

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/307433853>

Perbandingan Ekstraksi Ciri Statistik dan EFD dalam Pengenalan Tulisan Tangan Aksara Batak Toba

Conference Paper · November 2015

CITATION

1

READS

329

3 authors, including:



M. Jimmy Hasugian

Universitas Kristen Maranatha

13 PUBLICATIONS 23 CITATIONS

SEE PROFILE



Novie Theresia Pasaribu

Universitas Kristen Maranatha

14 PUBLICATIONS 14 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Klasifikasi Kondisi Jantung [View project](#)



AR MODEL UNTUK SINYAL ECG NORMAL DAN ARRHYTHMIA [View project](#)

Perbandingan Ekstraksi Ciri Statistik dan EFD dalam Pengenalan Tulisan Tangan Aksara Batak Toba

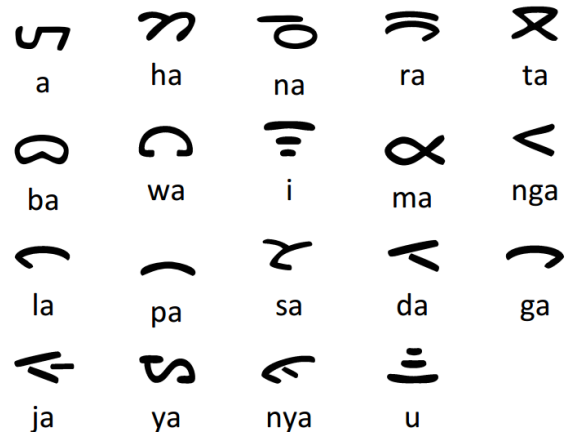
Meilan Jimmy Hasugian¹, Novie Theresia Br. Pasaribu²

^{1,2} Jurusan Teknik Elektro, Universitas Kristen Maranatha, Bandung
 jhasugian@maranatha.edu¹, novie.theresia @maranatha.edu²

Abstrak—Ekstraksi ciri statistik dan *Elliptic Fourier Descriptors* merupakan salah satu ekstraksi ciri yang cukup sering digunakan. Dalam penelitian ini kedua ekstraksi ciri tersebut digunakan untuk proses pengenalan tulisan tangan aksara Batak Toba. Ekstraksi ciri statistik menunjukkan hasil yang kurang baik, sehingga dilakukan penyeleksian ciri (*feature selection*). Penyeleksian ciri ini meningkatkan pengenalan sekitar 10-60%. Dari pengujian, diperoleh pengenalan tulisan tangan berdasarkan ciri EFD berkisar antara 40-100% dengan aksara “ra”, “ba”, “la”, “pa”, “ga”, “nya” mencapai 100%.

Kata Kunci—aksara Batak Toba, ekstraksi ciri statistik, ekstraksi ciri statistik terseleksi, *Elliptic Fourier Descriptor*

sembilan belas } Kesembilan belas aksara tersebut dapat dilihat pada gambar berikut:



Gbr. 1 *Ina ni surat* Aksara Batak Toba versi Modern

i. PENDAHULUAN

Aksara Batak Toba merupakan salah budaya Indonesia yang patut dilestarikan. Namun di kalangan suku Batak Toba penggunaan aksara ini masih sangat kurang. Salah satu upaya yang diusulkan penulis untuk pelestariannya dengan aksara Batak Toba dalam penelitian ini.

Aksara Batak Toba termasuk rumpun tulisan Brahmi (India), khususnya kelompok tulisan India Selatan. Aksara Batak Toba juga dapat diklasifikasikan sebagai abugida yaitu tulisan fonetis yang setiap bunyi bahasanya dapat dilambangkan secara akurat.^[1]

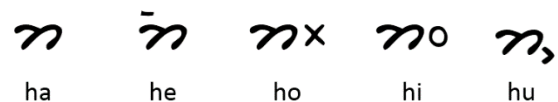
Aksara Batak Toba didominasi oleh *stroke* yang berbentuk lengkungan, yang memiliki kemiripan dengan tulisan Arab. Hal ini merupakan salah satu motivasi penggunaan ekstraksi ciri statistik dan *Elliptic Fourier Descriptor* (EFD) dalam proses pengenalan (bdk. Abandah, G.A, dkk. [2]).

Penelitian ini merupakan kelanjutan dari penelitian sebelumnya [8] untuk meningkatkan hasil pengenalan tulisan tangan aksara Batak Toba. Upaya ini dicapai dengan melakukan perbandingan proses pengenalan dari ekstraksi ciri statistik dan ekstraksi ciri EFD.

