

PENGENALAN NADA ALAT MUSIK MENGGUNAKAN EKSTRAKSI CIRI PERATAAN SEGMENT BERBASIS DST, DAN PENGLASIFIKASI SVM

Linggo Sumarno

Program Studi Teknik Elektro, Universitas Sanata Dharma Yogyakarta
email: lingsum@usd.ac.id

ABSTRACT

*Musical instrument tone recognition by a computer is an attempt to make a computer can imitate the human ability in recognizing the tone of musical instruments. This paper proposes a method of feature extraction in a musical instrument tone recognition. In more detail, the proposed feature extraction method is a DST (Discrete Sine Transform) based segment averaging. The instruments used are pianica and bellyra. Pianica is a representation of a musical instrument that has polyphonic tones, while belira is a representation of a musical instrument that has monophonic tones. The classifier used in this tone recognition is SVM (Support Vector Machine). The experimental results showed that the proposed feature extraction method required only 8 coefficients of feature extraction in order to represent a tone, both polyphonic or monophonic tones. Then, the SVM classifier used only required a linear kernel function. The use of 8 coefficients of feature extraction and a linear kernel function had been able to give the highest recognition rate up to 100%.
Keywords: DST, segment averaging, SVM, tone recognition*

INTISARI

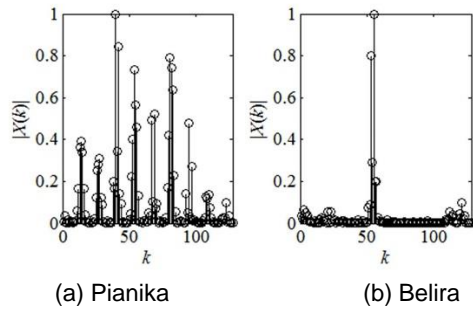
Pengenalan nada alat musik oleh komputer merupakan suatu upaya untuk membuat komputer dapat meniru kemampuan manusia dalam mengenali nada alat musik. Makalah ini mengusulkan suatu metode ekstraksi ciri dalam suatu pengenalan nada alat musik. Secara lebih detail, metode ekstraksi ciri yang diusulkan adalah perataan segmen berbasis DST (Discrete Sine Transform). Alat musik yang digunakan adalah pianika dan belira. Pianika merupakan representasi dari alat musik yang mempunyai nada polifonik, sedangkan belira merupakan representasi dari alat musik yang mempunyai nada monofonik. Pengklasifikasi yang digunakan dalam pengenalan nada ini adalah SVM (Support Vector Machine). Hasil percobaan memperlihatkan bahwa metode ekstraksi ciri yang diusulkan hanya memerlukan 8 koefisien ekstraksi ciri untuk merepresentasikan suatu nada, baik nada polifonik maupun monofonik. Kemudian, pengklasifikasi SVM yang digunakan hanya memerlukan fungsi kernel linear. Penggunaan 8 koefisien ekstraksi ciri dan fungsi kernel linear tersebut sudah dapat memberikan tingkat pengenalan tertinggi hingga 100%.
Kata kunci: DST, pengenalan nada, perataan segmen, SVM

1. PENDAHULUAN

Menurut Forster (2010), berdasarkan sudut pandang persepsi manusia, terdapat dua karakteristik aural dari nada-nada yang dikeluarkan alat musik. Dua karakteristik tersebut adalah skala (*scale*) dan warna nada (*timbre*). Karakteristik skala terkait dengan tinggi rendahnya nada, sedangkan karakteristik warna nada terkait dengan jenis alat musik. Pada dasarnya terdapat dua jenis warna nada, yaitu monofonik dan polifonik. Gambar 1 memperlihatkan dua jenis warna nada tersebut bila dilihat dalam ranah DST. Sebagaimana terlihat pada Gambar 1, nada belira merupakan salah satu contoh nada monofonik, karena hanya memperlihatkan satu puncak lokal signifikan. Sementara itu, nada pianika merupakan salah satu contoh nada polifonik, karena memperlihatkan lebih dari satu (*multiple*) puncak lokal signifikan.

Biasanya, seorang pemain musik yang terlatih dapat mengenali dengan mudah nada-nada yang terdengar dari suatu alat musik. Untuk dapat mengenali nada-nada tersebut, pada dasarnya orang tersebut mengambil karakteristik dari nada yang terdengar dan kemudian membandingkannya dengan sejumlah karakteristik dari nada-nada yang pernah didengarnya di masa yang lampau (McAdams, 2010). Dewasa ini, komputer dapat dikembangkan menjadi suatu sistem pengenalan nada, yang mampu menirukan kemampuan pengenalan nada dari pemain musik tersebut. Pada suatu sistem pengenalan nada, pada dasarnya terdapat dua bagian utama. Bagian yang pertama adalah bagian yang mengambil karakteristik dari nada, atau yang disebut juga bagian ekstraksi ciri. Bagian kedua adalah bagian yang mengklasifikasi hasil ekstraksi ciri dengan berdasarkan sejumlah ekstraksi ciri

yang sebelumnya sudah dilatihkan pada sistem pengenalan nada tersebut.



Gambar 1. Representasi warna nada untuk nada 'C' dalam ranah DST ternormalisir $|X(k)|$, untuk Pianika dan belira, dengan menggunakan frekuensi pencuplikan 5000 Hz dan DST 128 titik.

