

Rancangan Sistem Identifikasi Jenis Burung Kicau Berdasarkan Suara Burung dengan Mel Frequency Cepstrum Coefficiens (MFCC)

Terry Anda Putra Nurarinda¹, Julian Sahertian², Umi Mahdiyah³

^{1,2}Teknik Informatika, Fakultas Teknik, Universitas Nusantara PGRI Kediri

E-mail: ¹terryanda013@gmail.com, ²julian.sahertian@gmail.com, ³umimahdiyah@gmail.com

Abstrak – Identifikasi Jenis Burung Kicau Berdasarkan Suara Burung Menggunakan Pengenalan Suara. Perkembangan teknologi yang semakin pesat memberikan kemudahan bagi masyarakat pecinta burung kicau untuk secara otomatis mengidentifikasi jenis burung berdasar suara kicaunya. Burung selain memiliki keindahan fisik juga memiliki keindahan suara. Teknologi selain berperan penting sebagai media untuk bertukar informasi mengenai seputar tips dalam merawat burung, juga diharapkan untuk dapat memberikan fungsi lain, yaitu mendeteksi suara burung, sehingga masyarakat yang memiliki hobi serupa dapat dengan mudah membedakan jenis burung berdasarkan suara kicaunya. Penelitian ini menghasilkan sebuah model identifikasi otomatis jenis burung. Data yang digunakan berupa data rekaman suara yang didapat dari komunitas penghobi burung di setiap perlombaan burung kicau. Berdasarkan pengetahuan tentang suara kicau burung untuk setiap jenis burung terdapat perbedaan yang cukup signifikan, maka dapat disimpulkan bahwa identifikasi jenis burung dapat dilakukan berdasarkan suara kicaunya. Sehingga hal ini dapat memberikan kemudahan pada komunitas pecinta burung kicau.

Kata Kunci — MFCC, Pengenalan Suara Burung, Rekayasa Perangkat Lunak.

1. PENDAHULUAN

Perkembangan teknologi yang semakin pesat dapat memberikan kemudahan bagi masyarakat dalam menyalurkan hobi, misalnya hobi mendengarkan suara burung dalam komunitas kicau mania. Menurut Anissa 2012 burung selain memiliki keindahan fisik juga memiliki keindahan suara [1]. Di Indonesia burung berkicau merupakan salah satu jenis burung yang paling diminati oleh masyarakat pecinta burung kicau, hal tersebut dapat dilihat dari banyaknya masyarakat yang memulai hobinya dalam memelihara burung dan juga perlombaan suara kicau burung yang sering dan rutin diselenggarakan pada setiap daerah.

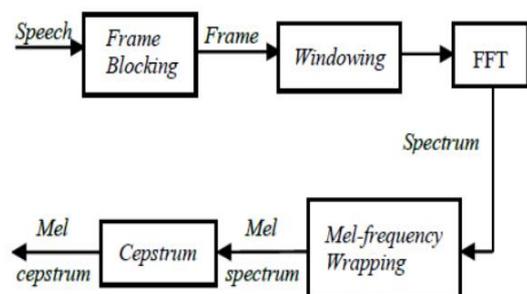
Berdasarkan pengetahuan tentang suara kicau burung untuk setiap jenis burung terdapat perbedaan yang cukup signifikan, maka dapat disimpulkan bahwa klasifikasi jenis burung dapat dilakukan berdasarkan suara kicaunya. Sehingga hal ini dapat memberikan kemudahan pada komunitas pecinta burung kicau. Dengan banyaknya jenis burung kicau di Indonesia tentunya akan membuat kesulitan bagi para pecinta burung kicau untuk mengidentifikasi jenis burung yang disukai. Oleh karena itu diperlukan sejumlah cara untuk mengidentifikasi jenis-jenis burung kicau.

Rancangan penelitian ini terinspirasi oleh penelitian yang dilakukan Destian Tri Handoko, Patmi Kasih, 2018, mengenai pengenalan suara untuk keamanan PC [2]. Dalam rancangan penelitian ini dibuat suatu rancangan sistem yang dapat mengidentifikasi jenis burung berdasarkan suara burung menggunakan menggunakan metode Mel

Frequency Cepstrum Coefficiens (MFCC). Diharapkan rancangan sistem tersebut dapat memberikan kemudahan pada komunitas pecinta burung kicau dalam mengidentifikasi jenis burung berdasarkan suara kicaunya.

