

PENGUNAAN ALGORITMA LEARNING VECTOR QUANTIZATION DALAM MENGENALI SUARA MANUSIA UNTUK KENDALI QUADROTOR

Veronica Indrawati¹, Yudianto Gunawan²

Jurusan Teknik Elektro, Fakultas Teknik, Universitas Surabaya

Jl. Raya Kalirungkut, Surabaya

031-2981157, ext 87

E-mail: veronica@staff.ubaya.ac.id, Yudiantogunawan92@gmail.com

ABSTRAK

Paper ini membahas tentang bagaimana suatu kata yang diucapkan manusia dapat dikenali oleh sebuah sistem, dengan tujuan untuk mengendalikan gerak quadrotor melalui instruksi berupa kata-kata "Hover", "Landing", "Right", "Left", "Forward", "Backward" yang diberikan operator quadrotor.

Pada paper sebelumnya yang berjudul "Aplikasi Learning Vector Quantization Untuk Pengenalan Suara Manusia Dengan Menggunakan Mel Frequency Cepstral Coefficient", yang membahas pemrosesan sinyal suara hingga didapatkan koefisien Mel, dan diproses dengan LVQ (Learning Vector Quantization) untuk satu responden saja. Dalam paper ini koefisien-koefisien dari proses MFCC (Mel Frequency Cepstral Coefficient) telah didapatkan dan dilakukan penelitian untuk memperoleh konfigurasi net yang optimal untuk algoritma pembelajaran LVQ dalam mengenali suara manusia untuk mengendalikan gerak quadrotor. Dengan menggunakan algoritma pembelajaran yang tepat maka kelemahan dari penelitian sebelumnya dapat diminimalisir dan akurasi dapat ditingkatkan lebih baik lagi.

Kata Kunci: LVQ, MFCC

ABSTRACT

This paper discusses how a man spoken word can be recognized by a system, with the aim to control Quadrotor motion through the instruction of the words "Hover", "Landing", "Right", "Left", "Forward", "Backward" given by Quadrotor's operator.

In a previous paper entitled "Application Learning Vector Quantization For Human Voice Recognition Using Mel frequency cepstral coefficient", which discusses the speech signal processing to obtain the coefficient of Mel, and processed with a LVQ (Learning Vector Quantization) for one respondent. In this paper the coefficients of the MFCC (Mel Frequency cepstral coefficient) has been obtained and conducted research to obtain the optimal configuration of the net for LVQ learning algorithm to recognize the human voice to control the motion of Quadrotor. By using the appropriate learning algorithms weaknesses of previous research can be minimized and the accuracy can be improved.

Keywords: LVQ, MFCC

1. PENDAHULUAN

Pada paper sebelumnya telah dilakukan penelitian pemrosesan sinyal suara manusia hingga didapatkan koefisien mel atau biasa yang disebut dengan MFCC. Algoritma yang digunakan dalam mengolah MFCC pada penelitian sebelumnya yaitu dengan menggunakan neural network LVQ dan Feed Forward (FF). Hasil dari penelitian sebelumnya membuktikan bahwa LVQ lebih cocok untuk dipergunakan dalam mengolah MFCC dibandingkan dengan FF.

Pada penelitian sebelumnya, sistem baru dicobakan untuk satu responden. Untuk pengembangan penelitian sebelumnya maka pada penelitian kali ini akan bertujuan untuk meminimalisir kesalahan yang terjadi sehingga sistem dapat mengenali suara lebih akurat lagi.

1.1 Tujuan

Tujuan dari penelitian ini yaitu untuk membuat sistem pengenalan suara yang lebih baik lagi agar

dapat digunakan secara lebih maksimal oleh semua orang yang akan mengoperasikan quadrotor melalui instruksi-instruksi kata-kata yang diberikan, dengan menggunakan LVQ sebagai pengolah MFCC dengan tetap menggunakan MATLAB.

