

Deteksi Suku Minangkabau dan Sunda Berdasarkan Citra Sidik Bibir dengan Metode *Image Registration Gray Level Co-Occurrence Matrix (GLCM)* Dan Klasifikasi *Decision Tree (DT)*

Lara Pratiwi ¹⁾, Bambang Hidayat ²⁾, Yurika Ambar Lita ³⁾

^{1),2)}Fakultas Teknik Elektro, Universitas Telkom, Bandung

³⁾Fakultas Kedokteran Gigi, Universitas Padjajaran, Bandung

Email : larapратиwi@student.telkomuniversity.ac.id

Abstrak. Banyak ilmu yang mempelajari prinsip dan cara klasifikasi makhluk hidup. Salah satunya ilmu forensik. Ilmu forensik adalah bidang ilmu pengetahuan yang digunakan untuk mengidentifikasi individu seperti jenis kelamin, suku, ras, dan usia, yang digunakan untuk membantu proses penegakkan keadilan. Setiap individu memiliki pola ciri yang unik, tidak ada seorangpun yang memiliki pola ciri yang persis sama dengan individu lain. Pola ciri inilah yang dapat digunakan untuk mengidentifikasi tiap individu. Salah satunya yaitu pola sidik bibir. Sama halnya dengan pola sidik jari yang bersifat stabil dari waktu ke waktu. Sidik bibir memiliki kelebihan yaitu polanya sangat sulit untuk hilang kecuali dalam kasus-kasus tertentu. Selain itu, sidik bibir dapat dengan mudah ditemukan pada benda-benda yang pernah digunakan, seperti peralatan makan yang digunakan sehari-hari. Gambaran sulci pada mukosa bibir atas dan bawah inilah yang digunakan dalam mengidentifikasi individu. Penelitian ini bertujuan untuk mengidentifikasi suku minangkabau dan suku sunda berdasarkan registrasi citra dengan metode *image registration gray level co-occurrence matrix (GLCM)* dan *decision tree (DT)* melalui pengolahan sampel citra bibir menggunakan aplikasi *Matrix Laboratory (MATLAB)* sehingga mendapatkan akurasi sebesar 85% menggunakan parameter-parameter *co-occurrence* bernilai $d=2$, *quantization* bernilai 8, dan dua buah statistik yaitu statistik *contrast* dan *energy*.

Kata Kunci: sidik, bibir, *GLCM*, *DT*, forensik, Suku, Minangkabau, Sunda

1. Pendahuluan

Belakangan ini banyak terjadi kasus kematian yang tidak wajar, salah satunya yang disebabkan oleh tindak kriminalitas. Pembunuhan, pencurian, dan tindakan asusila merupakan salah satu contoh tindak kriminalitas yang mengharuskan adanya tindakan lebih lanjut. Identifikasi manusia merupakan suatu proses yang sangat sulit untuk dilakukan. Suku, jenis kelamin, ras, dan umur adalah salah satu contoh faktor yang dapat digunakan untuk memperkecil ruang lingkup identifikasi makhluk hidup.

Penentuan identitas identitas baik korban maupun tersangka merupakan hal terpenting dalam penyelidikan. Selain itu, identifikasi individu dapat juga digunakan sebagai prasyarat dalam pembuatan surat kematian, alasan pribadi, sosial dan hukum[1]. Ilmu yang mempelajari identifikasi individu untuk kepentingan hukum dan keadilan disebut dengan ilmu forensik[2]. Salah satu cabang dari ilmu forensik adalah odontologi forensik atau ilmu kedokteran gigi forensik. Dalam ilmu kedokteran gigi forensik membahas mengenai penelitian bukti yang berhubungan dengan gigi dan mulut[3].

Salah satu caranya yaitu melalui identifikasi sidik bibir tau *cheiloscopy*, karena sidik bibir bersifat unik tiap individu dan sangat sulit untuk hilang kecuali dalam kasus-kasus tertentu. Bahkan beberapa peneliti pernah melaporkan bahwa pola sidik bibir manusia bersifat tetap seumur hidup semenjak bayi berusia 6 minggu pada masa kehamilan[1][2].

Penelitian ini bertujuan untuk mengidentifikasi suku minangkabau dan suku sunda berdasarkan registrasi citra dengan metode *image registration gray level co-occurrence matrix (GLCM)* dan *decision tree (DT)* melalui pengolahan sampel citra bibir menggunakan aplikasi *Matrix Laboratory (MATLAB)*.