Suatu pendekatan untuk melaksanakan bagian ekstraksi ciri di atas, adalah dengan menggunakan ranah transformasi. FFT (*Fast Fourier Transform*) dan DCT (*Discrete Cosine Transform*) adalah dua contoh metode transformasi yang dapat dipakai untuk mengubah sinyal nada dari ranah waktu ke ranah transformasi. Pada dasarnya terdapat dua cara melaksanakan ekstraksi ciri dalam ranah transformasi. Cara ekstraksi ciri yang pertama adalah dengan berdasarkan sinyal-sinyal fundamental (Izzudin *et al*, 2005; Mitre *et al*, 2006; Gaffar *et al*, 2012). Selanjutnya, cara ekstraksi ciri yang kedua adalah dengan tidak berdasarkan sinyal-sinyal fundamental (Tjahyanto *et al*, 2013; Sumarno, 2016; Sumarno & Iswanjono, 2017).

Penelitian-penelitian sebelumnya tentang ekstraksi ciri dalam pengenalan nada, kebanyakan hanya terkait dengan nada-nada polifonik, seperti misalnya yang dilakukan oleh Tjahyanto *et al* (2013) dan Sumarno (2016). Masih sangat sedikit yang terkait dengan pengenalan nada untuk nada-nada polifonik dan juga monofonik. Penelitian sebelumnya oleh Sumarno & Iswanjono (2017), mengusulkan suatu metode ekstraksi ciri perataan segmen berbasis FFT, yang dapat digunakan baik untuk nada-nada polifonik dan juga monofonik. Metode ekstraksi ciri tersebut masih memerlukan sejumlah 16 koefisien ekstraksi ciri untuk merepresentasikan suatu nada. Terlihat masih ada peluang untuk lebih menurunkan lagi jumlah koefisien ekstraksi ciri tersebut.

SVM adalah suatu metode klasifikasi yang asalnya dari teori belajar statistis (Vapnik, 1995; Vapnik 1998). Pada awalnya SVM hanya digunakan untuk klasifikasi dua kelas. Dalam perkembangan selanjutnya, SVM dapat digunakan untuk klasifikasi multikelas. Nurdiyah & Muwahhid (2016) membandingkan metode klasifikasi KNN (*K-Nearest Neighbor*), yang merupakan salah satu metode dari *template matching*, dengan metode klasifikasi SVM. Hasil yang diperoleh dari perbandingan tersebut adalah, SVM memiliki kinerja yang lebih unggul daripada KNN, karena SVM dapat memberikan tingkat pengenalan hingga 93,2%, sedangkan KNN hanya 91,5%.

Tulisan ini mengusulkan suatu sistem pengenalan nada alat musik, yang metode ekstraksi cirinya menggunakan ranah transformasi DST, serta tidak berdasarkan sinyal-sinyal fundamental pada ranah transformasi tersebut. Selain itu, pengenalan nada alat musik tersebut pengklasifikasinya menggunakan metode SVM. Sebagai catatan, sistem pengenalan nada alat musik yang diusulkan diuji untuk nada-nada polifonik dan monofonik.

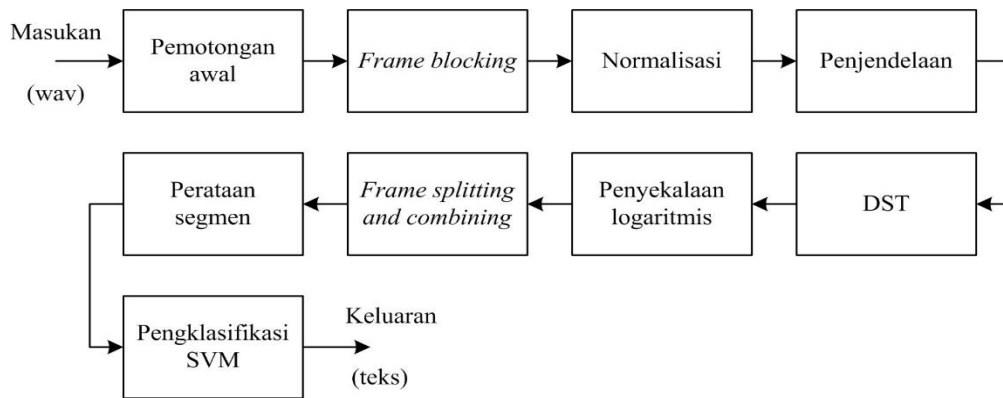
2. METODE PENELITIAN

2.1 Sistem Keseluruhan

Sistem keseluruhan dari pengenalan nada yang dikembangkan dalam penelitian ini diperlihatkan secara diagram blok dalam Gambar 2. Masukan dari sistem adalah suatu sinyal nada terisolasi dalam format wav. Keluaran dari sistem adalah suatu teks yang mengindikasikan nada yang dikenali. Berikut ini penjelasan masing-masing blok yang terdapat dalam Gambar 2.

2.1.1 Masukan

Masukan dari keseluruhan sistem pengenalan nada adalah suatu sinyal nada terisolasi dalam format wav. Sinyal nada ini didapatkan dari dua alat musik yang dimainkan yaitu pianika dan belira. Dalam penelitian ini, untuk setiap alat musik direkam delapan sinyal nada yaitu C, D, E, F, G, A, B, dan C'. Perekaman sinyal nada dilakukan dengan frekuensi pencuplikan 5000 Hz. Pada dasarnya frekuensi pencuplikan ini sudah memenuhi teorema sampling Shannon (Tan & Jiang, 2013) sebagai berikut.