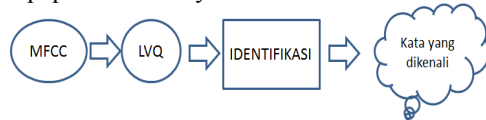
1.2 Batasan Masalah

Dalam menyelesaikan penelitian ini maka diberikan batasan masalah sebagai berikut:

- Sistem hanya dapat mengenali kata yang terdapat dalam database yang akan digunakan untuk mengatur gerak quadrotor, yaitu kata 'backward', 'forward', 'hover', 'landing', 'right', dan 'left'.
- Bahasa pemrograman menggunakan MATLAB.
- Sistem hanya dapat mengenali kata dengan pengucapan tempo nada yang mirip dalam database.
- Noise yang ada bukan merupakan suara manusia.

2. DASAR TEORI

Secara garis besar pada penelitian ini blok diagram dari sistem yang dibuat dapat dilihat pada Gambar 1 berikut ini. Pemrosesan sinyal suara manusia hingga didapatkan MFCC telah dibahas pada paper sebelumnya.



Gambar 1. Blok diagram sistem

Dari blok diagram ini hasil MFCC yang hanya berupa koefisien-koefisien akan diolah dengan menggunakan neural network LVQ untuk mengenali kata yang diucapkan oleh responden. Paper ini membahas percobaan-percobaan untuk mendapatkan net LVQ yang optimal untuk dapat mengenali suara manusia.

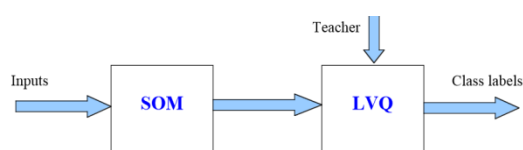
2.1 MFCC

MFCC merupakan suatu metode ekstraksi yang banyak digunakan dalam proses pengenalan suara. MFCC banyak digunakan karena merupakan hasil pengadopsian dari sistem pendengaran manusia. Dalam proses ekstraksi suara MFCC akan mengubah sinyal analog input menjadi sekumpulan koefisien yang biasa disebut dengan koefisien mel atau MFCC. Koefisien yang dihasilkan memiliki informasi-informasi penting yang terkandung dalam sinyal analog input.

2.2 LVQ

LVQ merupakan salah satu dari jenis neural network yang menggunakan pembelajaran jenis supervised learning. LVQ merupakan hasil pengembangan dari algoritma SOM (Self Organizing Map) yang dikembangkan oleh Teuvo Kohonen sehingga berdasarkan hal ini maka LVQ hanya memiliki 1 buah hidden neuron saja dalam proses pembelajarannya.

Pembelajaran LVQ ini cukup ideal karena dengan algoritma SOM maka input-input yang diberikan pada saat proses pembelajaran yang memiliki sifat yang mirip akan saling berkelompok sendiri sehingga proses pengenalan input terhadap output (teacher) akan diklasifikasi sendiri sehingga hasil yang diberikan pada proses pembelajaran ini akan memiliki hasil yang lebih akurat. Untuk lebih jelasnya bagaimana proses pembelajaran menggunakan LVQ maka dapat dilihat pada Gambar 2



Gambar 2. Proses LVQ^[9]

Tujuan digunakannya LVQ ini yaitu untuk mengelompokkan input terhadap output dalam klasifikasi vektor agar dapat meminimalkan proses terjadinya kesalahan dalam klasifikasi.

3. HASIL PENELITIAN

Hasil penelitian pada saat penelitian sebelumnya menggunakan data training dan data uji yang direkam pada hari yang sama dari satu responden. Dengan merekam suara pada hari yang sama maka karakter dari suara akan memiliki tempo nada yang mirip sehingga sistem akan mudah untuk mengenali kata yang suara yang diucapkan. Hasil dari penelitian sebelumnya dapat dilihat pada Tabel 1 berikut