VII. STUDI LITERATUR

Dalam bahasa Batak, aksara Batak sering disebut *sia sia* atau *surat sampulu sia* karena jumlah aksara (*ina ni surat*) adalah sembilan belas^[1] {*sampulu sia* artinya

Tulisan aksara Batak Toba di atas adalah versi modern, yang masih lazim ditemukan dalam tulisan-tulisan sekarang serta yang dimasukkan ke dalam kurikulum pembelajaran sebagai muatan lokal di beberapa sekolah di Kabupaten Toba Samosir, Sumatera Utara^[2]. Penulisan aksara Batak Toba sama dengan penulisan huruf latin yaitu dari kiri ke kanan. Semua *ina ni surat* berupa konsonan yang berakhir dengan bunyi [a]. Bunyi [a] yang melekat pada *ina ni surat* dapat diubah menjadi vokal lain dengan menambahkan anak ni surat seperti yang tampak pada Gbr 2.



Gbr. 2 Perubahan bunyi vokal dengan penambahan *anak ni surat*

Sejauh ini ada beberapa publikasi yang telah membahas pengenalan aksara Batak Toba seperti Panggabean, M. dkk^[4] dan Sitinjak, SO^[5]. Ulasan mengenai kedua publikasi di atas dapat dilihat pada [8].

Tulisan ini dibuat untuk mengembangkan apa yang dipaparkan dalam tulisan [8] karena hasil pengenalan yang belum memuaskan. Tulisan ini membahas

perbandingan penggunaan ekstraksi ciri *statistical* dan EFD (*elliptic fourier descriptor*) serta menentukan ekstraksi ciri yang signifikan dalam peningkatan kinerja pengenalan tulisan tangan aksara Batak Toba.

Metode yang digunakan untuk proses pengenalan adalah metode CoDED (*Combination of Data using Euclidean Distance*).

VIII. EKSTRAKSI CIRI

Ekstraksi ciri yang digunakan dibagi menjadi 2 kategori yaitu statistik dan EFD. Masing-masing terdiri dari 14 dan 26 ciri.

A. Ciri Statistik^{[3][6]}

Ada empat belas ciri statistik yang digunakan pada penelitian ini, yaitu :

1. Area citra (A)

$$A = \sum_x \sum_y B(x, y) \quad (1)$$

Keterangan :

$B(x, y)$: Nilai intensitas citra biner

2. Lebar citra (W)
3. Tinggi citra (H)
4. Rasio Lebar terhadap Tinggi citra (W/H)
5. Rasio Kuadran Kanan Atas terhadap Area citra (UR/A)
6. Rasio Kuadran Kiri Atas terhadap Area citra (UL/A)
7. Rasio Kuadran Kiri Bawah terhadap Area citra (LL/A)
8. Rasio Kuadran Kanan Bawah terhadap Area citra (LR/A)
9. Pusat massa koordinat x (\bar{x})
10. Pusat massa koordinat y (\bar{y})
11. Momen pusat ternormalisasi ($\eta_{2,0}$)
12. Momen pusat ternormalisasi ($\eta_{0,2}$)
13. Normalisasi Pusat Massa koordinat x (\bar{x}_N)
14. Normalisasi Pusat Massa koordinat y (\bar{y}_N)

B. Elliptic Fourier Descriptors^{[3][6][7]}

Fourier descriptor telah banyak digunakan oleh para ahli^[7]. EFD (*Elliptic Fourier Descriptors*) dikembangkan oleh Kuhl dengan mencari *fourier coefficient* dari *chain-code* sebuah citra kontur^[7]. Keempat deskriptor untuk orde- n dapat dicari menggunakan rumus berikut:

$$a_n = \frac{T}{2n^2\pi^2} \sum_{i=1}^m \frac{\Delta x_i}{\Delta t_i} [\cos \phi_i - \cos \phi_{i-1}] \quad (2)$$

$$b_n = \frac{T}{2n^2\pi^2} \sum_{i=1}^m \frac{\Delta x_i}{\Delta t_i} [\sin \phi_i - \sin \phi_{i-1}] \quad (3)$$

$$c_n = \frac{T}{2n^2\pi^2} \sum_{i=1}^m \frac{\Delta y_i}{\Delta t_i} [\cos \phi_i - \cos \phi_{i-1}] \quad (4)$$

$$d_n = \frac{T}{2n^2\pi^2} \sum_{i=1}^m \frac{\Delta y_i}{\Delta t_i} [\sin \phi_i - \sin \phi_{i-1}] \quad (5)$$

