

Pengembangan Model Sistem Identifikasi Pembicara dengan Kombinasi Teknik Ekstraksi Ciri Suara Mel-Frequency Cepstral Coefficients (MFCC) dan Principal Component Analys (PCA)

Titin Pramiyati, Jayanta

Jurusan Sistem Informasi, Jurusan Teknik Informatika

UPNVJ, Jakarta

Email: antasena30@yahoo.com, titin.harsono@gmail.com

ABSTAK

Pesatnya kemajuan teknologi informasi, memberi kemudahan dalam mengembangkan sistem cerdas berbasis data biometrik (data yang melekat pada diri seseorang), seperti sistem identifikasi berbasis suara, sidik jari, pola tangan, retina dan iris mata. Keunggulan penggunaan data biometrik karena sifat tidak dapat dihilangkan, dilupakan, atau dipindahkan dari satu orang ke orang lain. Pengembangan sistem pengamanan yang berbasis data biometrik menjadi pilihan, baik untuk pengamanan akses pada layanan sistem informasi dan transaksi bisnis. Pengembangan Sistem Identifikasi Pembicara (SIP) yang akurat dan cepat, sangat dibutuhkan terlebih pada transaksi bisnis. Untuk mencapai tingkat akurasi dan kecepatan pada SIP, maka pada proses ekstraksi ciri suara, digunakan kombinasi teknik ekstraksi ciri MFCC dan PCA, yang dilakukan secara serial. Hasil percobaan menunjukkan kombinasi teknik ekstraksi ciri ini dapat mereduksi data ciri sehingga kecepatan identifikasi dan tingkat akurasi menunjukkan hasil yang baik.

Kata kunci: Sistem Identifikasi Pembicara, MFCC, PCA

1. PENDAHULUAN

Data biometrik seperti suara merupakan data yang melekat pada diri seseorang, yang dapat dijadikan data masukan sebuah sistem identifikasi pembicara menggunakan informasi spesifik yang terdapat dalam gelombang suara untuk mengenali identitas seseorang, teknik ini dapat diterapkan untuk mengidentifikasi dan memverifikasi identitas seseorang ketika akan mengakses suatu jasa layanan yang dikembangkan menggunakan gelombang suara yang disampaikan melalui jaringan telepon [2].

Kemampuan SIP dalam mengidentifikasi seseorang bergantung pada keberhasilan mengidentifikasi dengan benar dan cepat, terlebih jika SIP digunakan pada transaksi bisnis yang membutuhkan otentifikasi dalam proses transaksi. Untuk memenuhi kemampuan SIP baik pada tingkat akurasi maupun kecepatan. Untuk mencapai kemampuan SIP, maka data masukan yang digunakan pada SIP menjadi hal penting, oleh karenanya pembentukan data masukan dengan ukuran kecil dan tetap membawa informasi spesifik suara yang baik perlu dilakukan sehingga SIP tetap dapat melakukan identifikasi seseorang melalui suara dengan akurat dan cepat.

Klasifikasi pola suara merupakan dasar kerja SIP, oleh karenanya sistem menggunakan Jaringan Syaraf Tiruan Probabilistik atau yang lebih dikenal sebagai PNN, yang dapat membentuk model referensi dari proses pembelajaran berdasarkan data pelatihan yang

ditetapkan, sehingga PNN memiliki kemampuan untuk melakukan pencocokan pola [7] dan melakukan klasifikasi data yang belum diberikan pada saat pembelajaran sebelumnya [8].

Data yang digunakan pada SIP baik untuk data pelatihan atau data pengujian merupakan data ciri hasil pengolahan gelombang suara pembicara, proses pengolahan gelombang suara menjadi data ciri disebut ekstraksi ciri. Metode yang banyak digunakan pada ekstraksi ciri adalah MFCC dan PCA. Kedua metoda memiliki keunggulan masing-masing, keunggulan MFCC pada pencapaian tingkat akurasi identifikasi. Penelitian dengan menggunakan MFCC dengan nilai parameter koefisien *mel* 12, lebar waktu *frame* 20 ms, *overlap* 50% dari lebar *frame*, memberikan nilai akurasi 94,1% [9].