Gambar 2. Diagram blok keseluruhan sistem pengenalan nada alat musik yang dikembangkan.

$$f_s \geq 2f_{\max} \quad (1)$$

dengan f_{\max} adalah komponen frekuensi tertinggi dari nada, dan f_s adalah frekuensi pencuplikan. Berdasarkan observasi, komponen frekuensi tertinggi dari nada C' untuk pianika dan belira masing-masing adalah 1584 Hz dan 2097 Hz. Berdasarkan observasi pula, perekaman nada selama 2 detik sudah mencukupi, untuk mendapatkan daerah tunak (*steady state*) dari sinyal nada. Daerah tunak ini digunakan untuk keperluan *frame blocking*. Sebagai catatan, sinyal nada hasil perekaman dapat dibagi menjadi tiga daerah, yaitu daerah hening, transisi, dan tunak. Hanya pada daerah tunak terdapat informasi nada yang akurat.



(a) Pianika



(b) Belira

Gambar 3. Pianika dan belira yang digunakan dalam penelitian (Sumarno & Iswanjono, 2017).

Dalam penelitian ini, alat musik yang digunakan untuk membangkitkan sinyal nada adalah pianika Yamaha P-37D dan belira Isuzu ZBL-27 (lihat Gambar 3). Untuk menangkap sinyal nada tersebut digunakan mikrofon AKG Perception 120.

2.1.2 Pemotongan awal

Pemotongan awal adalah suatu proses pemotongan daerah hening (*silence*) dan daerah transisi dari suatu sinyal nada. Daerah hening perlu dipotong karena pada daerah ini tidak terdapat informasi nada. Selanjutnya, daerah transisi juga perlu dipotong, karena pada daerah ini informasi nadanya tidak akurat. Informasi nada yang akurat baru bisa diperoleh setelah sinyal nadanya berada pada keadaan tunak. Pada penelitian ini, berdasarkan hasil observasi, pertama kali daerah hening dipotong dengan menggunakan nilai ambang amplitudo $[0,5]$ dari nilai maksimum sinyal nada. Berawal dari bagian paling kiri dari sinyal, jika amplitudo sinyal kurang dari $[0,5]$ dari nilai maksimum sinyal nada, maka sinyal tersebut dipotong. Setelah pemotongan daerah hening, selanjutnya dilakukan pemotongan daerah transisi. Berdasarkan hasil observasi juga, daerah transisi dipotong selama 300 milidetik dari bagian kiri sinyal.

2.1.3 Frame blocking

Frame blocking adalah suatu proses pemotongan suatu *frame* sinyal, dari suatu sinyal masukan yang panjang (Meseguer, 2009). *Frame blocking* ini bertujuan untuk mengurangi banyaknya data sinyal yang akan diproses. Pada dasarnya, data sinyal yang lebih sedikit, akan makin mempercepat komputasi dari sistem pengenalan nada. Penelitian ini menggunakan panjang *frame blocking* 2^n , dengan n adalah bilangan bulat positif. Nantinya, panjang *frame blocking* ini akan dievaluasi untuk mencari panjang *frame blocking* terkecil yang menghasilkan tingkat pengenalan tertinggi.

2.1.4 Normalisasi

Normalisasi adalah proses pengaturan nilai maksimum suatu runtun data sinyal menjadi 1 atau -1. Normalisasi ini bertujuan

untuk mengeliminasi adanya perbedaan nilai maksimum dari sinyal-sinyal hasil *frame blocking*. Pada dasarnya, normalisasi ini akan makin memperjelas perbedaan diantara kelas-kelas sinyal yang berbeda.

2.1.5 Penjendelaan

Penjendelaan (*windowing*) adalah suatu proses untuk mengurangi diskontinuitas pada tepi-tepi sinyal (Meseguer, 2009). Diskontinuitas ini muncul akibat adanya pemotongan sinyal pada proses *frame blocking* di atas. Diskontinuitas yang terlalu besar akan memunculkan sinyal-sinyal tambahan yang disebut sinyal-sinyal harmonik pada sinyal hasil transformasi. Penelitian ini menggunakan jendela Hamming, yang merupakan suatu jendela yang secara luas digunakan dalam bidang pengolahan sinyal (Oppenheim, 1989). Secara matematis, jendela Hamming $h(n)$ dengan panjang jendela N , dirumuskan sebagai berikut.

$$h(n) = 0,54 - 0,46 \cos\left(2\pi \frac{n}{N-1}\right), \quad 0 \leq n \leq N-1 \quad (2)$$

Penelitian ini menggunakan panjang jendela 2^n , dengan n adalah bilangan bulat positif. Panjang jendela ini sama dengan panjang *frame blocking* di atas.

