

Perbandingan Ekstraksi Ciri *Full, Blocks, dan Row Mean Spectrogram Image* dalam Mengidentifikasi Pembicara

Comparison of the Feature Extraction Methods's Full, Blocks, and Row Mean Spectrogram Image in Speaker Identification

La Ode Hasnuddin S Sagala^{*1}, Agus Harjoko²

¹Program Studi Ilmu Komputer, FMIPA UGM, Yogyakarta

²Jurusan Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta

e-mail: ^{*1}hasnuddinsagala@yahoo.com, ²aharjoko@ugm.ac.id

Abstrak

Pada sebuah sistem identifikasi pembicara, pemilihan metode ekstraksi ciri dan ukuran ciri yang digunakan mempengaruhi tingkat keakuratan identifikasi. Berkaitan dengan hal itu, dalam penelitian ini akan dijabarkan perbandingan tiga metode ekstraksi ciri CBIR yaitu row mean image, full image, dan blocks image. Ketiga metode tersebut digunakan untuk mengidentifikasi pembicara dengan menitikberatkan pada ukuran selection feature vector yang digunakan.

Data suara diperoleh dari rekaman suara menggunakan handphone. Rekaman suara berasal dari 10 orang narasumber dengan rincian 5 pria dan 5 wanita. Setiap narasumber mengucapkan lima buah kalimat yaitu Selamat Pagi, Selamat Siang, Selamat Sore, Selamat Malam, dan Dengan Siapa serta diulangi delapan kali tiap kalimat. Rekaman suara yang digunakan terlebih dahulu dikonversi menjadi image spectrogram menggunakan STFT. Spectrogram yang terbentuk kemudian diteruskan ke kekre transform lalu diekstraksi cirinya. Penggunaan kekre transform bertujuan untuk menyeleksi dan mengambil kemungkinan-kemungkinan ciri yang optimal serta juga meringankan proses komputasi.

Menggunakan data reference 250 image spectrogram dan data testing 150 image spectrogram memberikan hasil bahwa metode ekstraksi ciri full image memperoleh persentase identifikasi lebih tinggi yaitu 93,3% dengan ukuran fitur 32x32.

Kata kunci---Identifikasi pembicara, Spektrogram, Transformasi kekre, Full image, Blocks Image, Row mean image

Abstract

On a speaker identification system, selection extraction feature methods and feature size are used affect the accuracy of identification. In that regard, this study will present comparison three extraction feature CBIR methods namely full image, blocks image, and row mean image. The third methods are used for identifying the speaker with emphasis on the selection feature vector is used.

Sound data obtained from and recorded used mobile phone voice recording. Sound recordings are from 10 speakers with details of 5 men and 5 women. Every speaker pronounces the five sentences namely Selamat Pagi, Selamat Siang, Selamat Sore, Selamat Malam, and Dengan Siapa as well as each sentence was repeated eight times. The sound recordings used converted into spectrogram image using STFT. Spectrogram are formed then forwarded to in transform kekre for extracted feature. Use kekre transform aims to select and take the possibilities optimal characteristics also alleviate the computing process.

Using reference data 250 spectrogram image and testing data 150 spectrogram image provides results that the full image feature extraction methods obtain a higher percentage of identification is 93.3% with a feature size of 32x32.

Keywords---Speaker identification, Spectrogram, Kekre Transform, Full Image, Blocks Image, Row Mean Image

1. Pendahuluan

Pada sistem identifikasi pembicara atau *speaker identification*, proses ekstraksi ciri memainkan peranan penting dalam menghasilkan tingkat keakuratan yang baik. Terdapat banyak metode ekstraksi ciri untuk identifikasi pembicara telah diperkenalkan oleh para peneliti. Salah satunya adalah metode ekstraksi ciri berbasis *content* atau biasa dikenal CBIR (*Content Based Image Retrieval*). CBIR (*Content Based Image Retrieval*) merupakan metode ekstraksi ciri yang menggunakan *content* yang terdapat pada *image* sebagai fitur. *Content* yang digunakan dapat berupa warna, tekstur, *shape*, atau informasi-informasi lain yang mendukung namun *content* yang sering digunakan oleh para peneliti adalah warna, tekstur, dan *shape*.

Penerapan tehnik CBIR ke permasalahan identifikasi pembicara terlebih dahulu dilakukan proses konversi sinyal suara menjadi *image* dengan menggunakan *short time fourier transform* atau STFT. Gambar yang dihasilkan oleh STFT berupa *spectrogram* yang merupakan penampakan spektrum-spektrum frekuensi yang diplot terhadap waktu dan amplitudo. *Spectrogram* yang digunakan dalam identifikasi pembicara memiliki keuntungan dapat menganalisa suara seseorang sekalipun orang tersebut mencoba untuk menghilangkan karakter asli dari suaranya (Al-Azhar, 2011). Penggunaan *content* warna sangat memungkinkan jika diterapkan pada *spectrogram*. Namun penggunaannya terkadang mempengaruhi proses komputasi dan ukuran fitur apalagi jika gambarnya dalam bentuk RGB maka diperlukannya proses setiap *plane*. Pendekatan CBIR ke dalam transform domain dapat menjadi sebuah solusi pengganti *content* berdasarkan warna (Kekre dan Thepade, 2009). Penggunaan metode transform bermakna adanya proses transformasi sebelum proses ekstraksi ciri dilakukan.