2.1. Gray Level Co-occurrence Matrix (GLCM)

Metode GLCM merupakan suatu metode yang melakukan analisis terhadap suatu piksel pada citra dan mengetahui tingkat keabuan yang sering terjadi. Metode ini juga untuk tabulasi tentang frekuensi kombinasi nilai piksel yang muncul pada suatu citra[3]. Adanya matriks tersebut berdasarkan kondisi bahwa suatu matriks piksel akan mempunyai nilai perulangan sehingga terdapat pasangan aras kebuanya[4]. Proses-proses GLCM diantaranya:

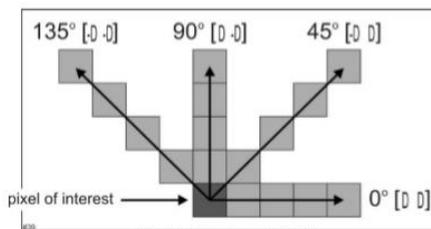
1) *Quantization*

Merupakan konversi dari nilai *grayscale* citra ke menjadi rentang level tertentu. *Quantization* bertujuan untuk mengurangi angka perhitungan dan merigankan proses komputasi. Sebagai contoh, ditetapkan delapan rentang dengan nilai kuantisasi 0-7 dimana seriap rentang mewakili 32 nilai keabuan.

Tabel 1. Quantization

No	Nilai Kuantisasi	Nilai Rentang
1.	0	0-31
2.	1	32-63
3.	2	64-95
4.	3	96-127
5.	4	128-159
6.	5	160-191
7.	6	192-223
8.	7	224-255

2) *Co-Occurrence*



Gambar 1. Arah Kookurasi GLCM[5]

Kookurensi dapat diartikan sebagai kejadian bersama, berarti banyaknya kejadian pada satu level piksel yang bertetangga dengan nilai piksel lainnya berdasarkan jarak (*d*) dan orientasi suatu sudut (Θ). Jarak dipresentasikan sebagai piksel sedangkan orientasi dipresentasikan sebagai derajat, yang terbentuk dari empat arah sudut dengan interval 45° seperti yang terlihat pada gambar 1 dan jarak antar piksel ditentukan sebesar 1 piksel [5].

3) *Symmetric*

Merupakan kemunculan posisi piksel yang sama dan didapat dari hasil penjumlahan matriks *co-occurrence* dengan matriks *transpose*-nya[6].

4) *Normalization*

Merupakan hasil pembagian setiap angka pada matriks *symmetric* dengan jumlah seluruh angka pada matriks tersebut[6].

5) Ciri Statistik

Variasi ciri statistik yang digunakan diantaranya *contrast*, *correlation*, *energy*, dan *homogeneity*[7].

a) Statistik *contrast*

Menunjukkan ukuran penyebaran elemen-elemen matriks citra. Jika terletak jauh dari diagonal uama, maka akan bernilai besar. Berikut persamaan untuk menghitung nilai *contrast*:

$$Contrast = \sum_i \sum_j (i - j)^2 C(i, j) \tag{1}$$

Dimana, *i* = baris, *j* = kolom

b) Statistik *correlation*

Menyatakan ukuran ketergantungan linear derajat kerajat keabuan citra sehingga dapat menunjukkan adanya struktur linear dalam citra. Perhitungannya dapat menggunakan persamaan:

$$Correlation = \sum_i \sum_j \frac{(i \ x \ j) \times C(i,j) - (\mu_x \times \mu_y)}{\sigma_x \times \sigma_y} \quad (2)$$

Dimana, i = baris, j = kolom, μ = mean, σ = variasi

c) Statistik *energy*

Menunjukkan pengukuran keseragaman atau *angular second moment*. *Energy* akan bernilai tinggi ketika nilai piksel mirip dengan piksel lain, sebaliknya akan bernilai kecil ketika nilainya heterogen. Berikut dijelaskan persamaan dalam menghitung *energy*:

$$Energy = \sum_i \sum_j C^2(i,j) \quad (3)$$

Dimana, i = baris, j = kolom

d) Statistik *homogeneity*

Bernilai tinggi ketika semua piksel mempunyai nilai yang sama yang sangat sensitif terhadap nilai diagonal. Dalam perhitungannya dapat menggunakan persamaan sebagai berikut:

$$Homogeneity = \sum_i \sum_j \frac{C(i,j)}{1 + |i+j|} \quad (4)$$

Dimana, i = baris, j = kolom

2.2. Decision Tree (DT)

Decision tree adalah sebuah *flowchart* seperti pohon dimana setiap titik menunjukkan suatu kasus pada suatu atribut, setiap cabang menunjukkan hasil dari kasus tersebut, dan daun menunjukkan kelas-kelas atau distribusi kelas. *Decision tree* berguna untuk mengeksplorasi data yang sudah melewati tahanan *preprocessing* dan menemukan model yang tersembunyi dari data dengan sebuah target variabel sehingga dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan *record* yang lebih kecil [8]. Secara umum, untuk membangun algoritma C4.5 yaitu [9]:

- Pilih atribut sebagai akar.
- Buat cabang untuk masing-masing nilai.
- Bagi kasus dalam cabang.
- Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, dibutuhkan nilai *gain* tertinggi dari atribut-atribut yang ada [9]. Untuk menghitung *gain* dapat menggunakan persamaan [9]:

$$Gain(S,A) = Entropy(S) - \sum_i^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (5)$$

Untuk menghitung nilai *entropy* digunakan rumus:

$$Entropy(S) = \sum_{i=1}^n p_i * \log_2 p_i \quad (6)$$

Keterangan :

S : Himpunan kasus

A : Atribut

n : Jumlah partisi S

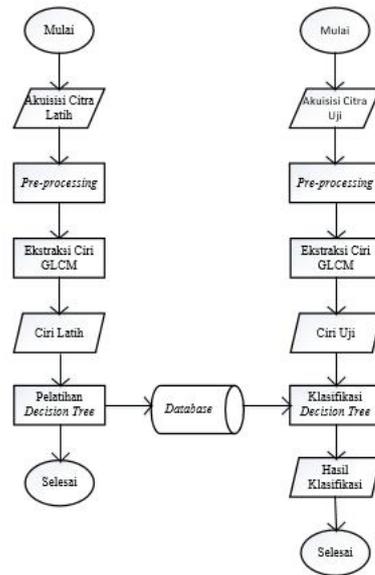
p_i : Proporsi dari S_i terhadap S

$|S_i|$: Jumlah kasus pada partisi ke-i

$|S|$: Jumlah kasus dalam S

2.3. Perancangan Sistem

Dalam model sistem ini dijelaskan tentang alur dalam program aplikasi identifikasi suku minangkabau dan suku sunda berbasis pengolahan citra digital. Adapun tahapan dalam proses identifikasi yaitu:



Gambar 2. Diagram Alir Desain Sistem

3. Data dan Analisis

Setelah dilakukan beberapa skenario dalam pengujian sistem, didapatkan beragam hasil untuk setiap pengujian. Beberapa hasil pengujian tersebut, dapat digambarkan dalam sebuah grafik untuk mendapatkan gambaran dari skenario yang paling baik dilakukan jika menggunakan metode GLCM dan klasifikasi *Decision Tree*. Berikut skenario-skenario yang dilakukan:

3.1. Pengaruh parameter *quantization*

Setelah dilakukan pengujian sistem yang membandingkan nilai *quantization* pada citra yang menggunakan *image registration* dengan citra yang tidak menggunakan *image registration* terlebih dahulu didapat hasil sebagai berikut:

Tabel 2. perbandingan nilai *quantization* pada image registration dan non image registration

<i>Quantization</i>	<i>Image Registration</i>		<i>Non Image Registration</i>	
	Akurasi (%)	Waktu (s)	Akurasi (%)	Waktu (s)
8	85.3659	174.11	81.6667	479.4007
16	69.1057	183.8579	66.6667	572.6664
32	76.4228	179.434	69.1667	591.2987

Berdasarkan tabel 2, pada saat menggunakan *image registration*, *quantization* yang paling baik digunakan ketika bernilai 8 dibandingkan dengan yang bernilai 16 ataupun 32. Sedangkan untuk waktu komputasinya, tidak menunjukkan perbedaan yang signifikan. Sedangkan ketika tidak menggunakan *image registration*, nilai akurasi yang dihasilkan setelah pengujian sistem menunjukkan angka yang lebih kecil, akan tetapi waktu komputasi yang dibutuhkan jauh lebih besar. Waktu komputasi yang lebih besar disebabkan karena penggunaan *image registration* akan mempercepat proses dari *decision tree*.