dengan

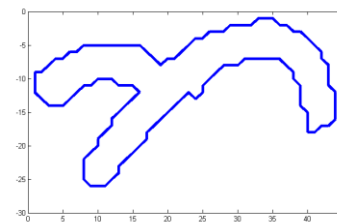
$$\phi_i = \frac{2n\pi t_i}{T}$$

$$\Delta x_i = x(i) - x(i-1) \quad \Delta y_i = y(i) - y(i-1)$$

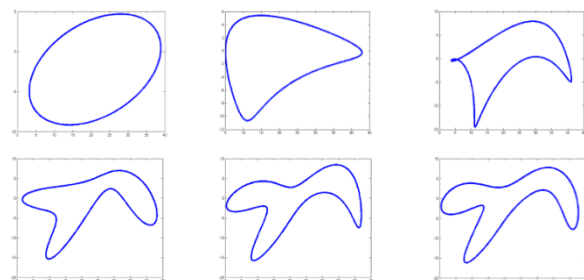
$$\Delta t_i = \sqrt{\Delta x_i^2 + \Delta y_i^2}$$

$$t_i = \sum_{j=1}^i \Delta t_j \quad T = t_m = \sum_{j=1}^m t_j \quad (6)$$

Dalam penelitian ini orde $n = 6$ dipilih karena hasil rekonstruksi dari *fourier coefficient* telah cukup baik.



Gbr. 3 Kontur dari Aksara Batak Toba "ha"



Gbr. 4 Rekonstruksi fourier coefficient orde 1-2-3 (atas) orde 4-5-6 (bawah) dari Aksara Batak Toba "ha"

IX. HASIL DAN PEMBAHASAN


Data tulisan tangan yang digunakan dalam penelitian ini berasal dari 10 naracoba, seperti tampak pada gambar berikut^[8]:

Aksara	Responden									
	1	2	3	4	5	6	7	8	9	10
a	5	5	5	5	5	5	5	5	5	5
ha	7	7	7	7	7	7	7	7	7	7
na	10	10	10	10	10	10	10	10	10	10
ra	11	11	11	11	11	11	11	11	11	11
ta	X	X	X	X	X	X	X	X	X	X
ba	8	8	8	8	8	8	8	8	8	8
wa	C	C	C	C	C	C	C	C	C	C
i	11	11	11	11	11	11	11	11	11	11
ma	8	8	8	8	8	8	8	8	8	8
nga	V	V	V	V	V	V	V	V	V	V
la	7	7	7	7	7	7	7	7	7	7
pa))))))))))
sa	N	N	N	N	N	N	N	N	N	N
da	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
ga	J	J	J	J	J	J	J	J	J	J
ja	4	4	4	4	4	4	4	4	4	4
ya	S	S	S	S	S	S	S	S	S	S
nya	E	E	E	E	E	E	E	E	E	E
u	11	11	11	11	11	11	11	11	11	11

Gbr. 5 Data Aksara tulisan tangan *ina ni surat* Aksara Batak Toba dari sepuluh responden

Contoh ekstraksi ciri statistik dan EFD dapat dilihat pada gambar berikut.


Ciri Statistik	
Ket	Nilai
A	424
W	44
H	26
W/H	1.692308
UR/A	0.382075
UL/A	0.353774
LR/A	0.051887



Ciri Statistik	
Ket	Nilai
LL/A	0.212264
\bar{x}	22
\bar{y}	11
$\eta_{2,0}$	0.358157
$\eta_{0,2}$	0.074888
\bar{x}_N	0.022727
\bar{y}_N	-0.11538

Gbr. 6 Nilai ekstraksi ciri Statistik aksara "ha"

Ciri Elliptic Fourier Descriptor	
Ket	Nilai
A0	21.04131
A1	-14.2326
A2	-1.5538
A3	-1.57496
A4	-2.05081
A5	-1.16582
A6	0.292471
B1	10.73186
B2	-4.2086
B3	-0.86903
B4	0.379797
B5	0.701348
B6	0.473671



Ket	Nilai
C0	-2.21666
C1	1.682814
C2	1.615367
C3	-1.47244
C4	0.240556
C5	-1.11542
C6	0.55309
D1	6.7125
D2	1.661775
D3	-3.86215
D4	2.087821
D5	1.334538
D6	0.466949

Gbr. 7 Nilai ekstraksi ciri EFD aksara "ha"

Proses pengenalan Aksara Batak Toba dilakukan dengan menggunakan metode **CoDED** (*Combination of Data using Euclidean Distance*) adalah sebagai berikut :