Salah satu contoh penelitian mengenai ekstraksi ciri CBIR pendekatan *transform domain* adalah yang dilakukan Kulkarni dan Kekre (2011) mengenai identifikasi pembicara. DCT dan Walsh Hadamard transform digunakan sebagai metode transformasi yang akan diimplementasikan ke dalam gambar sebelum diekstraksi cirinya. Setelah proses transformasi, *row mean image* akan menjadi metode untuk mengambil fitur-fitur yang terdapat pada *image*. Penggunaan ukuran fitur 64 sampai dengan 8192 akan menjadi perbandingan dalam proses ekstraksi cirinya. Hasilnya menunjukkan bahwa ukuran fitur yang kecil yaitu berukuran 512 menunjukkan pengenalan yang lebih baik.

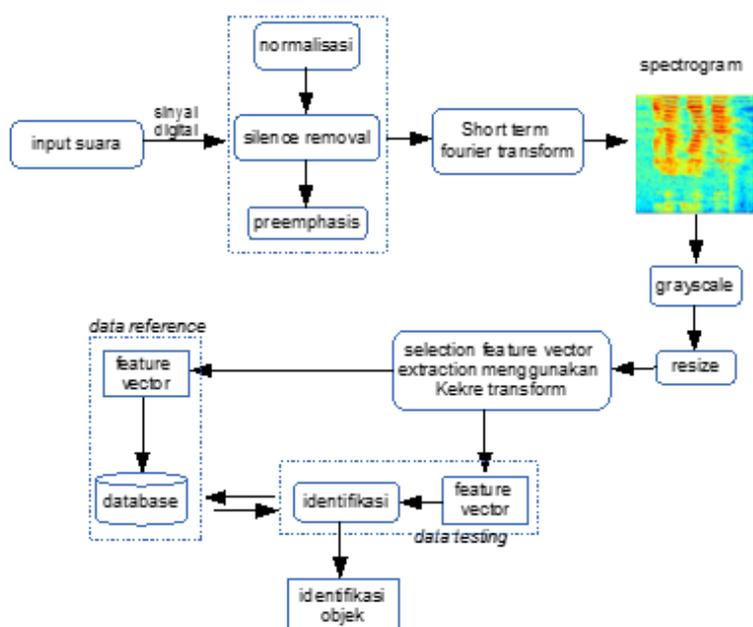
Persentase pengenalan yang baik pada ukuran fitur yang kecil juga di buktikan pada penelitian (Sarode, 2011) mengenai pengenalan wajah. Penelitiannya juga menggunakan DCT dan Walsh sebagai metode transformasinya namun metode ekstraksi ciri yang digunakan adalah *full image*. Hasilnya menunjukkan bahwa dengan menggunakan variasi ukuran fitur 4x4 sampai dengan 128x128, ukuran fitur 8x8 memberikan pengenalan yang baik dibandingkan ukuran fitur yang lain. Selain penggunaan *row mean image* dan *full image* yang dapat memberikan persentase pengenalan yang besar pada ukuran fitur yang kecil, *blocks image* juga dapat memberikan pengenalan yang baik pada ukuran fitur yang kecil seperti pada penelitian (Sarode, 2010) mengenai identifikasi pembicara, akan tetapi cuma menggunakan satu metode transform yakni DCT.

Adanya perbandingan metode ekstraksi ciri yang digunakan namun tetap memberikan hasil pengenalan yang baik pada ukuran fitur yang kecil telah mendorong untuk mengkaji lebih dalam dari ketiga metode ekstraksi ciri tersebut. Oleh karena itulah, melalui penelitian ini akan ditinjau bagaimana perbandingan ketiga ekstraksi ciri yaitu *full image*, *row mean image*, dan *blocks image* dalam mengidentifikasi pembicara dengan meninjau pada ukuran *selection feature vector* akan tetapi menggunakan kekre transform sebagai metode

transformasinya serta manhattan distance sebagai metode *feature matching*-nya (Kulkarni, 2012).

2. Metode Penelitian

Proses ekstraksi fitur akan lebih difokuskan pada penelitian ini. Terdapat tiga metode ekstraksi fitur *CBIR transform domain* yang akan dibandingkan keakuratannya dalam mengidentifikasi pembicara yaitu *full image*, *blocks image*, dan *row mean image*. Karena menggunakan pendekatan *transform domain* maka sebelum pembentukan *feature vector*, sampel yang akan diproses terlebih dahulu di transformasikan menggunakan kekre transform. Proses yang dilakukan dalam penelitian ini digambarkan dalam bentuk gambar alir seperti tampak pada Gambar 1. Berdasarkan pada Gambar 1 memperlihatkan sistem terbagi dalam dua kelompok yaitu *reference speech signal* dan *test speech signal* dimana proses pembentukan *feature vector* tiap kelompok adalah sama.