2.1.6 DST

DST merupakan suatu proses untuk mengubah sinyal dari ranah waktu ke ranah transformasi yang disebut sebagai ranah DST. Seperti halnya DCT, DST merupakan suatu jenis transformasi yang juga diturunkan dari DFT (*Discrete Fourier Transform*). DST dari suatu runtun $y(n)$ yang panjangnya N titik, secara matematis dirumuskan sebagai berikut.

$$Y(k) = \sqrt{\frac{2}{n+1}} \sum_{n=0}^{N-1} y(n) \sin\left[\frac{\pi(n+1)(k+1)}{N+1}\right] \quad (3)$$

dengan $0 \leq n \leq N-1$. Penelitian ini menggunakan panjang DST 2^n , dengan n adalah bilangan bulat positif. Panjang DST ini sama dengan panjang *frame blocking* dan jendela Hamming di atas. Selain hal tersebut, penelitian ini menerapkan perhitungan nilai absolut dari hasil DST, karena proses berikutnya penyekalaan logaritmis, tidak memungkinkan adanya perhitungan nilai negatif.

2.1.7 Penyekalaan logaritmis

104 Sumarno, Pengenalan Nada Alat Musik Menggunakan Ekstraksi Ciri Perataan Segmen Berbasis DST, dan Pengklasifikasi SVM

Penyekalaan logaritmis adalah suatu proses untuk mengurangi adanya perbedaan nilai-nilai puncak pada suatu vektor data. Pada dasarnya, hasil penyekalaan logaritmis memperlihatkan adanya jumlah puncak lokal maksimum yang semakin banyak. Berdasarkan penelitian-penelitian sebelumnya dari (Sumarno, 2016) serta (Sumarno & Iswanjono, 2017), ekstraksi ciri yang menggunakan perataan segmen memperlihatkan hasil yang lebih baik untuk runtun data yang mempunyai jumlah puncak lokal maksimum yang banyak. Penyekalaan logaritmis ini secara matematis diperlihatkan sebagai berikut.

$$Y_{out}(k) = \log(Y_{in}(k) + 1) \quad (4)$$

dengan $Y_{out}(k) = \{Y_{out}(0), Y_{out}(1), \dots, Y_{out}(N-1)\}$ dan $Y_{in}(k) = \{Y_{in}(0), Y_{in}(1), \dots, Y_{in}(N-1)\}$ masing-masing adalah runtun *frame* data keluaran dan runtun *frame* data masukan, serta N adalah panjang *frame*. Adanya penambahan '1' pada rumusan di atas adalah untuk menghindari hasil logaritma yang mendekati minus tak hingga, jika ada nilai pada runtun $X_{in}(k)$ yang mendekati nol.

2.1.8 Frame splitting and combining

Frame splitting and combining adalah suatu proses untuk mengurangi panjang *frame* data. Pada dasarnya, hasil *frame splitting and combining* memperlihatkan adanya distribusi data yang lebih rapat, dalam suatu *frame*. Secara algoritma, *frame splitting and combining* ini diperlihatkan sebagai berikut.

Algoritma *frame splitting and combining*

1. Misalkan pada suatu *frame* data masukan terdapat runtun $Y_{in}(k) = \{Y_{in}(0), Y_{in}(1), \dots, Y_{in}(N-1)\}$, dengan N adalah bilangan genap yang menyatakan panjang *frame* data masukan.
2. Bagi dua runtun tersebut menjadi dua runtun data:
3. $V_1(k) = \{Y_{in}(0), Y_{in}(1), \dots, Y_{in}((N/2)-1)\}$
dan
 $V_2(k) = \{Y_{in}(N/2), Y_{in}((N/2)+1), \dots, Y_{in}(N-1)\}$
4. Gabungkan runtun hasil pembagian $V_1(k)$ dan $V_2(k)$ tersebut, menjadi suatu *frame* data keluaran
 $Y_{out}(k) = \{Y_{out}(0), Y_{out}(1), \dots, Y_{out}((N/2)-1)\}$, sebagai berikut.

$$Y_{out}(k) = \frac{V_1(k) + V_2(k)}{2} \quad (5)$$

Sebagai catatan, proses *frame splitting and combining* akan menghasilkan *frame* keluaran yang panjangnya separuh dari *frame* masukannya. Dalam penelitian ini panjang *frame* masukan sama dengan panjang *frame blocking*.

2.1.9 Perataan segmen

Perataan segmen adalah suatu proses untuk mengurangi jumlah data dari suatu *frame* data sinyal. Pada dasarnya, *frame* data sinyal dengan jumlah data yang berkurang ini masih memperlihatkan bentuk dasar dari pola sinyal aslinya. Penelitian ini menggunakan perataan segmen yang diinspirasi dari Setiawan (2015). Secara algoritma, perataan segmen ini diperlihatkan sebagai berikut.

Algoritma perataan segmen

1. Misalkan suatu *frame* data masukan $Y_{in}(k) = \{Y_{in}(1), Y_{in}(2), \dots, Y_{in}(M-1)\}$ dengan $M = 2^p$ dan $p \geq 0$.
2. Carilah panjang segmen L dengan $L = 2^q$ untuk $0 \leq q \leq p$.
3. Bagilah runtun data $X(k)$ dengan menjadi sejumlah segmen sepanjang L . Hasil pembagian tersebut akan memberikan sejumlah Z segmen sebagai berikut.

$$Z = \frac{M}{L} \quad (6)$$

dan juga runtun data $S(u) = \{S(0), S(1), \dots, S(L-1)\}$ di setiap segmen.

