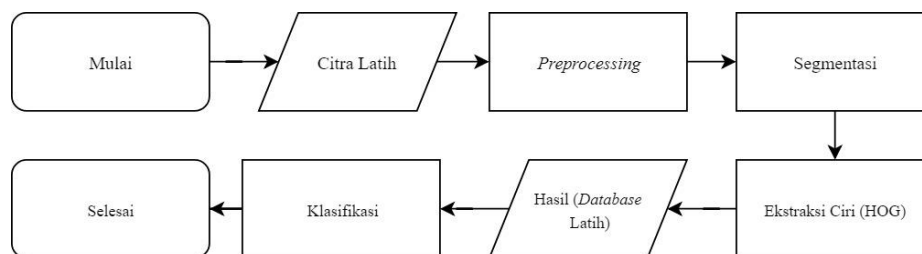


Gambar 1. Gambaran umum sistem

Berdasarkan gambar diatas, sistem yang diteliti pada penelitian ini terdiri dari empat tahap. Tahap pertama, yaitu tahap *preprocessing* dimana pada tahap ini dilakukan untuk mempermudah sistem dalam mengenali citra masukan. Tahap kedua, yaitu segmentasi citra yang bertujuan untuk mengoptimalkan dan mempermudah untuk menganalisis citra. Selanjutnya adalah tahap ekstraksi ciri untuk mendapatkan ciri citra wajah menggunakan metode *Histogram of Oriented Gradient* (HOG). Dan tahap terakhir adalah tahap klasifikasi menggunakan metode *Linear Discriminant Analysis* (LDA).

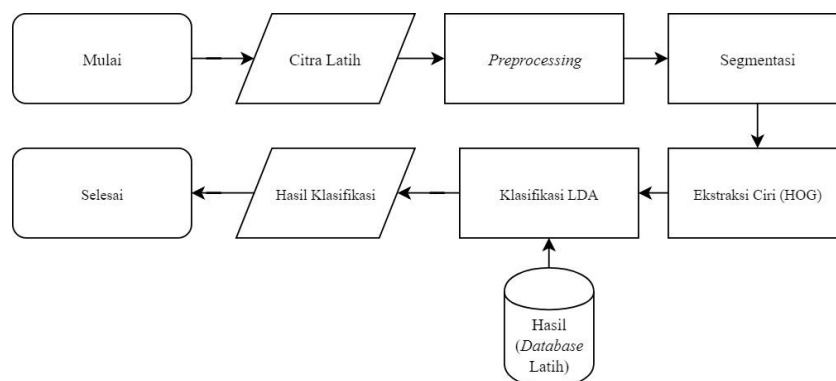
2.1. Tahap Pelatihan dan Tahap Pengujian

Tahap pelatihan dan pengujian dilakukan menggunakan metode *Histogram of Oriented Gradient* (HOG) sebagai proses ekstraksi ciri dan metode *Linear Discriminant Analysis* (LDA) sebagai proses klasifikasinya. Pada tahap pelatihan, proses awal yang dilakukan adalah *preprocessing* dan segmentasi data citra latih masukan. Selanjutnya akan dilakukan ekstraksi ciri terhadap data latih yang sudah melalui proses *preprocessing* dan segmentasi. Proses ekstraksi ciri akan menghasilkan ciri dari data latih yang akan dijadikan database latih untuk klasifikasi pada tahap pengujian. Diagram alir tahap pelatihan dapat dilihat pada gambar 2.



Gambar 2. Diagram alir tahap pelatihan

Tahap selanjutnya adalah tahap pengujian yang dilakukan untuk menguji data citra untuk diklasifikasikan oleh sistem. Tahapan proses pengujian digambarkan pada gambar 3.



Gambar 3. Diagram alir tahap pengujian

2.2. Analisis dan Pengujian Sistem

Pengujian sistem dilakukan dengan cara merubah parameter dari ekstraksi ciri *Histogram of Oriented Gradient* (HOG) untuk mendapatkan performansi yang terbaik. Pada bagian ini akan dilihat pengaruh dari perubahan parameter terhadap akurasi dan waktu komputasi yang diperoleh.