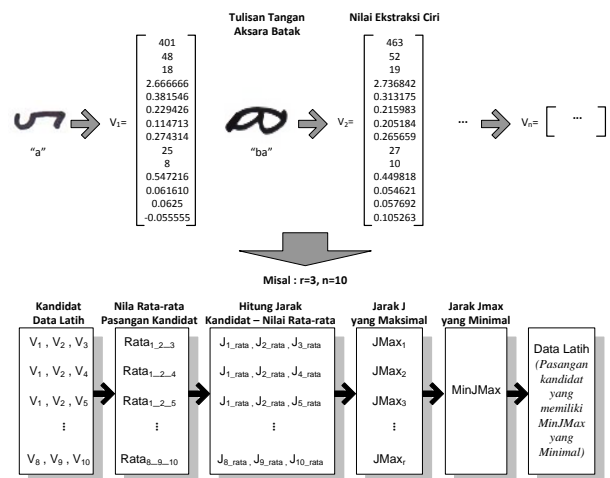
1. Penentuan Data Referensi

Penentuan data referensi (latih) yang tepat, akan memengaruhi hasil pengenalan tulisan tangan. Data referensi ini adalah data yang memiliki kemiripan (*similarity*) yang sangat dekat satu sama lain. Dalam penelitian ditentukan 3 data sebagai referensi untuk tiap aksara, sementara 7 data lainnya digunakan sebagai data uji. Skema penentuan data referensi adalah sebagai berikut:

- Menentukan semua kombinasi yang mungkin disusun jika terdiri dari 3 data (*class*). Dari 10 data, maka terdapat 120 *class* yang dapat disusun.
- Menentukan *internal-class similarity* (ICS) dari 120 *class*. ICS dihitung dengan mencari *Euclidean Distance* dari tiap *class member* terhadap nilai rata-rata *class*. Nilai *Euclidean distance* ini disebut sebagai nilai *radius* ICS
- Data atau *class* yang dipilih sebagai referensi adalah *class* dengan nilai *radius* ICS terkecil.

Langkah-langkah di atas dilakukan tiap aksara yang ada, sehingga akan diperoleh 19 vektor ciri (*featurevector*) untuk tiap aksara yang akan digunakan sebagai referensi.

Pada Gbr. 8 dapat dilihat skematik penentuan data referensi dengan metode CoDED.

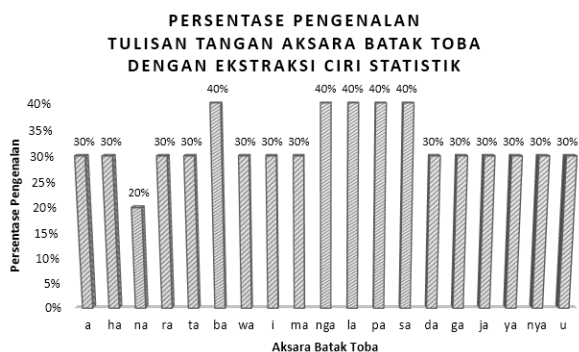


Gbr. 8 Contoh Skematik Penentuan Data Referensi Metode CoDED

2. Pengujian Data

Pengujian dilakukan dengan menentukan *Euclidean distance* dari data yang ada terhadap 19 vektor ciri aksara yang telah diperoleh pada tahap 1. Nilai *Euclidean distance* terkecil dianggap sebagai penentuan aksara dari data yang diuji.

Dengan menggunakan metode CoDED di atas, persentase keberhasilan pengenalan aksara tulisan tangan *ina ni surat* Aksara Batak Toba dapat dilihat pada gambar berikut.

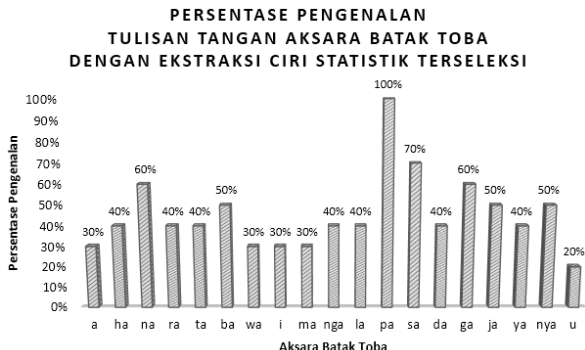


Gbr. 9 Persentase keberhasilan pengenalan Aksara Batak Toba berdasarkan ciri Statistik

Tampak dari diagram di atas, persentasi keberhasilan pengenalan aksara, berkisar antara 30% - 40%. Hasil ini masih dapat ditingkatkan dengan melakukan pemilihan ekstraksi ciri yang digunakan dalam proses pengenalan.