2. METODE PENELITIAN

Menurut Sanjaya 2014 metode MFCC merupakan cara yang paling sering digunakan untuk pemrosesan suara, karena dianggap cukup baik dalam mengolah sinyal suara [3]. Cara kerja MFCC didasarkan pada perbedaan frekuensi yang dapat ditangkap oleh telinga manusia sehingga mampu merepresentasikan sinyal suara sebagaimana manusia merepresentasikannya, terdapat lima tahapan dalam MFCC seperti dapat dilihat pada Blok diagram proses MFCC pada gambar berikut :



Gambar 1. Blok Diagram MFCC

2.1 Tahapan Metode MFCC

Berikut ini adalah tahapan metode MFCC diantaranya yaitu :

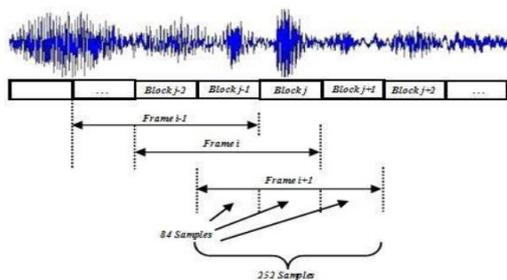
1) Preemphasis

Proses pengolahan sinyal wicara pre emphasis filter diperlukan setelah proses sampling. Tujuan dari pemfilteran ini adalah untuk mendapatkan bentuk spectral frekuensi sinyal wicara yang lebih halus. Filter pre-emphasis didasari oleh hubungan input/output dalam domain waktu yang dinyatakan dalam persamaan berikut :

$$y(n) = x(n) - \alpha x(n - 1) \quad \dots (1)$$

2) Frame Blocking

Pada proses ini, sinyal suara disegmentasi menjadi beberapa frame yang saling tumpang tindih. Proses ini dilakukan agar tidak ada sebagian sinyal yang hilang. Proses ini akan berlanjut sampai seluruh sinyal sudah masuk ke dalam satu atau lebih frame ini.



Gambar 2. Ilustrasi Frame Blocking Pada Sinyal Suara

3) Windowing

Proses windowing bertujuan untuk meminimalisasi ketidakberlanjutan sinyal pada awal dan akhir setiap frame. Jika kita definisikan window sebagai $w(n)$, $0 \leq n \leq N - 1$, dimana N adalah jumlah sampel pada setiap frame-nya, maka hasil dari *windowing* adalah sinyal:

$$y_1(n) = x_1(n) w(n), 0 \leq n \leq N - 1 \quad \dots (2)$$

4) Fast Fourier Transform (FFT)

Proses FFT merupakan proses yang berguna untuk mengonversi setiap *frame* dengan N sampel dari domain waktu menjadi domain frekuensi, sebagaimana didefinisikan sebagai berikut :

$$X_n = \sum_{k=0}^{N-1} x_k e^{-\frac{2\pi jkn}{N}} \quad \dots (3)$$

dimana $n = 0, 1, 2, \dots, N - 1$ dan $j = \sqrt{-1}$. Hasil dari tahapan ini biasanya disebut dengan *spectrum* atau *periodogram*.

5) Mel-Frequency Wrapping

Persepsi sistem pendengaran manusia terhadap frekuensi sinyal suara tidak dapat diukur dalam skala linear. Untuk setiap nada dengan frekuensi aktual, f , diukur dalam Hz, sebuah *subjective pitch* diukur dalam sebuah skala yang disebut "mel". Skala *mel-frequency* ialah sebuah frekuensi rendah yang bersifat linear di bawah 1000 Hz dan sebuah frekuensi tinggi yang bersifat logaritmik di atas 1000 Hz. Persamaan berikut menunjukkan hubungan skala mel dengan frekuensi dalam Hz:

$$F_{mel} = \begin{cases} 2595 \times \log_{10} \left(1 + \frac{F_{HZ}}{700} \right) & F_{HZ} > 1000 \\ F_{HZ} & F_{HZ} < 1000 \end{cases}$$

6) Cepstrum

Mel-Frequency Cepstrum kemudian didapat dari DCT (*Discrete Cosine Transform*) untuk mendapatkan kembali sinyal dalam domain waktu. Hasilnya disebut sebagai Mel-Frequency Cepstral Coefficient (MFCC). MFCC bisa didapat dari pendekatan persamaan :

$$C_i = \sqrt{\frac{2}{N} \sum_{j=1}^p m_j \cos \left(\frac{\pi}{N} (j - 0.5) \right)} \quad \dots (4)$$

Di mana adalah hasil akumulasi dari *kuadrat magnitude* DFT yang dikalikan dengan *Mel-Filter Bank*. Setelah itu didapatlah MFCC. Pada sistem pengenalan suara, biasanya hanya 13 cepstrum koefisien pertama yang digunakan.