Penelitian pengenalan suara (*speaker verification*) melalui jaringan telepon tetap, yang juga menggunakan MFCC dengan nilai parameter koefisien *mel* 20, lebar waktu *frame* 30 ms, *overlap* 50% dari lebar *frame*, memberikan nilai akurasi pengujian sebesar 97,43%, [3]. Penelitian identifikasi pembicara (*speaker identification*), menggunakan MFCC dengan nilai parameter koefisien *mel* 20, lebar waktu *frame* 40 ms, *overlap* 50% dari lebar waktu *frame*, hasil penelitian memberikan nilai akurasi 96% [4].

Metoda PCA memiliki keunggulan dalam mereduksi dimensi data ciri, *Bulent Bolat* dan *Tulay Yildirim*, melakukan penelitian menggunakan kombinasi algoritma replikasi data dan PCA secara serial untuk

mendapatkan data ciri, penerapan PCA terhadap hasil replikasi data dapat mereduksi ukuran data ciri hingga 30% dengan hasil penelitian memperlihatkan bahwa penerapan metode PCA tetap memberikan nilai akurasi identifikasi sebesar 98,6%. [1].

Dengan melihat pada keunggulan kedua metoda dan kebutuhan akan sistem identifikasi pembicara yang memberikan tingkat akurasi dan kecepatan identifikasi yang baik, maka penggunaan kedua metoda dalam proses ekstraksi ciri perlu dilakukan.

Hasil yang diperoleh dari pengembangan SIP yang menggunakan data ciri yang diperoleh dari kombinasi teknik ekstraksi ciri MFCC dan PCA, menggunakan pendekatan *text-dependen* dengan teks “*Sembilan*” sebagai kata kunci yang diucapkan, dengan nilai parameter yang digunakan pada metode MFCC, adalah koefisien *mel* 20; lebar waktu *frame* 40 ms; lebar *overlap* 50% dari lebar waktu *frame*; dan sistem dilakukan secara *off-line*, kemudian data ciri yang diperoleh di reduksi dengan PCA, hasil ekstraksi ciri dapat mereduksi dimensi vektor suara hingga lebih dari 90%, dan akurasi identifikasi pembicara pada prototipe sistem adalah sebesar 96%.

Makalah ini juga akan membahas latar belakang dan konsep yang digunakan dalam sistem identifikasi pembicara yang telah ada, kemudian akan dibahas proses Penggunaan metode yang dipilih serta hasil yang diperoleh, dan di bagian akhir berupa kesimpulan dan saran.

2. METODA dan DATA CIRI

2.1 Kombinasi MFCC dan PCA pada Ekstraksi Ciri Suara

Kombinasi teknik MFCC dan PCA pada proses ekstraksi ciri dilakukan secara serial, dimana teknik MFCC dilakukan terlebih dahulu untuk mendapatkan data ciri suara. Gelombang suara yang dipakai direkam pada frekuensi 16000 Hz dan kuantisasi amplitudo 16 bit, menggunakan alat bantu mikrofon standar PC dan komputer personal yang telah diprogram untuk dapat melakukan poses rekam suara. Sumber suara berasal dari 10 pembicara dewasa dengan rentang usia 21 tahun hingga 51 tahun, pembicara terdiri atas 5 orang wanita dan 5 orang pria. Setiap pembicara diminta mengucapkan kata sandi “*Sembilan*” sebanyak 10 kali dengan durasi rekam 2 (dua) detik untuk satu kali pengucapan kata sandi.

MFCC akan memecah setiap gelombang suara menjadi beberapa *frame*, kemudian menentukan nilai *overlap* antar *frame*. Nilai lebar *frame* yang digunakan adalah 40 ms dengan *overlap* 20 ms atau 50% dari lebar *frame*. Untuk mendapatkan jumlah *frame* yang dihasilkan, dihitung terlebih dahulu jumlah data pada setiap *frame* dengan menggunakan rumus;

$$Fr/1000 * Fs \quad (1)$$

Fr adalah lebar waktu *frame*, dan *Fs* adalah rentang frekuensi suara dalam 1 (satu) detik, maka akan didapat $(40/1000 * 16000) = 640$ sampel data untuk setiap *frame*.