4. Hitung *frame* data keluaran $Y_{out}(v) = \{Y_{out}(1), Y_{out}(2), \dots, Y_{out}(Z-1)\}$ dengan melakukan operasi rerata disetiap segmen sebagai berikut.

$$Y_{out}(v) = \frac{1}{L} \sum_{u=0}^{L-1} S_v(u) \quad (7)$$

Dalam penelitian ini, panjang segmen L di atas dievaluasi dengan nilai-nilai 1, 2, 4, ..., $2^{\log_2(M/2)}$, dengan M adalah panjang *frame* hasil proses *frame splitting and combining*. Panjang segmen ini dievaluasi untuk mendapatkan sejumlah Z hasil ekstraksi ciri yang terkecil, yang menghasilkan tingkat pengenalan tertinggi. Sebagai catatan, jika N adalah panjang *frame* data masukan pada proses *frame splitting and combining*, maka persamaan (6) dapat dituliskan lagi menjadi sebagai berikut.

$$Z = \frac{N}{2L} \quad (8)$$

2.1.10 Klasifikasi SVM

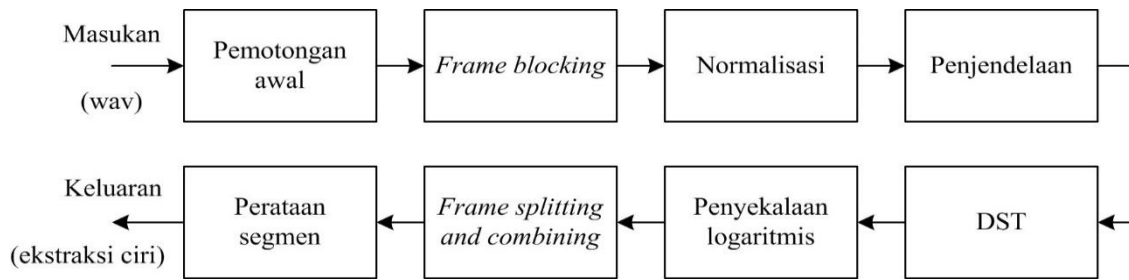
Klasifikasi adalah suatu proses untuk menentukan kelas pola dari data sinyal masukan. Salah satu metode untuk melakukan klasifikasi adalah SVM. Pada dasarnya SVM adalah pengklasifikasi linear. Dalam pelatihan SVM, akan dicari suatu *hyperplane* terbaik, yang mampu memisahkan dua set data, yang berasal dari dua kelas pola yang berbeda. Secara matematis, *hyperplane* tersebut merupakan suatu fungsi diskriminan linear.

Dalam kenyataan di dunia nyata, belum tentu dua set data dari dua kelas pola yang berbeda dapat dipisahkan dengan suatu *hyperplane*. Dengan kata lain, dua set data tersebut tidak *linearly separable*. Oleh karena itu, supaya dua set data tersebut *linearly separable* perlu dilakukan transformasi data. Untuk melakukan transformasi ini digunakan suatu fungsi yang disebut sebagai fungsi *kernel* (Boser *et al*, 1992). Fungsi *kernel* linear dan polinomial adalah dua contoh fungsi *kernel* yang umumnya digunakan. Penelitian yang dilaksanakan mengevaluasi kedua fungsi *kernel* tersebut.

Pada awalnya SVM dikembangkan untuk kasus klasifikasi dua kelas pola. Selanjutnya, SVM dikembangkan untuk kasus klasifikasi multikelas pola. Penelitian yang dilaksanakan merupakan kasus klasifikasi multikelas pola. Untuk kasus multikelas ini akan digunakan skema OVA (*one-vs-all*). Pemilihan OVA ini dilandasi bahwa, secara kinerja OVA ini pada dasarnya setara dengan skema-skema yang lainnya. Akan tetapi OVA ini menawarkan adanya kesederhanaan konseptual dan komputasi (Rifkin & Klatau, 2004).

2.2 Pelatihan Pengklasifikasi SVM

Pengklasifikasi SVM yang diperlihatkan pada Gambar 2 di atas, memerlukan sejumlah data dalam proses pelatihannya. Sejumlah data ini dibangkitkan menggunakan ekstraksi ciri nada yang diusulkan dalam penelitian ini, yang diperlihatkan dalam Gambar 4. Sebagaimana terlihat pada Gambar 4, masukannya adalah sinyal nada yang merupakan sinyal nada terisolasi dalam format wav, sedangkan keluarannya ekstraksi ciri dari sinyal nada masukannya.



Gambar 4. Diagram blok ekstraksi ciri nada yang diusulkan.

Pada penelitian yang dilaksanakan, untuk keperluan pelatihan pengklasifikasi SVM, dilakukan pengambilan 10 sampel untuk setiap sinyal nada (C, D, E, F, G, A, B, atau C'), untuk setiap alat musik (pianika atau belira). Dengan demikian secara keseluruhan terdapat 180 sinyal nada untuk keperluan pelatihan ini. Selanjutnya, dengan menggunakan model ekstraksi ciri yang diperlihatkan pada Gambar 4 di atas, masing-masing sinyal nada tersebut diproses ekstraksi cirinya. Hasil dari ekstraksi ciri ini kemudian digunakan untuk pelatihan pengklasifikasi SVM.

2.3 Nada uji

Nada uji merupakan sinyal nada yang digunakan untuk menguji kinerja dari sistem pengenalan nada yang dikembangkan. Pada penelitian yang dilaksanakan, untuk keperluan nada uji, dilakukan pengambilan 10 sampel yang lain untuk setiap sinyal nada (C, D, E, F, G, A, B, atau C'), untuk setiap alat musik (pianika atau belira). Dengan demikian secara keseluruhan terdapat 180 sinyal nada untuk keperluan nada uji ini.