Gambar 1. Sistem Identifikasi Pembicara

2.1 Input Suara

Proses pertama dalam sistem ini adalah pengambilan data rekaman suara menggunakan handphone Nexian bertipe NX-G381i. Menggunakan *sampling frequency* 8000 Hz dan *resolution* 16bps, suara pembicara yang bersifat analog akan diubah menjadi sinyal bersifat diskrit atau berbentuk sinyal digital.

Rekaman suara yang diambil berasal dari 10 pembicara dengan rincian 5 pembicara wanita dan 5 pembicara pria. Proses perekamannya dimulai dengan mengucapkan 5 buah kalimat oleh setiap pembicara dan setiap kalimat diulangi sebanyak 8 kali serta disimpan dalam bentuk file WAV. Kalimat tersebut adalah Selamat Pagi, Selamat Siang, Selamat Sore, Selamat Malam dan Dengan Siapa. Rekaman suara yang terbentuk akan dibagi menjadi dua bagian yaitu 5 rekaman sebagai *reference signal* dan 3 rekaman sebagai *test signal* untuk setiap kalimat.

2.2 Proses preprocessing

Setelah rekaman suara berbentuk sinyal digital dan disimpan dalam bentuk file WAV, rekaman tersebut kemudian dikenakan proses *preprocessing*. Proses *preprocessing* bertujuan untuk memperbaiki rekaman suara agar menghasilkan sinyal rekaman suara yang baik dan pengaruh-pengaruh *noise* pada rekaman dapat dikurangi. Pada penelitian ini, proses *preprocessing* terdiri dari tiga bagian yaitu normalisasi, *silence removal*, dan *pre-emphasis*.

Perbedaan intensitas suara dan jarak media rekaman saat proses perekaman terkadang mempengaruhi nilai amplitudo yang dihasilkan. Oleh karena itu, langkah pertama yang dilakukan pada proses *preprocessing* adalah proses normalisasi. Proses normalisasi bertujuan untuk menyamakan/ menyeragamkan besar amplitudo setiap sinyal rekaman suara dengan cara membagi data ke- i pada sampel rekaman sinyal suara dengan nilai amplitudo maksimum suara yang bersangkutan. Sehingga setiap sampel rekaman suara akan memberi nilai amplitudo maksimum sebesar 1 (satu). Proses normalisasi dilakukan dengan menggunakan persamaan (1).

$$S[i] = \frac{S[i]}{\max_{1 \leq i \leq N} |S[i]|}, \quad i = 1, 2, 3, \dots, N \quad (1)$$

dimana, $S[i]$ merupakan data ke- i pada sampel suara dan $\max_{1 \leq i \leq N} |S[i]|$ merupakan amplitudo maksimum pada sinyal suara tersebut.

Setelah proses normalisasi, setiap sampel rekaman suara yang terbentuk pasti mengandung daerah *silence*. Daerah *silence* pada suara biasanya terletak pada awal dan akhir rekaman suara serta tidak mengandung kalimat yang diucapkan oleh pembicara. Adanya daerah *silence* terkadang mempengaruhi penggunaan waktu pada proses komputasi sehingga menghasilkan waktu yang lama. Oleh karena itu, tahap kedua pada proses *preprocessing* adalah proses *silence removal* yang berfungsi untuk mendeteksi serta menghapus daerah *silence*. Proses *silence removal* pada penelitian ini menggunakan pustaka yang dibuat oleh [7] yang dimana penentuan daerahnya berdasarkan besar *signal energy* dan *spectral centroid* yang dihasilkan setiap sampel rekaman suara.

Setelah nilai amplitudo setiap suara diseragamkan dan daerah *silence* di hapus, sampel suara yang akan digunakan seringkali juga terdapat *noise* yang bercampur dengan kalimat yang diucapkan pembicara saat proses perekaman. Oleh karena itu, proses terakhir dari *preprocessing* adalah proses *pre-emphasis*. Proses *pre-emphasis* bertujuan untuk memperbaiki sinyal suara dari gangguan *noise* [8] tetapi tetap mempertahankan frekuensi-frekuensi tinggi dari rekaman suara [9] yang di dalamnya mengandung kalimat yang diucapkan pembicara saat proses perekaman. Persamaan (2) menunjukkan rumus yang digunakan dalam proses *pre-emphasis*.