2.2.1 Pengujian Pengaruh Parameter Cell Size

Pengujian ini bertujuan untuk mengetahui pengaruh parameter *cell size* pada metode ekstraksi ciri HOG terhadap akurasi dan waktu komputasi sistem. Hasil perbandingan dapat dilihat pada tabel dibawah.

Tabel 1. Hasil Pengujian Pengaruh Parameter *Cell Size*

<i>Cell Size</i>	Akurasi	Waktu Komputasi (s)
2x2	65,7%	33,94
4x4	72,2%	38,57
8x8	82,2%	39,74
16x16	75,5%	29,73
32x32	76,6%	29,34

Berdasarkan pengujian parameter *cell size* akurasi tertinggi didapatkan ketika *cell size* 8x8, yaitu 82,2% dengan waktu komputasi 39,74 detik.

2.2.2 Pengujian Pengaruh Block Size

Pengujian ini bertujuan untuk mengetahui pengaruh parameter *block size* pada metode ekstraksi ciri HOG terhadap akurasi dan waktu komputasi sistem. Berdasarkan pengujian yang telah dilakukan sebelumnya maka parameter *cell size* yang digunakan pada pengujian tahap ini adalah 8x8 karena memiliki akurasi tertinggi. Hasil perbandingan dapat dilihat pada tabel dibawah.

Tabel 2. Hasil Pengujian Pengaruh Parameter *Block Size*

<i>Block Size</i>	Akurasi	Waktu Komputasi (s)
2x2	76,6%	30,26
4x4	74,4%	34,15
8x8	82,2%	37,42

Berdasarkan pengujian parameter *block size* akurasi tertinggi didapatkan ketika *block size* 8x8. Akurasi yang didapatkan adalah 82,2 % dan waktu komputasi sebesar 37,42 detik. Semakin besar ukuran *block size* maka waktu komputasi yang diperlukan semakin lama.

2.2.3 Pengujian Pengaruh Bin Numbers

Pengujian ini bertujuan untuk mengetahui pengaruh parameter *bin numbers* yang ada pada HOG terhadap akurasi dan waktu komputasi sistem. Berdasarkan dua tahap pengujian yang sudah dilakukan maka *cell size* yang digunakan adalah 8x8 dan *block size* yang digunakan adalah 8x8 karena memiliki akurasi tertinggi. Hasil perbandingan pengaruh parameter *bin numbers* dapat dilihat pada tabel berikut.

Tabel 3. Hasil Pengujian Pengaruh Parameter *Block Size*

<i>Bin Numbers</i>	Akurasi	Waktu Komputasi (s)
9	82,2%	33,46
12	80%	37,22
15	80%	43,61

Berdasarkan pengujian parameter *bin numbers* didapatkan akurasi tertinggi ketika bin 9 dengan akurasi 82,2%. Bin 12 dan 15 memiliki akurasi yang sama, yaitu 80%. Dapat dilihat pada hasil pengujian semakin banyak jumlah bin waktu komputasi yang diperlukan akan semakin besar.

2.2.3 Pengujian Ciri Statistik

Pada bagian ini dilakukan pengujian dengan menggunakan 6 ciri statistik orde pertama. Ciri statistik yang digunakan antara lain *mean*, *variance*, *standard deviation*, *skewness*, *kurtosis*, dan *entropy*. Berikut merupakan tabel hasil pengujian 1 ciri statistik dan kombinasi dari 2 hingga 6 ciri statistik.

Tabel 4. Hasil Pengujian Ciri Statistik

Ciri Statistik	Akurasi	Waktu Komputasi (s)
<i>Mean</i>	80%	0,56
<i>Variance</i>	81,1%	0,14
<i>Standard Deviation</i>	81,1%	0,15
<i>Skewness</i>	48,8%	0,09
<i>Kurtosis</i>	35,5%	0,08
<i>Entropy</i>	75,5%	0,13
<i>Mean dan variance</i>	83,3%	0,09
<i>Mean, variance, standard deviation</i>	78,8%	0,08
<i>Mean, variance, standard deviation, dan skewness</i>	81,1%	0,11
<i>Mean, variance, standard deviation, skewness, dan kurtosis</i>	81,1%	0,09
<i>Mean, variance, standard deviation, skewness, kurtosis, dan entropy</i>	82,2%	0,9

Berdasarkan pengujian ciri yang sudah dilakukan dapat disimpulkan bahwa kombinasi antara *mean* dan *variance* memiliki akurasi tertinggi yaitu 83,3%.