7) Mel Scale

Mel Scale / skala mel berhubungan dirasakan frekuensi, atau lapangan, dari nada murni untuk frekuensi yang diukur sebenarnya. Manusia jauh lebih baik di perubahan kecil yang cerdas di lapangan pada frekuensi rendah daripada mereka pada frekuensi tinggi. Menggabungkan skala ini membuat fitur kami cocok lebih dekat apa yang manusia mendengar. Rumus untuk mengkonversi dari frekuensi ke skala Mel adalah:

$$M(f) = 1125 \ln(1 + f/700)$$

Untuk kembali dari Mel ke frekuensi :

$$M^{-1}(m) = 700 \left(\exp \left(\frac{m}{1125} \right) - 1 \right)$$

2.2 DTW (Dynamic Time Warping)

Menurut Candra D 2017 metode DTW adalah sebuah algoritma yang menghitung jalur pembengkokan optimal antara dua deretan waktu. Algoritma menghitung kedua nilai jalur pembengkokan antara dua deretan dan jarak di antara mereka [4]. Algoritma ini dimulai dengan perhitungan jarak lokal antara elemen dari dua

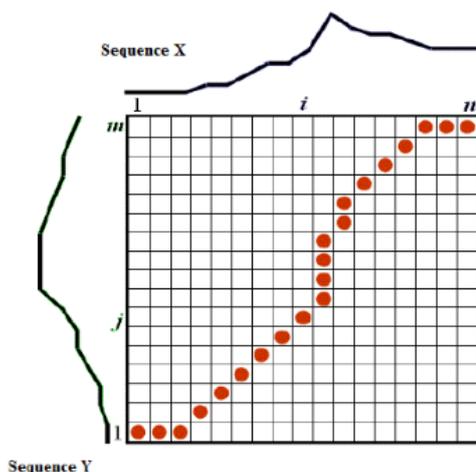
urutan menggunakan berbagai jenis jarak. Metode yang paling sering digunakan untuk perhitungan jarak adalah jarak absolut antara nilai-nilai dari dua elemen (jarak euclidean)

Dua ucapan dari kata yang sama oleh pengguna yang sama dapat memiliki waktu yang berbeda. Sebagai contoh, two dapat dilafalkan dengan to atau too. DTW menyelesaikan masalah ini dengan menyelaraskan kata-kata dengan benar dan menghitung jarak minimum antara dua kata. Sebuah matriks jarak lokal dibentuk untuk semua segmen dalam kata sampel dan template kata.

Waktu keselarasan ucapan yang berbeda adalah masalah inti untuk pengukuran jarak dalam pengenalan ucapan. Pergeseran kecil mengakibatkan identifikasi yang salah. Dynamic Time Warping adalah metode yang efisien untuk memecahkan masalah keselarasan waktu. Algoritma DTW ditujukan untuk menyelaraskan dua sekuen vector dengan membelokkan sumbu waktu berulang-ulang sampai kecocokan optimal antara dua sekuen ditemukan. algoritma ini melakukan sebagian pemetaan linear dari sumbu axis untuk menyelaraskan kedua sinyal.

Anggap saja terdapat dua urutan vector di ruang n-dimensi [4]:

$x = [x_1, x_2, \dots, x_n]$ dan $y = [y_1, y_2, \dots, y_n]$ Dua urutan tersebut selaras di sisi kotak, dengan satu di atas dan lainnya di sisi kiri. Kedua urutan mulai di bagian bawah kiri grid.



Gambar 3. Jarak Grid Global

ditempatkan, membandingkan unsur-unsur yang sesuai dari dua sekuens. Jarak antara dua titik dihitung melalui jarak Euclidean.

$$\text{Dist}(x,y) = \sqrt{|x-y|} = \sqrt{[(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2]^{1/2}}$$

Pencocokan terbaik atau keselarasan antara dua sekuens ini adalah jalan melalui grid, yang meminimalkan total jarak antara mereka, yang disebut sebagai jarak global. Keseluruhan jarak (jarak Global) dihitung dengan menemukan dan pergi melalui semua rute yang mungkin melalui grid, masing-masing menghitung jarak keseluruhan.