$$y[i] = S[i] - \alpha S[i - 1] \quad , \quad 0,9 \leq \alpha \leq 1 \quad (2)$$

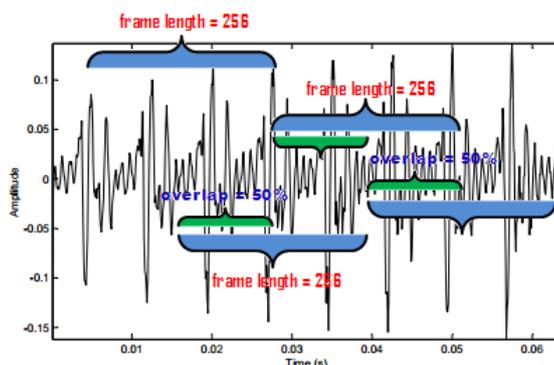
dimana $y[i]$ merupakan data ke- i pada sampel suara hasil preemphasis dan $S[i]$ merupakan data ke- i pada sampel suara sebelum proses *preemphasis*. Nilai α yang digunakan adalah 0.97

2.3 Short Term Fourier Transform (STFT)

Rekaman suara yang telah diperbaiki kualitasnya dan diminimalisir *noise*-nya kemudian ditransformasikan ke dalam bentuk *image* menggunakan *short term fourier transform*

(STFT). STFT atau *short term fourier transform* merupakan salah satu analisis fourier yang biasa digunakan untuk mengubah sinyal suara menjadi *image* berbentuk *spectrogram*. *Spectrogram* yang terbentuk merupakan penampakan spektrum-spektrum frekuensi yang diplot terhadap waktu dan amplitudo. Pembentukan *spectrogram* menggunakan STFT secara sederhana terdiri atas tiga proses yaitu *frame blocking*, *windowing*, dan *fast fourier transform*[10].

Pada penelitian ini, *frame blocking* merupakan langkah yang pertama kali dilakukan dalam proses pembentukan *spectrogram*. Proses pembentukan frame dalam penelitian ini menggunakan *overlap* 50% dengan panjang frame pada satu sinyal suara adalah 256 seperti yang ditunjukkan pada Gambar 2.



Gambar 2. Proses frame blocking

Proses *frame blocking* yang dilakukan dalam penelitian ini berfungsi untuk membagi sinyal suara menjadi potongan-potongan kecil (atau frame) yang mempermudah proses perhitungan dan analisis suara. Namun disisi lain. Proses *frame blocking* dapat menyebabkan sinyal suara bersifat *discontinue/non-stationary* yang dapat menimbulkan frekuensi-frekuensi baru saat dilakukan proses *fast fourier transform* atau FFT. Oleh karena itu, untuk mencegah hal tersebut perlu adanya proses mengubah sinyal suara dari bersifat *discontinue* menjadi *continue*. Salah satu caranya adalah dengan menggunakan proses hamming window. Persamaan (3) menunjukkan rumus hamming window yang digunakan.

$$w[i] = 0.52 - 0.46 \cos\left(\frac{2\pi i}{N-1}\right) \quad , \quad 0 \leq i \leq N-1 \quad (3)$$

dimana, $w[i]$ merupakan hasil windowing

Hasil kali antara frame sinyal dengan windowing akan menjadi input untuk proses transformasi fourier. *Fast fourier transform* merupakan metode transform yang digunakan untuk mengubah sinyal suara dalam domain waktu menjadi sinyal suara dalam domain frekuensi. Hasil dari transform ini berupa spektrum-spektrum frekuensi suara.

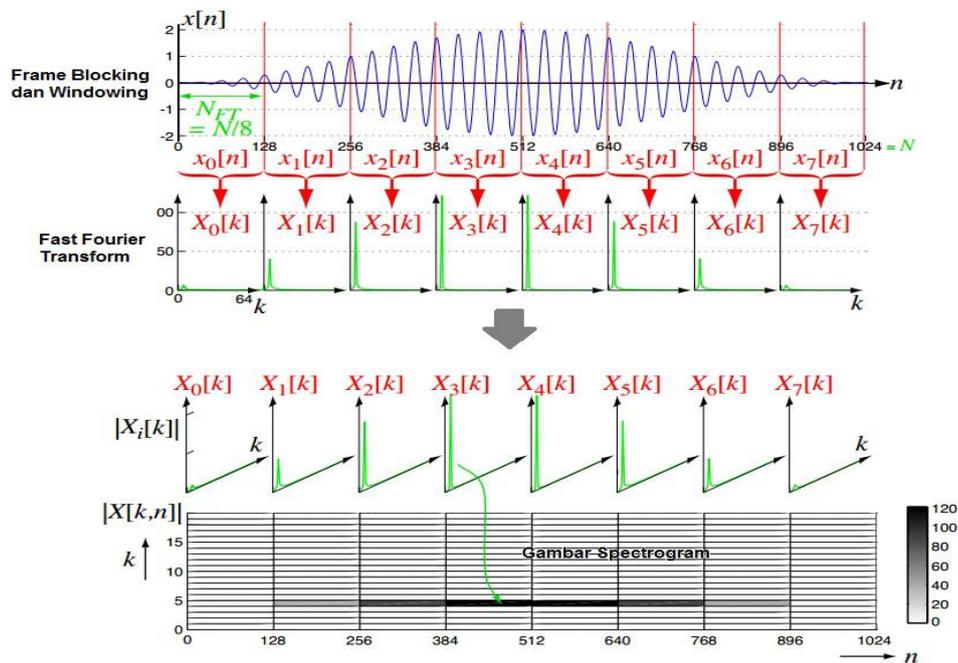
Rumus FFT seperti ditunjukkan pada persamaan (4) dan untuk mencari nilai spektrum frekuensinya menggunakan persamaan (5)