3. HASIL DAN PEMBAHASAN

3.1 Hasil-Hasil Pengujian

Hasil-hasil pengujian untuk sistem pengenalan nada yang dikembangkan untuk penggunaan fungsi *kernel* linear, pada berbagai kombinasi panjang *frame blocking* dan panjang segmen, diperlihatkan pada Tabel 1 dan 2. Sebagai catatan yang pertama, jumlah koefisien ekstraksi ciri untuk setiap

kombinasi panjang *frame blocking* dan panjang segmen dapat dilihat pada Tabel 3. Persamaan (8) dipakai sebagai dasar untuk membuat Tabel 3 tersebut. Sebagai catatan yang kedua, panjang segmen mengacu pada panjang segmen yang digunakan pada proses perataan segmen.

Berdasarkan observasi Tabel 1, 2, dan 3 secara simultan, dapat adanya dilihat dua hal berikut. Hal yang pertama adalah, secara umum untuk panjang *frame blocking* makin besar, tingkat pengenalan akan makin tinggi (Tabel 1 dan 2) dan jumlah koefisien ekstraksi ciri makin besar (Tabel 3). Persamaan (8) dipakai sebagai dasar untuk membuat Tabel 3 tersebut. Sebagai catatan yang kedua, panjang segmen mengacu pada panjang segmen yang digunakan pada proses perataan segmen.

Berdasarkan observasi Tabel 1, 2, dan 3 secara simultan, dapat adanya dilihat dua hal berikut. Hal yang pertama adalah, secara umum untuk panjang *frame blocking* makin besar, tingkat pengenalan akan makin tinggi (Tabel 1 dan 2) dan jumlah koefisien ekstraksi ciri makin besar (Tabel 3). Selanjutnya hal yang kedua adalah, secara umum untuk panjang segmen makin kecil, tingkat pengenalan juga makin tinggi (Tabel 1 dan 2) serta jumlah koefisien ekstraksi ciri juga makin besar (Tabel 3). Dengan demikian dari dua hal tersebut dapat dikatakan bahwa, tingkat pengenalan yang makin tinggi disebabkan karena jumlah koefisien ekstraksi ciri makin besar.

Tabel 1. Hasil pengujian alat musik pianika, untuk penggunaan fungsi *kernel* linear, pada berbagai kombinasi panjang *frame blocking* dan panjang segmen.

Hasil yang diperlihatkan: Tingkat pengenalan (%).

Panjang <i>frame blocking</i> (titik)	Panjang segmen (titik)									
	1	2	4	8	16	32	64	128	256	512
64	100	100	100	88,75	70,00	48,75	-	-	-	-
128	100	100	100	98,75	93,75	62,50	36,25	-	-	-
256	100	100	100	100	100	82,50	65,00	36,25	-	-
512	100	100	100	100	100	100	76,25	55,00	35,00	-
1024	100	100	100	100	100	100	100	73,75	55,00	28,75

Tabel 2. Hasil pengujian alat musik belira, untuk penggunaan fungsi *kernel* linear, pada berbagai kombinasi panjang *frame blocking* dan panjang segmen. Hasil yang diperlihatkan: Tingkat pengenalan (%).

Panjang <i>frame blocking</i> (titik)	Panjang segmen (titik)									
	1	2	4	8	16	32	64	128	256	512
64	100	100	98,75	92,50	55,00	22,50	-	-	-	-
128	100	100	100	100	81,25	55,00	22,50	-	-	-
256	100	100	100	100	98,75	90,00	56,25	20,00	-	-
512	100	100	100	100	100	100	81,25	56,25	20,00	-
1024	100	100	100	100	100	100	98,75	75,00	47,50	18,75

Tabel 3. Jumlah koefisien ekstraksi ciri yang untuk berbagai panjang *frame blocking* dan panjang segmen. Hasil yang diperlihatkan: Jumlah koefisien ekstraksi ciri (koefisien)

Panjang <i>frame blocking</i> (titik)	Panjang segmen (titik)									
	1	2	4	8	16	32	64	128	256	512
64	32	16	8	4	2	1	-	-	-	-
128	64	32	16	8	4	2	1	-	-	-
256	128	64	32	16	8	4	2	1	-	-
512	256	128	64	32	16	8	4	2	1	-
1024	512	256	128	64	32	16	8	4	2	1

Pada dasarnya, jika jumlah koefisien ekstraksi ciri makin besar, akan makin banyak koefisien ekstraksi ciri yang digunakan untuk membedakan antara kelas pola yang satu dengan kelas pola yang lainnya. Ini berarti, akan makin mudah dibedakan antara kelas pola yang satu dengan kelas pola yang lain. Dengan makin mudah dibedakannya kelas pola yang satu dengan kelas pola yang lain tersebut, pada akhirnya akan makin meningkatkan tingkat pengenalan.