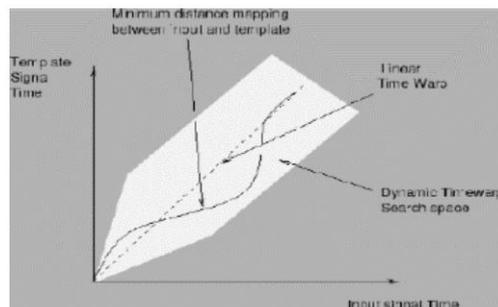
Jarak global minimum dari jumlah jarak (jarak Euclidean) antara unsur-unsur individual di jalan dibagi dengan jumlah dari fungsi pembobotan. Untuk setiap urutan cukup panjang jumlah kemungkinan jalan melalui grid akan sangat besar. mengukur jarak global diperoleh dengan menggunakan rumus rekursif.

$$GD_{xy} = LD_{xy} + \min(GD_{x-y-1}, GD_{x-1-y}, GD_{x-y-1})$$

Dimana:

GD = Global Distance (overall distance)

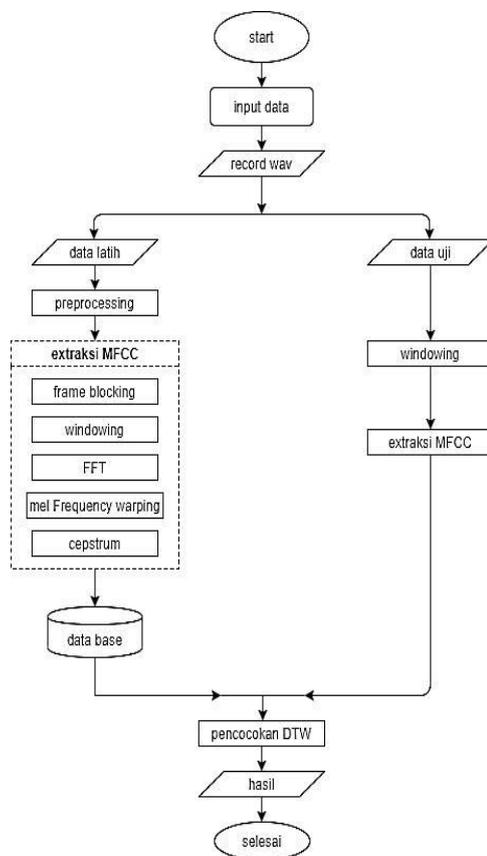
LD = Local Distance (Euclidean distance)



Gambar 4. Dynamic Time Warping

3. HASIL DAN PEMBAHASAN

3.1 Rancangan Sistem



Gambar 5. Rancangan Sistem

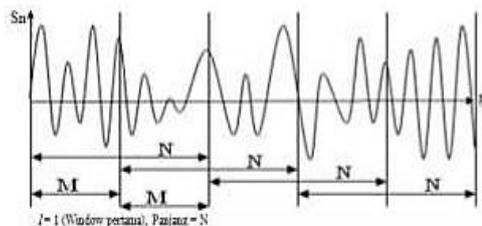
Berikut adalah rancangan sistem untuk identifikasi jenis suara burung. Pada rancangan sistem diatas terdiri dari proses data suara latih dan proses suara uji.

1) Proses Data Latih

Pada proses ini suara yang dijadikan sebagai suara latih didapatkan dari proses perekaman atau suara tersebut dipilih dari direktori. Pada saat input suara, suara masih berupa sinyal analog dimana sistem tidak dapat mengidentifikasinya, maka perlu dilakukan konversi suara menjadi digital dalam bentuk format (.wav). Kemudian dilakukan Proses ekstraksi ciri sebagai berikut :

a. Frame Blocking

Frame Blocking adalah pembagian suara menjadi beberapa frame yang nantinya dapat memudahkan dalam perhitungan dan analisa suara, satu frame terdiri dari beberapa sampel tergantung tiap berapa detik suara akan disampel dan berapa besar frekuensi samplingnya.



Gambar 6. Proses Frame Blocking

b. Windowing

Proses setiap frame untuk meminimalisir diskontinuitas sinyal pada permulaan dan akhir setiap frame. Konsepnya adalah meruncingkan sinyal ke angka nol pada permulaan dan akhir setiap frame, maka hasil dari proses ini adalah sinyal :

$$y(n) = x(n)w(n), 0 \leq n \leq N - 1$$

Dengan \$y(n)\$ = sinyal hasil windowing sampel ke-\$n\$

\$x(n)\$ = nilai sampel ke-\$n\$

\$w(n)\$ = nilai window ke-\$n\$

\$N\$ = jumlah sampel dalam frame

Secara khusus (dalam masalah ini), secara empiris, digunakan hamming window, yang mempunyai bentuk,

$$w(n) = 0,54 \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1$$

c. Fast Fourier Transform (FFT)