$$X[k] = \sum_{i=1}^N x[i] * w[i] \left(\cos\left(\frac{2\pi ki}{N}\right) - j \sin\left(\frac{2\pi ki}{N}\right) \right) \quad (4)$$

$$|F(k)| = |R^2 + I^2|^{1/2} \quad (5)$$

dimana, $X[k]$ merupakan hasil proses FFT, $x(i)$ merupakan data ke- i pada sampel suara, $F(k)$ merupakan spektrum frekuensi, R merupakan bilangan *real* hasil perhitungan dan I

merupakan bilangan *imajiner* hasil perhitungan. Secara keseluruhan, Gambar 3 menunjukkan proses pembentukan *spectrogram*.



Gambar 3. Pembentukan *Spectrogram*

2.4 Grayscale dan Resize

Proses grayscale merupakan proses konversi gambar dari RGB menjadi *grayscale* dengan tujuan untuk mempermudah komputasi dan proses *resize* bertujuan untuk menormalisasikan ukuran tiap *spectrogram* yang terbentuk. Ukuran normalisasi yang digunakan adalah 256x256.

2.5 Selection Feature Vector Extraction Menggunakan Kekre

Proses ekstraksi ciri dalam penelitian ini mengikutsertakan kekre transform sebagai metode untuk menyaring kemungkinan-kemungkinan fitur yang memberikan *property* yang baik [11]. Penentuan *feature vectornya* menggunakan tiga metode yaitu *full image*, *row mean image*, dan *blocks image* yang diimplementasikan menggunakan kekre transform seperti pada persamaan (6).

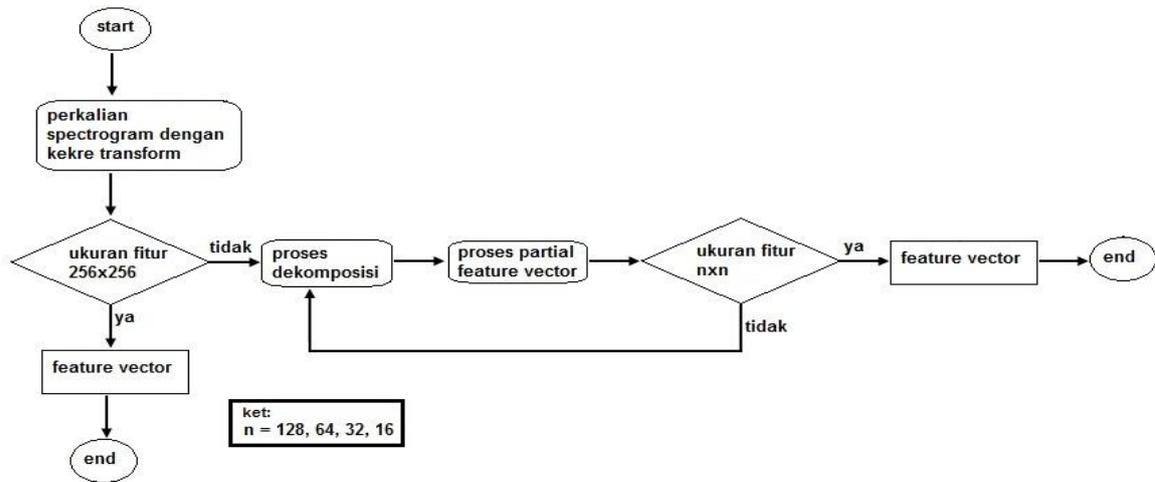
$$[A] = [K][I][K]^T \tag{6}$$

dimana $[A]$ merupakan hasil transformasi, $[K]$ merupakan matriks kekre, $[I]$ merupakan gambar, dan $[K]^T$ adalah matriks *kekre transpose*.

Full image

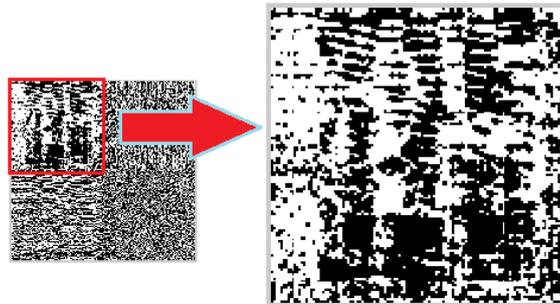
Ekstraksi ciri yang pertama dilakukan adalah ekstraksi ciri *full image*. Variasi ukuran ciri yang digunakan adalah 256x256, 128x128, 64x64, 32x32, dan 16x16. Ukuran ciri 256x256 tidak akan mengalami proses dekomposisi sedangkan ukuran selain 256x256 akan

mengalami proses dekomposisi. Proses ekstraksi ciri *full image* seperti ditunjukkan pada Gambar 4.