Kemudian untuk mendapat jumlah *frame* digunakan rumus;

$$K = ((L - N)/(N - D)) + 1 \quad (2)$$

K adalah jumlah *frame* yang dihasilkan, *L* rentang frekuensi suara, *N* jumlah data dalam 1 (satu) *frame*, dan *D* adalah titik awal terjadinya *overlapping* dalam *frame*, maka akan dihasilkan $((16000 - 640)/(640 - 320)) + 1 = 49$ *frame*.

Setelah *frame* berhasil dibentuk kemudian proses dilanjutkan dengan pembobotan terhadap setiap *frame*, proses pembobotan dilakukan dengan *window hamming*. Proses dilanjutkan untuk mendapatkan data ciri, koefisien *mel* yang digunakan pada pembentukan data ciri ini adalah 20 [3][4].

Proses ekstraksi ciri dengan MFCC di atas menghasilkan vektor data ciri dengan ukuran 1.881 ciri, ukuran ini masih terlalu besar untuk dipakai sebagai data masukan. Untuk memperkecil ukuran vektor data ciri ini, digunakan teknik lain yaitu PCA.

PCA (*Principal Component Analysis*) atau yang dikenal sebagai metode *Karhunen-Loeve*, merupakan salah satu teknik analisis peubah ganda yang sering digunakan untuk mereduksi dimensi data tanpa harus kehilangan nilai informasi berarti. Peubah hasil transformasi PCA merupakan kombinasi linier dari peubah asli, tidak berkorelasi antar sesama, tersusun berdasarkan informasi yang dikandungnya disebut sebagai vektor *eigen* atau nilai komponen utama.

Untuk mengetahui berapa banyak komponen utama dan seberapa efektif dimensi data dapat dijadikan data sampel, maka dilakukan perhitungan proporsi nilai *eigen*, yaitu membagi jumlah *r* nilai *eigen* dengan jumlah seluruh nilai *eigen*. Hasil perhitungan proporsi nilai *eigen* disebut rasio yang diekspresikan sebagai persentasi, dengan rumusan [6]:

$$Rasio = \frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^q \lambda_i} \quad (3)$$

Jika persentasi rasio 80%, dan 90% dari total nilai *eigen*, akan memberikan sebanyak *r* komponen utama untuk menggantikan data asli tanpa banyak kehilangan informasi [5].

Dengan menggunakan PCA, ukuran vektor data ciri yang semula berukuran 1.881 direduksi menjadi vektor data ciri berisi 99 nilai komponen utama, dengan persentasi rasio yang dipakai sebesar 95%, maka didapat sebanyak 36 nilai *eigen* mulai dari nilai *eigen* ke-1 sampai dengan nilai *eigen* ke-36 untuk membentuk vektor data ciri.

2.2 Pengelompokan Data Ciri

Data ciri yang telah dihasilkan disebar kedalam 2 kelompok data, yaitu: kelompok data pelatihan dan kelompok data pengujian. Penentuan komposisi dan jumlah data ciri yang akan dijadikan anggota kelompok data pelatihan dan kelompok data pengujian.

Pengelompokan data ciri didasarkan pada metode validasi yang akan dipakai, pada tulisan ini digunakan metode *leave-one-out*, sehingga pengelompokan data ciri dibagi menjadi 10 kelompok. Setiap kelompok data ciri terdiri dari 9 vektor ciri sebagai data pelatihan, dan 1 vektor ciri sebagai data pengujian, hasil pengelompokan data ciri adalah sebagai berikut:

Tabel. 1. Kelompok Data Pelatihan dan Pengujian

Pengujian ke.	Komposisi dan jumlah data ciri	
	Kelompok Data Pelatihan	Kelompok Data Pengujian
1	Vc1, Vc2, Vc3, Vc4, Vc5, Vc6, Vc7, Vc8, Vc9	Vc10
2	Vc1, Vc2, Vc3, Vc4, Vc5, Vc6, Vc7, Vc8, Vc10	Vc9
3	Vc1, Vc2, Vc3, Vc4, Vc5, Vc6, Vc7, Vc9, Vc10	Vc8
4	Vc1, Vc2, Vc3, Vc4, Vc5, Vc6, Vc8, Vc9, Vc10	Vc7
5	Vc1, Vc2, Vc3, Vc4, Vc5, Vc7, Vc8, Vc9, Vc10	Vc6
6	Vc1, Vc2, Vc3, Vc4, Vc6, Vc7, Vc8, Vc9, Vc10	Vc5
7	Vc1, Vc2, Vc3, Vc5, Vc6, Vc7, Vc8, Vc9, Vc10	Vc4
8	Vc1, Vc2, Vc4, Vc5, Vc6, Vc7, Vc8, Vc9, Vc10	Vc3
9	Vc1, Vc3, Vc4, Vc5, Vc6, Vc7, Vc8, Vc9, Vc10	Vc2
10	Vc2, Vc3, Vc4, Vc5, Vc6, Vc7, Vc8, Vc9, Vc10	Vc1