Tabel 1. Hasil penelitian Sebelumnya

Percobaan	Frame block	Windowing	TOBF	Neural Network	Hover	Landing	Right	Left	Forward	Backward	Rata-rata
1	25	Hamming	30	LVQ	76.7	100	96.7	90	83.3	83.3	88.33333
2	25	Hamming	32	LVQ	60	100	93.3	90	83.3	83.3	84.98333
3	25	Hamming	34	LVQ	76.7	100	93.3	90	83.3	83.3	87.76667
4	25	Hamming	36	LVQ	80	100	93.3	90	83.3	83.3	88.31667
5	25	Hamming	38	LVQ	83.3	100	93.3	90	83.3	83.3	88.86667
6	25	Hamming	40	LVQ	60	100	93.3	90	83.3	83.3	84.98333
7	30	Hamming	30	LVQ	83.3	100	93.3	90	83.3	83.3	88.86667
8	30	Hamming	32	LVQ	96.7	100	96.7	70	83.3	83.3	88.33333
9	30	Hamming	34	LVQ	90	100	93.3	90	83.3	83.3	89.98333
10	30	Hamming	36	LVQ	93.3	100	90	90	83.3	83.3	89.98333
11	30	Hamming	38	LVQ	86.7	100	90	90	83.3	83.3	88.88333
12	30	Hamming	40	LVQ	73.3	100	93.3	90	83.3	83.3	87.2
13	25	Hamming	30	FF	33.3	93.3	16.7	83.3	50	83.3	59.98333
14	25	Hamming	32	FF	26.7	46.7	70	90	80	80	65.56667
15	25	Hamming	34	FF	46.7	100	36.7	60	63.3	83.3	65
16	25	Hamming	36	FF	16.7	83.3	90	20	83.3	80	62.21667
17	25	Hamming	38	FF	56.7	76.7	80	90	76.7	83.3	77.23333
18	25	Hamming	40	FF	63.3	100	70	40	70	83.3	71.1
19	30	Hamming	30	FF	26.7	100	66.7	80	83.3	83.3	73.33333
20	30	Hamming	32	FF	76.7	66.7	23.3	86.7	50	83.3	64.45
21	30	Hamming	34	FF	40	36.7	43.3	16.7	36.67	83.3	42.77833
22	30	Hamming	36	FF	26.7	96.7	86.7	90	0	83.3	63.9
23	30	Hamming	38	FF	83.3	40	60	40	70	83.3	62.76667
24	30	Hamming	40	FF	30	100	50	80	20	76.7	59.45

Dengan menggunakan sistem yang sama saat diujikan pada pada responden lain beberapa hari selanjutnya, terjadi penurunan akurasi yang ekstrim sehingga akurasi rata-rata yang mencapai 89.98% ini turun menjadi di bawah 16% walaupun telah dilakukan training ulang dengan menambahkan jumlah data training yang berisikan rekaman di hari berikutnya. Jadi system masih tidak dapat mengenali suara yang responden lain dengan data training.

Berdasarkan hal ini maka dilakukan pengujian dengan melakukan training ulang untuk mengetahui apakah sistem dapat mengenali pemilik suara dari suatu kata yang sama yang diucapkan pada hari yang berbeda yang hasilnya dapat dilihat pada Tabel 2.

Tabel 2. Uji coba tempo nada

Kata	(hasil dalam %)	
	Responden1	Responden2
Hover	93	96.7
Landing	100	100
Right	76.7	73.3
Left	90	93
Forward	83.3	83.3
Backward	76.7	80

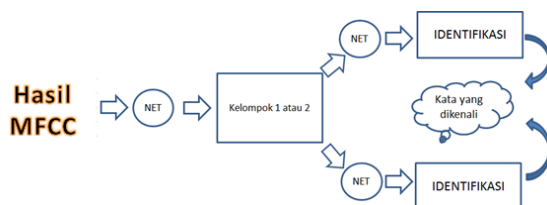
Berdasarkan hasil dari uji coba Tabel 2 ini maka dapat disimpulkan bahwa satu buah net LVQ tidak mampu untuk mengenali 2 variabel untuk dikenali (pemilik suara dan kata yang diucapkan) oleh karena itulah maka berikutnya kinerja dari net LVQ akan diperingan dengan sehingga sistem akan menggunakan lebih dari satu buah net.