Gambar 4. Bagan alir proses ekstraksi ciri *full image*

Berdasarkan Gambar 4, proses ekstraksi ciri *full image* memiliki dua perlakuan yang berbeda tergantung pada kondisi ukuran fitur yang akan digunakan. Pada ukuran 256x256, proses pembentukan feature vector diawali dengan proses perkalian spectrogram dengan kekre transform atau biasa dikenal dengan istilah proses transformasi. Sedangkan untuk ukuran selain 256x256, setelah proses transformasi akan mengalami proses dekomposisi dan proses partial feature vector. Gambar 5 menunjukkan hasil proses dekomposisi dan proses partial feature vector.



Gambar 5. Proses dekomposisi sekaligus proses partial feature vector

Proses dekomposisi bertujuan untuk menghasilkan komponen berfrekuensi rendah yang identik dengan bebas noise dan mengandung informasi global yang menjadi karakteristik dari gambar. Karena hasil proses dekomposisi berupa 4 subimage dan hanya subimage yang berada pada pojok kiri atas yang memiliki frekuensi rendah maka untuk memisahkannya dapat menggunakan partial feature vector.

Row mean image

Ekstraksi ciri yang kedua dilakukan adalah ekstraksi ciri *row mean image*. Ekstraksi ciri ini diperoleh dengan cara mengambil nilai *mean* (rata-rata) piksel tiap baris dan hasilnya disimpan sebagai ciri. Variasi ukuran yang digunakan adalah sama dengan variasi ukuran

yang digunakan pada *full image*. Proses ekstraksi ciri *row mean image* seperti ditunjukkan pada Gambar 6.



Gambar 6. Bagan alir ekstraksi ciri row mean image

Blocks Image

Ekstraksi ciri yang ketiga atau terakhir digunakan adalah ekstraksi ciri *blocks image*. Variasi ukuran yang digunakan sama dengan variasi ukuran yang digunakan *full image* dan *row mean image*. Seperti *full image*, ekstraksi ciri *blocks image* mengalami dua perlakuan yang berbeda yaitu untuk ukuran 256x256 prosesnya dimulai dengan membagi *image spectrogram* menjadi empat bagian *subimage* kemudian setiap *subimage* diimplementasikan ke dalam kekre transform lalu diambil *feature vector*-nya. Namun untuk ukuran selain 256x256 perlakuannya adalah sebagai berikut :

1. Membagi *image spectrogram* menjadi *subimage* dengan ukuran $256/n \times 256/n$, dimana n merupakan variasi ukuran yang telah ditetapkan.
2. Mencari nilai rata-rata dari setiap *subimage* yang bertujuan sebagai perwakilan nilai piksel setiap *subimage*.
3. Menggabungkan tiap *subimage* agar membentuk *image* baru.
4. *Image* baru tersebut kemudian di *block* menjadi 4 bagian *subimage*.
5. Setiap *subimage* diimplementasikan ke dalam kekre transform
6. Hasil perkaliannya akan berupa *feature vector* yang disimpan sebagai ciri.

2.6 Identifikasi

Proses terakhir dalam penelitian ini adalah proses identifikasi. Metode *manhattan distance* menjadi eksekutor untuk mengidentifikasi pembicara dengan cara mencari jarak terkecil antara sampel dengan semua data yang terdapat pada database. Jika ditemukan jarak terkecil maka data tersebut akan menjadi kesimpulan dalam proses identifikasi ini. Persamaan (7) merupakan rumus manhattan distance yang digunakan.

$$d_{ij} = \sum_{k=1}^n |x_{ik} - x_{jk}| \quad (7)$$

dimana, x_{ik} merupakan data reference, x_{jk} merupakan data testing, dan d_{ij} merupakan nilai jarak.

2.7 Accuracy of Identification

Setelah proses identifikasi setiap proses ekstraksi ciri, maka dilanjutkan dengan penentuan persentase besarnya jumlah pembicara yang dikenali. Tujuannya adalah untuk mengetahui ekstraksi ciri yang mana yang memberikan persentase identifikasi yang besar dan pada ukuran fitur berapa. Persamaan (8) menunjukkan rumus yang digunakan.

$$Accuracy(\%) = \frac{\text{number of matches}}{\text{number of samples tested}} \times 100\% \quad (8)$$