3.2. Pengaruh parameter *co-occurrence*

Tabel 3. Perbandingan nilai Co-occurrence pada image registration dan non image registration

<i>Co-occurrence</i>	<i>Image Registration</i>		<i>Non Image Registration</i>	
	Akurasi(%)	Waktu (s)	Akurasi (%)	Waktu (s)
[-1 -1; -1 0; -1 1; 0 1]	69.1667	196.1905	69.1667	431.5304
[-2 -2; -2 0; -2 2; 0 2]	71.5447	185.397	73.3333	975.3351

Berdasarkan table 3. yang berdasarkan parameter kuantisasi bernilai 8 dan statistik kontras, disaat kita menggunakan *image registration*, dapat kita lihat bahwa jarak mempengaruhi nilai akurasi

dan waktu komputasi. Nilai akurasi yang dihasilkan jika menggunakan jarak $d=2$ lebih baik daripada nilai akurasi yang dihasilkan oleh jarak $d=1$. Begitupun dengan nilai waktu komputasinya. Waktu oleh perangkat pada jarak $d=2$ lebih cepat dibandingkan dengan jarak $d=1$. Jadi kesimpulannya, jarak yang digunakan pada *co-occurrence* lebih memiliki kualitas yang baik jika menggunakan jarak $d=2$ dibandingkan jarak $d=1$.

Sedangkan disaat kita tidak menggunakan *image registration*, waktu komputasinya menjadi jauh lebih besar sehingga tidak efektif untuk digunakan, sedangkan nilai dari akurasinya tidak menunjukkan pola tersendiri. Ada yang nilai akurasinya mengalami kenaikan dan ada juga yang mengalami penurunan seperti disaat menggunakan parameter statistik *correlation*.

3.3. Pengaruh parameter statistik

3.3.1. Satu statistik

Pada penelitian ini, penulis membandingkan hasil yang didapat ketika penggunaan satu statistik yang berbeda dengan perlakuan yang sama ketika *quantization* bernilai 8, *co-occurrence* bernilai $d=2$, dan menggunakan *image registration*.

Tabel 4. Perbandingan nilai menggunakan satu statistik

Statistik	Akurasi (%)	Waktu Komputasi (s)
<i>Contrast</i>	71.5447	185.4
<i>Correlation</i>	64.2276	182.84
<i>Homogeneity</i>	67.4797	183.42
<i>Energy</i>	58.5366	181.37

3.3.2. Dua statistik

Pada penelitian ini, penulis membandingkan hasil yang didapat ketika menggunakan dua statistik yang berbeda dengan perlakuan yang sama ketika *quantization* bernilai 8, *co-occurrence* bernilai $d=2$, dan menggunakan *image registration*. Untuk melihat hasil pengujian yang didapat, dapat dilihat pada tabel 5 yaitu:

Tabel 5. Perbandingan nilai menggunakan dua statistik

Statistik	Akurasi (%)	Waktu (s)
Contrast dan energy	80.4878	173.0878
Correlation dan energy	75	176.3977
Homogeneity dan correlation	78.8618	173.7618
Homogeneity dan energy	82.5	174.5982
Contrast dan energy	85	174.1061

3.3.3. Tiga statistik

Pada penelitian ini, penulis membandingkan hasil yang didapat ketika menggunakan tiga statistik yang berbeda dengan perlakuan yang sama ketika *quantization* bernilai 8, *co-occurrence* bernilai $d=2$, dan menggunakan *image registration*. Berikut hasil yang didapat ketika dilakukan pengujian sistem.

Tabel 6. Perbandingan nilai menggunakan tiga statistik

Statistik	Akurasi (%)	Waktu (s)
Contrast dan correlation dan homogeneity	80	179.6774
Correlation dan energy dan homogeneity	79.1667	183.3673
Energy dan correlation dan contrast	81.667	177.4781
Contrast dan energy dan homogeneity	85	221.1654

Berdasarkan tabel 5 dan tabel 6, dapat ditarik kesimpulan bahwa statistik *contrast*, *energy*, dan *homogeneity* paling baik untuk digunakan bersama. Jika dibandingkan dengan penggunaan dua parameter statistik, penggunaan tiga parameter statistik membutuhkan waktu komputasi yang lebih besar, sehingga kurang efisien untuk digunakan.

3.3.4. Empat statistik

Pada penelitian ini, penulis membandingkan hasil yang didapat ketika menggunakan empat statistik yang berbeda dengan perlakuan yang sama ketika *quantization* bernilai 8. Pada tabel 4.7 terlihat nilai yang didapatkan ketika dilakukan pengujian sistem.