Percobaan berikutnya yaitu untuk mengetahui kemiripan dari masing-masing kata yang diucapkan dengan melakukan training satu banding satu yakni kata 'hover' akan diuji dengan 5 kata yang lain dan kata 'landing' akan diuji dengan 5 kata yang lain dan seterusnya yang hasilnya dapat dilihat pada Tabel 3 di bawah ini

Tabel 3. Percobaan 1 banding 1

Train NET	Kata yang diuji dikenali sebagai					
	Hover (H)	Landing (La)	Right (R)	Left (Le)	Forward (F)	Backward (B)
HLa	Hover	Landing	Landing	Landing	Hover	Hover
HR	Hover	Right	Right	Right	Hover	Hover
HLe	Hover	Left	Left	Left	Hover	Hover
HF	Hover	Forward	Forward	Forward	Forward	Forward
HB	Hover	Backward	Backward	Backward	Hover	Backward
LaH	Hover	Landing	Landing	Landing	Hover	Hover
LaR	Right	Landing	Right	Right	Right	Right
LaLe	Left	Landing	Left	Left	Left	Left
LaF	Forward	Landing	Landing	Landing	Forward	Forward
LaB	Backward	Landing	Landing	Landing	Backward	Backward
RH	Hover	Right	Right	Right	Hover	Hover
RLa	Right	Landing	Right	Right	Right	Right
RLe	Right	Left	Right	Left	Right	Right
RF	Forward	Right	Right	Right	Forward	Forward
RB	Backward	Right	Right	Right	Backward	Backward
LeH	Hover	Left	Left	Left	Hover	Hover
LeLa	Landing	Landing	Left	Left	Left	Left
LeR	Right	Left	Right	Left	Right	Right
LeF	Forward	Left	Left	Left	Forward	Forward
LeB	Backward	Left	Left	Left	Backward	Backward
FH	Hover	Forward	Forward	Forward	Forward	Forward
FLa	Forward	Landing	Landing	Landing	Forward	Forward
FR	Forward	Right	Right	Right	Forward	Forward
FLe	Forward	Left	Left	Left	Forward	Forward
FB	Forward	Backward	Backward	Backward	Forward	Backward
BH	Hover	Backward	Backward	Backward	Backward	Backward
BLa	Backward	Landing	Landing	Landing	Backward	Backward
BR	Backward	Right	Right	Right	Backward	Backward
BLe	Backward	Left	Left	Left	Backward	Backward
BF	Forward	Backward	Backward	Forward	Forward	Backward

Berdasarkan hasil dari Tabel 3 maka dapat disimpulkan bahwa kata 'hover', 'forward', dan 'backward' memiliki karakteristik yang mirip sedangkan kata 'landing', 'right' dan 'left' memiliki karakteristik yang mirip sehingga dibuatlah desain sebuah net yang dipergunakan untuk membagi kelompok karakteristik kata. Pembuatan desain ini dapat dilihat pada Gambar 2.



Gambar 2. Desain net akhir

Dengan menggunakan desain pada Gambar 2 ini maka hasil MFCC yang masuk akan dikenali oleh net pertama untuk mengetahui apakah kata yang dikenali masuk pada kelompok 1 ('hover', 'forward', 'backward') atau masuk pada kelompok 2 ('landing', 'right', 'left'). Setelah mengetahui kelompok tersebut maka akan dikenali oleh net yang ditujukan untuk mengenali kata dalam kelompok. Jadi jika kata yang masuk dikenali sebagai

kelompok 1 maka pada net berikutnya akan dilakukan identifikasi sehingga outputnya antara kata 'hover', 'forward' dan 'backward'.