3. Hasil dan Pembahasan

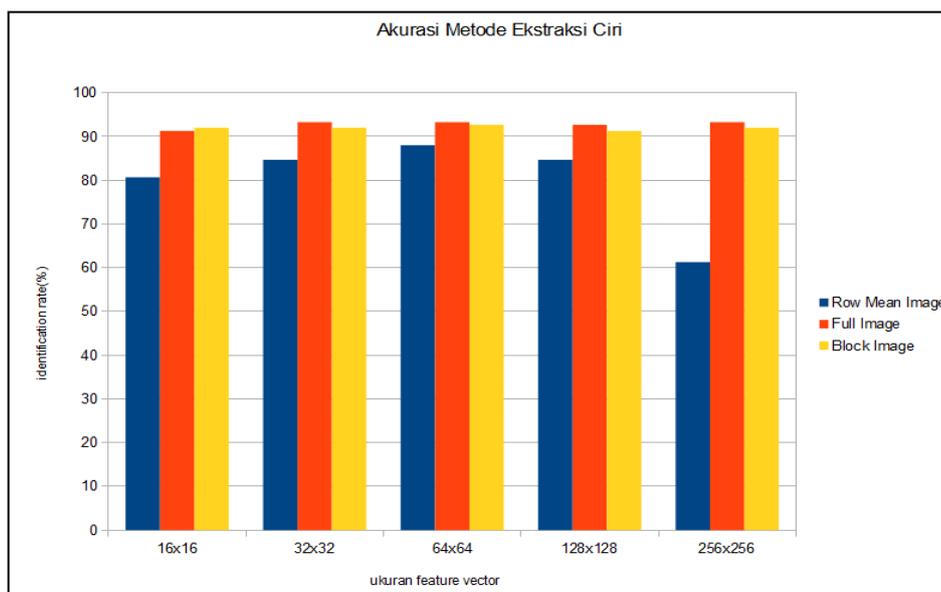
Pada penelitian ini, tiga metode ekstraksi ciri CBIR dengan pendekatan transform dibandingkan keakuratannya dalam mengidentifikasi pembicara dengan meninjau pada ukuran fitur yang digunakan. Berdasarkan pengamatan dengan menggunakan data *reference* 250 *image spectrogram* dan data *testing* 150 *image spectrogram* menunjukkan bahwa ekstraksi ciri *full image* memberikan persentase pengenalan lebih besar dibandingkan *row mean image* dan *blocks image*.

Pada ekstraksi ciri *full image* mengenali 140 pembicara atau setara dengan 93,3%. Ekstraksi ciri *blocks image* mengenali 139 pembicara atau setara dengan 92,7 % sedangkan ekstraksi ciri pada *row mean image* mengenali 132 pembicara atau setara dengan 88%. Tabel 1 menunjukkan hasil pengenalan setiap ekstraksi ciri.

Tabel 1. Hasil persentase setiap metode ekstraksi ciri

No	Ukuran	Dikenali (%)		
		Full image	Blocks image	Row mean image
1	16x16	91,3	92	80,7
2	32x32	93,3	92	84,7
3	64x64	93,3	92,7	88
4	128x128	92,7	91,3	84,7
5	256x256	93,3	92	61,3

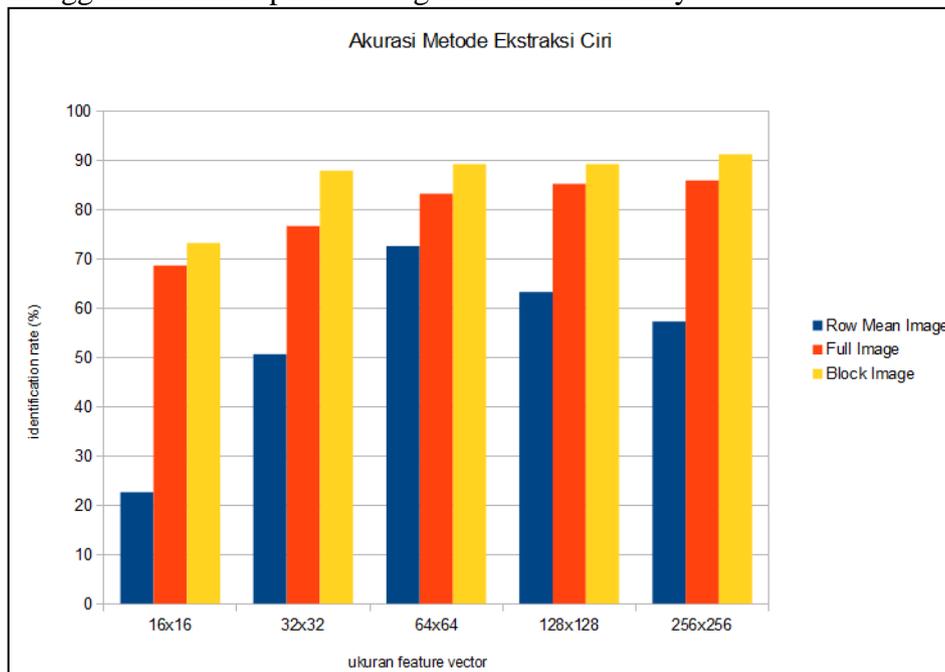
Berdasarkan Tabel 1, ekstraksi ciri *full image* memberikan pengenalan yang lebih besar pada ukuran fitur 32x32, 64x64, 256x256. Namun karena penelitian ini lebih memfokuskan pada ukuran fitur yang digunakan maka ukuran fitur 32x32 diambil sebagai ukuran fitur yang terbaik dalam penelitian ini dan menobatkan ekstraksi ciri *full image* sebagai ekstraksi ciri yang memberikan persentase pengenalan lebih besar. Gambar 6 menunjukkan perbedaan persentase akurasi tiap ekstraksi ciri.