Ket.: Vc singkatan dari Vektor ciri

3. HASIL PENGUJIAN SISTEM

Hasil pengujian sistem memperlihatkan bahwa keberhasilan identifikasi dari 10 kali pengujian setelah di generalisasi memberikan nilai akurasi tertinggi sebesar 96%, dengan data ciri yang telah direduksi hingga 98%. Dengan hasil yang demikian memperlihatkan penggunaan kombinasi teknik ekstraksi ciri MFCC dan PCA tetap memiliki tingkat akurasi yang baik.

Penggunaan MFCC dan PCA dalam proses ekstraksi ciri, memperlihatkan kecepatan identifikasi yang diperoleh baik, yaitu 0.85 detik.

4. KESIMPULAN dan SARAN

4.1 Kesimpulan

Kesimpulan dari pembahasan ini adalah, penggunaan kombinasi metode ekstraksi ciri MFCC dan PCA pada rasio nilai *eigen* 95 dapat mereduksi dimensi data hingga 98% lebih kecil dari dimensi data asli. Hasil reduksi ini memberikan kecepatan proses identifikasi yang baik.

Vektor data ciri yang telah mengalami reduksi, jika digunakan pada koefisien *mel* 20, lebar *frame* 40 ms, dan *overlap* 50%, memberikan tingkat akurasi hingga 95%, nilai tersebut lebih rendah 1% dibandingkan dengan nilai akurasi dari pengujian sistem dengan data ciri hasil metode MFCC dengan parameter yang sama.

4.2 Saran

Penelitian lanjutan terhadap penggunaan kombinasi teknik ekstraksi ciri MFCC dan PCA dapat dilakukan pada kelompok data ciri dengan koefisien *mel*, lebar *frame* yang berbeda, yang bertujuan untuk membandingkan tingkat akurasi dan kecepatan proses identifikasi. Dengan penelitian lanjutan ini, sistem identifikasi pembicara dapat diterapkan pada situasi nyata.

DAFTAR PUSTAKA

- [1] Bolat B, Yildirim T., "Performance increasing methods for probabilistic Neural Networks", *Pakistan Journal of Information and Technology*, vol 2 no 3, 2003, p. 250-255.
- [2] Furui S., "Recent advances in speaker recognition", *Pattern Recognition Letters* vol 18", 1997, p. 859 – 872.
- [3] Ganchev T, Tsopanoglou A, Fakotakis N, Kokkinakis G., "Probabilistic Neural Networks Combined with GMMs for Speaker Recognition over Telephone Channels", *14th Int. Conf. On Digital Signal Processing* vol 2, 2002, p:1081-1084.

- [4] Jayanta, “Pengembangan Model Jaringan Syaraf Tiruan Probabilistik (PNN) pada Identifikasi Pembicara”, 2007,
- [5] Johnson R. A, Wichern R.E., *Applied Multivariate Statistical Analysis*, Printice Hall, Inc., New Jersey, USA, 1998.
- [6] Kantardzic, M., *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons, Inc., New Jersey, USA, 2003.
- [7] Kusumadewi S., *Artificial Intelligence (Teknik dan Aplikasinya)*., Graha Ilmu, Yogyakarta, 2003.
- [8] Li Min Fu., *Neural Networks in Computer Intelligence*, The McGraw-Hill, Inc., New York, 1994.
- [9] Low R, Togneri R., *Speech Recognition Using the Probabilistic Neural Network*., <http://www.ee.uwa.edu/~roberto/research/speech/local/papers/-tr98-01.pdf>