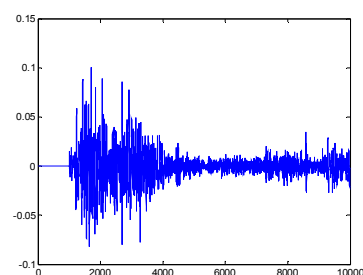
Dengan menggunakan desain seperti ini maka pengolahan MFCC dengan menggunakan LVQ yang sebelumnya hanya memiliki akurasi di bawah 16% untuk mengenali tempo nada yang berbeda akan memiliki kenaikan akurasi yang dapat dilihat pada Tabel 4.

Tabel 4. Pengaruh penambahan data training dengan tempo berbeda

Percobaan	NET (Data Train)	HASIL UJI (NET) (dalam %)			
		hari1	hari2	hari3	hari4
1	Hari 1	< 85	< 60	< 30	< 30
2	Hari 1+2	< 85	< 70	< 45	< 40
3	Hari 1+2+3	< 85	< 70	< 55	< 40
4	Hari 1+2+3+4	< 85	< 70	< 55	< 50

Berdasarkan hasil dari Tabel 4, maka dapat disimpulkan bahwa akurasi dari sistem yang didesain dengan menggunakan lebih dari satu net ini dapat ditingkatkan dengan ditambahkannya jumlah data training, akan tetapi ternyata desain sistem ini memiliki batas maksimal dalam mengolah jumlah data training, hal ini dapat dilihat pada Tabel 4 yang menunjukkan bahwa sistem hanya optimal untuk mengenali kata di 2 hari yang berbeda saja.

Percobaan akhir yang dilakukan pada penelitian kali ini yaitu menguji ketahanan sistem yang dibuat ini terhadap noise. Pada percobaan noise ini dilakukan 2 percobaan yaitu percobaan dengan noise berupa suara kipas angin dan percobaan menggunakan noise berupa suara manusia. Bentuk sinyal noise dari suara kipas angin dapat dilihat pada Gambar 3 berikut



Gambar 3. Sinyal noise dari kipas angin

Pada saat proses pengujian dengan adanya noise kipas ini ternyata dari 20 kali percobaan, 15 diantaranya dikenali dengan baik (tempo nada ditemukan terlebih dahulu). Berikutnya percobaan dilakukan dengan adanya gangguan suara manusia seperti lagu dan ruangan yang ramai (ada manusia yang berbicara), ternyata dari 20 kali percobaan hanya 2 saja yang dikenali.

Berdasarkan hasil dari penelitian yang telah dilakukan maka terdapat beberapa faktor yang mempengaruhi keberhasilan sistem antara lain:

1. Tempo nada yang diucapkan.
2. Kejelasan dari kata yang diucapkan.
3. Tingkat noise yang ada dalam ruangan.

4. KESIMPULAN

Berdasarkan dari hasil percobaan dan penelitian yang telah dilakukan maka dapat disimpulkan bahwa:

1. Tempo nada merupakan faktor penting dalam proses pengenalan suara dengan menggunakan MFCC.
2. Metode pengolahan menggunakan LVQ memiliki batasan maksimum dalam pembelajarannya.
3. Metode MFCC memiliki kelemahan tinggi terhadap noise berupa suara manusia.

5. SARAN

Adapun saran yang diberikan sehubungan dengan pelaksanaan penelitian ini adalah sebagai berikut:

1. Jumlah jenis kata yang digunakan dapat ditingkat lebih banyak lagi.
2. Perlu adanya dilakukan penelitian terhadap tempo nada.
3. Perlu adanya penelitian untuk menggabungkan metode-metode ekstraksi suara.

PUSTAKA

[1] Indrawati V., dan Yudianto G., "Aplikasi Learning Vector Quantization Untuk Pengenalan Suara Manusia Dengan Menggunakan Mel Frequency Cepstral Coefficient", 2013

[2] Davis, S. Mermelstein, P., "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences". In IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28 No. 4, pp. 357-366., 1980

[3] Siang, J. J., "Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan Matlab", Penerbit Andi, Yogyakarta, 2005.