Gambar 6. Diagram batang perbandingan akurasi metode ekstraksi ciri rekaman suara menggunakan handphone

Selain mengambil dari rekaman handphone, pada penelitian ini juga mencoba mengambil rekaman suara yang berasal dari microphone merk X-Tech dengan jenis XH-318. Suara yang direkam disampling dengan *frequency* 8000 Hz dan resolution 16bps. Percobaan ini cuma berfungsi sebagai pembandingan jika ditinjau dari media rekaman yang digunakan.

Hasil yang diperoleh menunjukkan bahwa ekstraksi ciri *blocks image* memberikan persentase yang besar dibandingkan dua metode ekstraksi ciri yang lain. Hal ini sangat berbeda jika terjadi pada rekaman suara yang menggunakan handphone, dimana pada rekaman suara tersebut justru yang menunjukkan persentase pengenalan lebih besar adalah ekstraksi ciri *full image*. Gambar 7 menunjukkan perbedaan persentase tiga metode ekstraksi ciri jika menggunakan microphone sebagai media rekamannya.



Gambar 7. Diagram batang perbandingan akurasi metode ekstraksi ciri rekaman suara menggunakan microphone

Adanya perbedaan yang dihasilkan baik yang menggunakan microphone ataupun handphone menunjukkan bahwa faktor alat rekaman yang digunakan juga mempengaruhi persentase identifikasi yang dihasilkan. Selain itu juga keragaman lingkungan, keragaman pembicara maupun kondisi perekaman dapat menjadikan tolak ukur mengenai persentase identifikasi yang baik.

4. Kesimpulan

Berdasarkan hasil pengamatan terhadap tiga metode ekstraksi ciri dalam mengidentifikasi pembicara maka dapat disimpulkan sebagai berikut :

- Ekstraksi ciri menggunakan *full image* memberikan persentase lebih besar dibandingkan *row mean image* dan *blocks image*.
- Ukuran fitur yang memberikan persentase pengenalan lebih besar adalah *full image* pada ukuran 32x32 sedangkan *blocks image* dan *row mean image* masing-masing pada ukuran 64x64.

Daftar Pustaka

- Al-Azhar, M.N., 2011, *Audio Forensic: Theory and Analysis*, Pusat Laboratorium Forensik Polri Bidang Fisika dan Komputer Forensik.
- Giannakopoulos, 2010, T., *A Method for Silence Removal and Segmentation of Speech Signals Implemented in Matlab*, Department of Informatics and Telecommunications, University of Athens, Greece.
- Ifeachor, Emmanuel, C. And Barrie, W.J., 2002, *Digital Signal Processing*, Prentice Hall, USA.
- Kekre, H.B. And Shah, K., 2009d, Performance Comparison of Kekre's Transform with PCA and Other Conventional Orthogonal Transforms for Face Recognition, Second Int. Conf. On Emerging Trends in Engineering and Technology, pp. 873-879.
- Kekre, H.B. And Thepade, D.S., 2009, Rendering Futuristic Image Retrieval System, *National Conference on Enhancements in Computer Communication and Information Technology*.
- Kulkarni, V. and Kekre, H.B., 2011a, Speaker Identification Usiang Row Mean of DCT and Walsh Hadamard Transform, *Int. J. on Computer Science and Engineering*, Vol 03(03), pp.1295-1301.
- Kulkarni, V., Kekre, H.B., Gaikar, P. and Gupta, N., 2012, Speaker Identification using Spectrogram of Varying Frame Sizes, *Int. J. of Computer Applications*, Vol 50(20), pp. 27-33.
- Manunggal, H.S., 2005, Perancangan dan Pembuatan Perangkat Lunak Pengenalan Suara Pembicara Dengan Menggunakan Analisa MFCC Feature Extraction, *Skripsi*, Tehnik Informatika Universitas Kristen Petra, Surabaya.
- Sarode, T., Kekre, H.B., Natu, S. and Natu, P., 2010, Performance Comparison Of 2-D On Full/Block spectrogram And 1-D DCT On Row Mean Of Spectrogram For Speaker Identification, *Int. Journal of Biometrics and Bioinformatics*, Vol 04(03), pp. 100-112.
- Sarode, T., Kekre, H.B., Natu, S. and Natu, P., 2011, Performance Comparison Of Face Recognition Using DCT and Walsh Transform with Full and Partial Feature Vector against KFCG VQ Algorithm, *Int. Conference and workshop on Emerging Trends in Technology*, pp. 22-30.
- Yee, C.S., 2009, Speaker Identification Using Artificial Network By Combination Of Resilient back-propagation And Gradient Descent, Universiti Teknologi Malaysia.