3.2 Jumlah Koefisien Ekstraksi Ciri Terkecil

Penelitian yang dilaksanakan bertujuan untuk mencari jumlah ekstraksi ciri terkecil, yang dapat merepresentasikan suatu nada, baik nada polifonik maupun nada monofonik. Tabel 1 dan 2 dapat digunakan untuk mencari jumlah koefisien terkecil tersebut, dengan cara mengeksplorasi kombinasi panjang *frame blocking* terkecil dan panjang segmen terbesar, yang menghasilkan tingkat pengenalan tertinggi (100%). Hasil eksplorasi memperlihatkan bahwa panjang *frame*

blocking terkecil yaitu 512 titik dan panjang segmen terkecil yaitu 32 titik, dapat menghasilkan tingkat pengenalan tertinggi. Berdasarkan Tabel 3, jumlah koefisien ekstraksi ciri yang bersesuaian dengan panjang *frame blocking* dan panjang segmen tersebut adalah 8 koefisien. Ini berarti model ekstraksi ciri yang diusulkan dalam penelitian ini sangat efisien, karena nada-nada polifonik maupun monofonik dapat direpresentasikan hanya dengan sejumlah 8 koefisien ekstraksi ciri.

Ada satu hal yang perlu dijadikan perhatian yaitu bahwa jumlah koefisien ekstraksi ciri terkecil, yaitu sejumlah 8 koefisien di atas, terkait dengan penggunaan fungsi *kernel* linear pada pengklasifikasi SVM. Untuk penggunaan fungsi *kernel* yang lain (fungsi polinomial) pada pengklasifikasi SVM, Tabel 1 dan 2 dikerjakan lagi untuk fungsi polinomial orde 2 dan orde 3. Hasil yang didapat dari penggunaan kedua fungsi polinomial tersebut, diperlihatkan pada Tabel 4

Tabel 4. Hasil penggunaan fungsi *kernel* linear dan polinomial yang memberikan tingkat pengenalan tertinggi 100%, untuk nada polifonik dan nada monofonik.

Fungsi <i>kernel</i> pada pengklasifikasi SVM	Linear	Polinomial	
		Orde 2 (Kuadratik)	Orde 3 (Kubik)
Jumlah koefisien ekstraksi ciri terkecil (koefisien)	8	16	64

Sebagaimana terlihat pada Tabel 4, penggunaan fungsi *kernel* linear memberikan hasil yang terbaik, karena hanya dengan menggunakan sejumlah 8 koefisien ekstraksi ciri (yang merupakan jumlah yang terkecil), dapat memberikan tingkat pengenalan tertinggi 100%, untuk nada polifonik maupun nada monofonik yang diujikan. Adanya tingkat pengenalan 100% ini mengindikasikan, bahwa kelas-kelas pola dari ekstraksi ciri sinyal-sinyal nada yang diujikan masuk kategori *linearly separable*. Dengan kata lain, penggunaan sejumlah 8 koefisien ekstraksi ciri dan fungsi *kernel* linear sudah mencukupi, karena sudah dapat memberikan ekstraksi ciri

sinyal-sinyal nada yang masuk kategori *linearly separable*.

3.3 Perbandingan dengan Ekstraksi Ciri beserta Pengklasifikasi yang Lain

Tabel 5 membandingkan kinerja beberapa metode ekstraksi ciri beserta pengklasifikasinya untuk nada-nada alat musik, baik yang polifonik maupun monofonik. Sebagaimana terlihat pada Tabel 5, metode ekstraksi ciri beserta pengklasifikasinya yang diusulkan dalam penelitian ini paling efisien, karena memberikan jumlah koefisien ekstraksi ciri yang paling kecil, untuk merepresentasikan nada yang polifonik maupun monofonik

Tabel 5. Perbandingan kinerja beberapa metode ekstraksi ciri/pengklasifikasi, untuk nada polifonik maupun monofonik. Hasil yang ditampilkan: Jumlah koefisien ekstraksi ciri (alat musik).

Ekstraksi ciri/pengklasifikasi	Jumlah koefisien ekstraksi ciri terkecil	
	Polifonik	Monofonik
<i>Spectral Features</i> /SVM (Tjahyanto <i>et al</i> , 2013)	34 (gamelan)	-
MFCC dan <i>Codebook/template matching</i> (Fruandta & Buono, 2011)	13 (piano pada keyboard)	-
DCT dan perataan segmen/ <i>template matching</i> (Sumarno, 2016)	8 (pianika), 16 (rekorder sopran)	-
FFT dan perataan segmen/ <i>template matching</i> (Sumarno & Iswanjono, 2017)	4 (pianika), 16 (rekorder tenor)	16 (belira)
DST dan perataan segmen/SVM (penelitian ini)	8 (pianika)	8 (belira)

4. KESIMPULAN DAN SARAN

Berikut ini beberapa hal yang dapat disimpulkan, dengan berdasarkan pada hal-hal yang telah dideskripsikan pada hasil percobaan dan pembahasannya.

1. Metode ekstraksi ciri yang diusulkan efisien untuk digunakan pada pengenalan nada alat musik yang polifonik, maupun yang monofonik. Hal ini disebabkan, metode ekstraksi ciri yang diusulkan hanya memerlukan sejumlah 8 koefisien ekstraksi ciri, untuk merepresentasikan suatu nada polifonik maupun monofonik. Sejumlah 8 koefisien tersebut diperoleh dengan penggunaan panjang *frame blocking* 512 titik dan panjang segmen 32 titik.
2. Pengklasifikasi SVM yang digunakan hanya memerlukan fungsi *kernel* linear. Penggunaan sejumlah 8 koefisien ekstraksi ciri dan fungsi *kernel* linear tersebut sudah mencukupi, untuk mendapatkan tingkat pengenalan tertinggi 100%. Adanya tingkat pengenalan 100% ini mengindikasikan

bahwa kelas-kelas pola dari ekstraksi ciri sinyal-sinyal nada yang diujikan, masuk kategori *linearly separable*.