Proses selanjutnya sebuah algoritma cepat untuk implementasi Discrete Fourier Transform (DFT) yang dioperasikan pada

sebuah sinyal waktu-diskret yang terdiri dari \$N\$ sampel sebagai berikut :

$$f(n) = \sum_{k=0}^{N-1} y_k e^{-2\pi jkn/N}, n = 0,1,2, \dots, N-1$$

d. Mel Frequency Warping

Selanjutnya sebuah pendekatan untuk simulasi spektrum dalam skala mel dengan menggunakan filter bank yang diletakkan secara seragam dalam skala mel. Skala ini didefinisikan oleh Stanley Smith, John Volkman dan Edwin Newman sebagai :

$$mel(f) = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right)$$

e. Cepstrum

Cepstrum biasa digunakan untuk mendapatkan informasi dari suatu sinyal suara yang diucapkan oleh manusia. Bila mel power spectrum coefficients dinotasikan sebagai \$S_k, k = 1,2, \dots, K\$, Min \$N\$. Do mendefinisikan koefisien dari MFCC (\$c_n\$) sebagai :

$$c_n = \sum_{k=1}^K (\log S_k) \cos\left[n\left(k - \frac{\pi}{K}\right)\right], n = 1,2, \dots, K$$

Hasil dari proses ekstraksi MFCC di atas didapatkan suatu vektor ciri yang disimpan di database. Selanjutnya database suara latih akan digunakan sebagai bahan untuk pencocokan data uji dan data latih dalam proses pengenalan jenis suara burung.

2) Rencana Pengujian

Pada rencana pengujian suara yang dijadikan sebagai data uji didapatkan dari proses perekaman atau suara tersebut dipilih dari direktori, kemudian data suara tersebut dijadikan bentuk format (.wav). Selanjutnya dilakukan proses ekstraksi ciri menggunakan metode MFCC untuk mendapatkan vector ciri yang akan dicocokkan dengan database. Pencocokan digunakan dengan menggunakan metode DTW yaitu vektor uji dicocokkan dengan vektor data suara latih dan didapatkan nilai minimum yang menunjukkan jenis suara dari jenis burung berkicau.

4. SIMPULAN

Dari hasil rancangan sistem identifikasi suara burung menggunakan metode Mel Frequency Cepstrum Coefficients dan Dynamic Time Warping sebagai pencocokan pola berhasil sesuai rencana.

5. HARAPAN

Harapan besar penulis adalah dapat mewujudkan pembuatan sistem identifikasi jenis

burung dapat dikembangkan lagi untuk penambahan suara jenis burung ke dalam database. Dan ditambah dengan metode yang lain agar hasil menjadi lebih akurat dan pengguna menjadi lebih yakin serta mendapatkan informasi yang lebih baik.

DAFTAR PUSTAKA

- [1] Annisa, R 2012. *Klasifikasi jenis burung berdasarkan kicau burung menggunakan wavelet packet decomposition dan jaringan syaraf tiruan self organizing map*. Universitas Telkom.
- [2] Destian, Tri Handoko., Patmi Kasih, *Voice Recognition untuk Sistem Keamanan PC Menggunakan Metode MFCC dan DTW*, Generation Journal, Volume 2, No.1. Januari 2018
- [3] Sanjaya, M. *Implementasi pengenalan pola suara menggunakan Mel Frequency Cepstrum Coefficients dan adaptif neuro-fuzzy inference system (anfis) sebagai kontrol lampu otomatis*, Volume 1, No.1, Desember 2014
- [4] Candra, D. 2017 *Implementasi Teknik Dynamic Time Warping (DTW) pada Aplikasi Speech to Text*, Volume 10, No. 1, 2017.
- [5] Syafria, F. Dkk. *Pengenalan Suara Paru-Paru dengan MFCC sebagai Ekstraksi Ciri dan Backpropagation sebagai Classifier*, *Jurnal Ilmu Komputer Agri-Informatika*, Volume 3 Nomor 1 halaman 28 – 37
- [6] Mustofa, A. 2007. *Sistem Pengenalan Penutur dengan Metode Mel-Frequency Wrapping*. *Jurnal Teknik Elektro*. Vol. 7(2): 88-96.
- [7] Putera, L. E. S. & Adi, C. K. *Klasifikasi Burung Berdasarkan Suara Kicau Burung Menggunakan Jaringan Syaraf Tiruan Propagasi Balik*, *Seminar Nasional Ilmu Komputer (SNIK 2016)* - Semarang, 10 Oktober 2016 ISBN: 978-602-1034-40-8.
- [8] Setyawan A. 2011. *Aplikasi Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients (MFCC) Melalui Jaringan Syaraf Tiruan (JST), Learning Vector Quantization (LVQ) untuk Mengoperasikan Cursor Komputer*.