3. Kesimpulan

Berdasarkan hasil pengujian pada sistem identifikasi dan klasifikasi ras melalui citra wajah didapatkan kesimpulan sebagai berikut:

1. Sistem yang dihasilkan mampu mengidentifikasi ras menggunakan metode ekstraksi ciri *Histogram of Oriented Gradient* (HOG) dan klasifikasi *Linear Discriminant Analysis* (LDA) dengan jumlah data uji sebesar 90 citra.
2. Berdasarkan percobaan yang sudah dilakukan terdapat beberapa parameter ekstraksi ciri metode *Histogram of Oriented Gradient* (HOG) yang mempengaruhi performansi dan akurasi sistem. Parameter tersebut antara lain *cell size*, *block size* dan *bin numbers*.
3. Parameter *Histogram of Oriented Gradient* (HOG) yang memiliki akurasi adalah kombinasi dari *cell size* 8x8, *block size* 8x8, dan *bin* 9 dengan akurasi 82,2% serta waktu komputasi 33,46 detik. Pada pengujian ciri statistik kombinasi yang memiliki akurasi tertinggi adalah *mean* dan *variance* dengan akurasi 83,3%.

Ucapan Terima Kasih

Ucapan terima kasih penulis ucapkan kepada Tuhan Yang Maha Esa karena berkat karunia-Nya penelitian ini dapat berjalan dengan baik. Penulis juga ucapkan terima kasih kepada semua pihak yang telah mendukung penelitian ini terutama tim riset dari Institut Teknologi Bandung dan Universitas Telkom.

Daftar Pustaka

- [1]. Kroeber and A. Louis, *Anthropology: race, language, culture, psychology, pre-history*. New York: Harcourt, 1948.
- [2]. H. Han, C. Otto, X. Liu, and A. K. Jain, "Demographic Estimation from Face Images: Human vs. Machine Performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1148-1161, Jun. 2015.
- [3]. S. Z. Li and A. K. Jain (eds.), *Handbook of Face Recognition*, 2nd ed. New York: Springer, 2011.
- [4]. N. Kumar, A. Berg, P. Belhumeur, and S. Nayar, "Describable Visual Attributes for Face Verification and Image Search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1962-1977, Oct. 2011.
- [5]. Ma, Correll, & Wittenbrink (2015). The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data. *Behavior Research Methods*, 47, 1122-1135.
- [6]. E. Abraham, S. Mishra, N. Tripathi, and G. Sukumaran, "HOG Descriptor based Registration (A New Image Registration Technique)," in *The 15th International Conference on Advanced Computing Technologies (ICACT)*, 2014.
- [7]. M. Ghorbani, M. M. Dehshibi, and A. T.Targhi, "HOG and LBP: Towards a Robust Face Recognition System," in *The Tenth International Conference on Digital Information Management*, 2015.
- [8]. N. A. A. Shashoa, N. A. Salem, I. N. Jleta, O. Abusaeeda, "Classification Depend on Linear Discriminant Analysis Using Desired Outputs," in *The 17th international conference on Sciences and Techniques of Automatic control & computer engineering*, 2016.
- [9]. A. Scholahuddin, R. E. Siregar, I. Supriana, and S. Hadi, "Penerapan Metode Linier Discriminant Analysis Pada Pengenalan Wajah Berbasis Kamera," in *Proc. Semnasunima 2010*, 30 Juni - 4 Juli 2010.
- [10]. MathWorks, "Discriminant Analysis Classification," *The MathWorks*, <https://www.mathworks.com/help/stats/discriminant-analysis.html>. [Diakses 11 Desember 2018, 10:12:44 WIB].