Tabel 7. Hasil pengujian sistem menggunakan empat statistik

<i>Co-occurrence</i>	<i>Image Registration</i>		<i>Non Image Registration</i>	
	Akurasi (%)	Waktu Komputasi (S)	Akurasi (%)	Waktu Komputasi (s)
[-1 -1; -1 0; -1 1; 0 1]	81.3008	172.9523	76.1667	759.4569
[-2 -2; -2 0; -2 2; 0 2]	82.1138	177.9884	80	633.2323

Berdasarkan tabel 7., dapat kita tarik kesimpulan bahwa pada saat penggunaan *image registration* maupun *non image registration*, nilai akurasi yang dihasilkan ketika menggunakan *co-occurrence* dengan $d=2$ akan memberikan hasil yang lebih baik dibandingkan dengan $d=1$. Kebalikan dengan waktu koputasi yang digunakan, $d=1$ membutuhkan waktu yang sedikit lebih cepat dibandingkan dengan $d=2$.

3. Kesimpulan

Setelah dilakukan pengujian sistem, didapat kesimpulan bahwa:

- Sistem yang dihasilkan mampu mengidentifikasi suku minangkabau dan suku sunda menggunakan metode GLCM dan *Decision Tree*.
- Hasil terbaik pengujian sistem yaitu menggunakan *image registration*, *quantization* bernilai 8, dan menggunakan dua buah statistik yaitu *contrast* dan *energy*.

Ucapan Terima Kasih

Alhamdulillahirrabbi lalamin puji syukur kepada Allah SWT, karena berkat rahmat, hidayah, dan ridha-Nya penulis dapat menyelesaikan makalah ini. Penulis sangat menyadari dalam penulisan ini tidak akan pernah selesai tanpa doa, dukungan, dorongan, dan bantuan dari berbagai pihak baik bersifat moral maupun material.

Daftar Pustaka

- [1] S. N. Qomariah, M. Novita, and E. Wulandari, "Hubungan antara Pola Sidik Bibir dengan Jenis Kelamin pada Mahasiswa Fakultas Kedokteran Gigi Universitas Jember," *eJurnal Pustaka Kesehatan*, vol. 2, no. 2, pp. 231–6, 2015.
- [2] J. M. Malahika, H. Karamoy, and R. J. Pusung, "Identifikasi Pola Sidik Bibir pada Pria dan Wanita Menggunakan Metode Watershed dan Klasifikasi Support Vector Machine (SVM) Sebagai Aplikasi Bidang Forensik," *J. Ris. Akunt. Going Concern*, vol. 3, no. 1, pp. 1177–1204, 2017.
- [3] S. A. Gadro, "Peran Odontologi forensik sebagai salah satu sarana pemeriksaan identifikasi jenazah tak dikenal," *Berkala Ilmu Kedokteran*, vol. 31, 1999.
- [4] A. Liana, B. Clara, I. B. Hidayat, Y. Malinda, and M. Kes, "Identifikasi Individu Berdasarkan Pola Sidik Bibir Menggunakan Metode Content Based Image Retrieval Based on Gray Level Co- occurrence Matrix dan Back Propagasi untuk Aplikasi Bidang Forensik," vol. 5, no. 1, pp. 480–487, 2018.
- [5] R. A. Surya, A. Fadlil, A. Yudhana, M. T. Informatika, P. T. Informatika, and U. A. Dahlan, "Ekstraksi Ciri Metode Gray Level Co-Occurrence Matrix (GLCM) dan Filter Gabor Untuk Klasifikasi Citra Batik Pekalongan," vol. 02, no. 02, pp. 23–26, 2017.
- [6] A. K. Putri, "Identifikasi Pola Sidik Bibir pada Identitas Manusia Menggunakan Metode Gray Level C0-occurrence Matrix dan Local Binary Pattern dengan Klasifikasi Dcision Tree Pattern sebagai Aplikasi Bidang Forensik Biometrik," 2018.
- [7] M. Ramadhani, H. B. D. K, and F. T. Elektro, "Klasifikasi Jenis Jerawat Berdasarkan Tekstur

- dengan Menggunakan Metode GLCM,” vol. 5, no. 1, pp. 870–876, 2018.
- [8] S. A. Zega, “Penggunaan Pohon Keputusan untuk Klasifikasi Tingkat Kualitas Mahasiwa Berdasarkan Jalur Masuk Kuliah,” pp. 7–13, 2014.
- [9] Z. Azmi, M. Dahria, P. Studi, S. Komputer, P. Studi, and S. Informasi, “Decision Tree Berbasis Algoritma Untuk Pengambilan Keputusan,” pp. 157–164, 1978.