[4] X. Huang, A. Acero, and H. Hon., "Spoken Language Processing: A guide to theory, algorithm, and system development". Prentice Hall, 2001.

[5] Ajcl., "Frames Representation of Speech Signal", <http://basic-programming.blogspot.com/2005/11/frames-representation-of-speech-signal.html>, 2005.

[6] Brookes, M., "VOICEBOX: a MATLAB toolbox for Speech Processing", www.ee.ic.ac.uk/hp/staff/dmb/voicebox.html, 2003

[7] Kyaw, T.T., "Robust Speaker Identification/Verification For Telephony Applications", <http://docs.exdat.com/docs/index-87374.html>, SIM University, 2009.

[8] Tiwari, Ganesh., "Silence Removal and End Point Detection MATLAB Code", <http://ganeshtiwariidotcomdotnp.blogspot.com/2011/08/silence-removal-and-end-point-detection.html>, 2011.

[9] John A. Bullinaria, "Introduction to Neural Networks Lecture 16-18", University of Birmingham, 2004

[10] Developer, NOKIA., "Digital Signal Processing", https://projects.developer.nokia.com/DSP/wiki/Mel_frequency_cepstral_coefficients, 2012.

[11] Suhartono, D., "Dasar Pemahaman Neural Network", <http://socs.binus.ac.id/2012/07/26/konsep-neuralnetwork/>, 2012

[12] K. Preetham, "Speech Recognition Using MFCC Feature Extraction". <http://www.scribd.com/doc/161437587/SPEECHRECOGNITIONUSING-MFCC-FEATURE-EXTRACTION>, 2013

[13] Manunggal, HS. 2005. "Perancangan dan Pembuatan Perangkat Lunak Pengenalan Suara Pembicara dengan Menggunakan Analisa MFCC Feature Extracton. Surabaya: Universitas Kristen Petra.

[14] Pah. Nemuel, D., Modul Kuliah, "Discrete Fourier Transform", Universitas Surabaya.

[15] Lyons. James, "Mel Frequency Cepstral Coefficient (MFCC) Tutorial", <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>, 2009-2012

[16] Sphinx, "Class Mel Frequency Filterbank", <http://cmusphinx.sourceforge.net/sphinx4/javadoc/edu/cmu/sphinx/frontend/frequencywarp/melfrequencyfilterbank.html>

[17] Kurniawan, A.T., "The Discrete Cosine Transform (DFT) Part 1- Introduction", <http://coretancempluk.wordpress.com/2010/02/05/the-discrete-fourier-transform-dft-part-1-introduction/>, 2010

[18] Wikipedia, "Window Function", http://en.wikipedia.org/wiki/window_function, 2013

[19] Kurniawan, D.E., "Konsep Learning Vector Quantization (LVQ)", <http://ikhs.wordpress.com/2011/07/03/konsep-learning-vector-quantization-lvq/>, 2011

[20] Budi, "Time Series Forecasting Dengan Seasonal Time Series Arima dan Feed Forward Neural Network", <http://statistikakomputasi.wordpress.com/2010/04/20/time-series-forecasting-dengan-seasonal-time-series-arima-dan-feed-forward-neural-network/>, 2010

- [21] Richards G. Lyons, 2004, “ Understanding Digital Signal Processing”, Prentice-Hall
- [22] Douglas F. Elliott, 1987, “Handbook of Digital Signal Processing, Engineering Applications”, Academic Press Inc.
- [23] Vijay K. Madisetti, 1999, “Digital Signal Processing Handbook”, Chapman & Hall.
- [24] B.A.Shenoi, 2006, “Introduction to Digital Signal Processing and Filter Design”, John Wiley & Sons.
- [25] Endi, “*Proses Produksi Suara Manusia*”, <http://dejavuanakselatan.blogspot.com/2011/01/proses-produksi-suara-manusia.html>, 2009
- [26] MATLAB Product help., “MATLAB Compiler: Introducing the MATLAB Compiler: Why Compile M-Files?”, 2001.