Pada penelitian ini dilakukan penyeleksian ekstraksi ciri statistik yang semula ada 14 ciri, menjadi 9 ciri, dengan membuang ciri: A, W, H, W/H, \bar{x} , dan \bar{y} . Pemilihan ciri tersebut yang dibuang karena data citra hasil pindai (*scan*) tidak dinormalisasi, serta beragamnya ketebalan penulisan tangan dari tiap responden.

Persentasi keberhasilan setelah menggunakan ciri statistik terseleksi dapat dilihat pada gambar berikut.

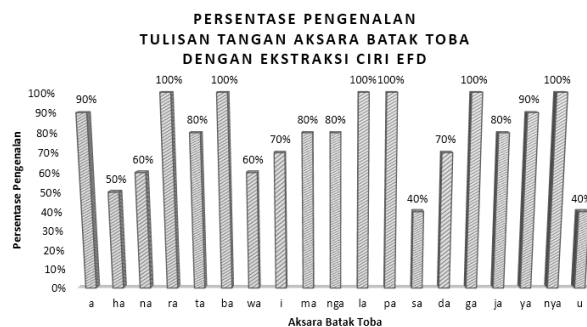


Gbr. 10 Persentase keberhasilan pengenalan Aksara Batak Toba berdasarkan ciri Statistik Terseleksi

Dari gambar di atas tampak bahwa persentase pengenalan menjadi meningkat apabila menggunakan ciri statistik yang telah diseleksi bahkan aksara "pa" yang semula 40% dikenali meningkat menjadi 100%.

Persentasi keberhasilan ini juga meningkat dibandingkan penelitian sebelumnya [8] dengan menggunakan metode CoDED.

Hasil pengenalan tulisan tangan dengan ekstraksi ciri EFD dapat dilihat pada Gbr 11. Dari diagram tersebut tampak bahwa persentasi keberhasilan berkisar antara 40% - 100%. Terdapat 6 dari 19 aksara yang berhasil dikenali 100%. Hasil ini jauh lebih baik dibandingkan pengenalan dengan ciri statistik



Gbr. 11 Grafik Persentase Pengenalan Tulisan Tangan Aksara Batak Toba dengan Ekstraksi Ciri EFD

Pada penelitian [8] ekstraksi ciri dari tiap aksara digabung antara ciri statistik dan EFD sehingga hasil pengenalan yang diperoleh kurang baik. Dengan pemisahan ciri ternyata dapat meningkatkan hasil pengenalan.

X. PENUTUP

Pengenalan tulisan tangan aksara Batak Toba menggunakan metode CoDED pada ciri EFD (*EllipticFourierDescriptor*) menunjukkan hasil yang lebih baik dibandingkan ciri statistik ataupun ciri statistik terseleksi. Ciri EFD membuat pengenalan tulisan tangan aksara Batak Toba mayoritas 100% (ada 6 dari 19 aksara), pengenalan 90% sebanyak 2 dari 19 aksara, lalu pengenalan 80% sebanyak 4 aksara.

REFERENSI

- [1] Kozok, Uli. Februari 2009. Surat Batak: Sejarah Perkembangan Tulisan Batak, Berikut Pedoman Menulis Aksara Batak dan Cap Sisimangaraja XII. KPG (Kepustakaan Populer Gramedia) & EFEO.
- [2] Hasil survei pribadi ke beberapa sekolah di Kabupaten Toba Samosir, Provinsi Sumatera Utara
- [3] Abandah, G., Fuad Jamour, Esam Qaralleh. 2014. Recognizing Handwritten Arabic Words using Grapheme Segmentation and Recurrent Neural Networks. International Journal on Document Analysis and Recognition (IJ DAR). Springer.
- [4] Panggabean, M. dan Rønningen, L. November 2009. Character Recognition of the Batak Toba Alphabet Using Signatures and Simplified Chain Code. Signal and Image Processing Applications (ICSIPA), IEEE, pp. 215 – 220.
- [5] Sitinjak, S. 2012. Pengenalan Tulisan Tangan Aksara Batak Toba Menggunakan Backpropagation. TESIS Program Studi Magister Teknik Informatika, Univ. Atmajaya, Yogyakarta.
- [6] Abandah, G., Khedher, M. 2009. Analysis of Handwritten Arabic Letters using Selected Feature Extraction Techniques. International Journal Computer Processing Language 22(1), 49–73
- [7] Kuhl, F., Giardina, C. 1982. Elliptic Fourier Features of a Closed Contour. Computer Graphic and Image Processing 18(3), 236–258
- [8] Pasaribu, N.T., M.J. Hasugian, Pengenalan Tulisan Tangan Ina ni surat Aksara Batak Toba. akan dipresentasikan pada SNTT 2015, Sekolah Vokasi UGM.