Untuk pengembangan lebih lanjut penelitian ini, disarankan untuk mengerjakan beberapa hal sebagai berikut.

1. Eksplorasi metode transformasi yang lain selain DST, untuk mendapatkan kinerja yang lebih baik dalam hal jumlah koefisien ekstraksi ciri lebih kecil, atau panjang *frame blocking* yang lebih kecil.
2. Eksplorasi fungsi *kernel* yang lain dalam pengklasifikasi SVM, untuk mendapatkan kinerja yang lebih baik dalam hal jumlah koefisien ekstraksi ciri yang lebih kecil.

UCAPAN TERIMA KASIH

Ucapan terima kasih dihaturkan penulis kepada Lembaga Penelitian dan Pengabdian Masyarakat Universitas Sanata Dharma Yogyakarta, atas peran sertanya dalam pendanaan penelitian yang dilaksanakan ini.

DAFTAR PUSTAKA

- Boser, B.E., Guyon, I.M., & Vapnik, V.N., (1992, Juli). A Training Algorithm for Optimal Margin Classifiers, *In Proceedings of the Fifth Annual Workshop on Computational Learning Theory (COLT '92)* (pp 144–152). ACM Press, Pittsburgh.
- Forster, C. (2010). *Musical Mathematics: On the Art of Science and Acoustic Instruments* (pp. vii-viii). California: Chronicle Books LLC.
- Fruandta A, & Buono A., (2011, Juni). Identifikasi Campuran Nada pada Suara Piano Menggunakan Codebook. *Prosiding Seminar Nasional Aplikasi Teknologi Informatika (SNATI) 2011* (pp. G8-G13). Universitas Islam Indonesia, Yogyakarta.
- Gaffar, I., Hidayatno, A., & Zahra, A.A. (2012). Aplikasi Pengkonversi Nada-nada Instrumen Tunggal Menjadi Chord Menggunakan Metode Pitch Class Profile. *Transient*. 1(3), 121-127.
- Izzudin, A., Santoso, T.B., & Dutono T. (2005). Pengenalan Nada-Nada Tunggal Pada Gitar Dengan Menggunakan Teknik Pengolahan Sinyal Digital. *EEPIS Journal Online System*. 10(1).
- McAdams, S. (1993). *Recognition of Auditory Sound Sources and Events. Thinking in Sound: The Cognitive Psychology of Human Audition* (pp. 146-198). Oxford: Oxford University Press.
- Meseguer, N. A. (2009). *Speech Analysis for Automatic Speech Recognition (MSc Thesis)*. Trondheim. NTNU.
- Mitre A, Queiroz M, & Faria R. (2006, Mei). Accurate and Efficient Fundamental Frequency Determination from Precise Partial Estimates. *Proceedings of the 4th AES Brazil Conference* (pp. 113-118). Sao Paulo, Brazil: Audio Engineering Society (AES).
- Noll, M. (1970). Pitch Determination of Human Speech by the Harmonic Product Spectrum, the Harmonic Sum Spectrum and a Maximum Likelihood Estimate. *Proceedings of the Symposium on Computer Processing in Communications*. Brooklyn, New York. Polytechnic Press. 19: 779-797.
- Nurdiyah, D. & Muwakhid, I.A. (2016). Perbandingan Support Vector Machine Dan K-Nearest Neighbor Untuk Klasifikasi Telur Fertil dan Infertil Berdasarkan Analisis Texture GLCM. *Transformatika*. 13(2), 29-34.
- Oppenheim, A.V. & Schaffer, R.W. (1989). *Discrete-Time Signal Processing* (pp. 447-448). New Jersey: Prentice-Hall Inc.
- Rifkin, R.M. & Klautau, A.B.R. (2004). In Defense of One-vs-All Classification. *The Journal of Machine Learning Research*. 5, 101–141.
- Setiawan, Y.R. (2015). *Pengenalan Ucapan Angka Menggunakan Transformasi Fourier Cepat dan Similaritas Kosinus (Skripsi)*. Yogyakarta. Universitas Sanata Dharma.
- Sumarno L. (2016). On The Performace of Segment Averaging of Discrete Cosine Transform Coefficients on Musical Instruments Tone Recognition. *ARPN Journal of Engineering and Applied Sciences*. 11(9), 5644-5649.
- Sumarno L., & Iswanjono. (2017). Feature Extraction of Musical Instrument Tones using FFT and Segment Averaging. *Telkomnika*. 15(3), 1280-1289.
- Tan, L. & Jiang, J. (2013). *Digital Signal Processing Fundamentals and Applications. Second Edition*. Oxford: Elsevier Inc. pp. 15-56.
- Tjahyanto, A., Suprpto, Y.K., & Wulandari, D.P. (2013). Spectral-based Features Ranking for Gamelan Instruments Identification using Filter Techniques. *Telkomnika*. 11(1), 95-106.
- Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*, New York: Springer-Verlag.
- Vapnik, V. N. (1998). *Statistical Learning Theory*. New York: Wiley.