

Pengenalan Pola Gerak Mulut Dengan Menggunakan Pendekatan SOM (Self Organizing Feature Map)

I Gede Aris Gunadi

Jurusan Pendidikan Fisika, FMIPA
Universitas Pendidikan Ganesha Singaraja
Singaraja-Bali, Indonesia
Aria_sukaat@yahoo.com

Agus Harjoko

Lab Elektronika dan Instrumentasi, FMIPA
Universitas Gajah Mada
Yogyakarta, Indonesia
aharjoko@ugm.ac.id

Abstract—Pengenalan pola gerak mulut merupakan salah satu acuan yang digunakan untuk pengenalan ekspresi emosi manusia berdasarkan ekspresi wajah. Emosi manusia dapat diklasifikasikan menjadi enam sampai tujuh kategori yaitu, marah, khawatir, sedih, jijik, bahagia, terkejut, dan normal. Dalam penelitian ini akan dilakukan pengenalan pola gerak mulut yang terjadi dalam sebuah adegan rekaman video. Rekaman video terlebih dahulu diekstraksi dalam bentuk image frame. Pada setiap image frame dilakukan segmentasi untuk mendapat area mulut. Metode segmentasi mulut yang digunakan adalah dengan menggunakan peta mulut, yang didasarkan pada warna bibir yang khas. Proses selanjutnya adalah melakukan *crooping* berdasarkan tepi mulut yang sudah tersegmentasi. Selanjutnya dilakukan penentuan feature tepi dari hasil *crooping* yang didapat. Pada Tahap akhir dilakukan pengenalan pola gerak mulut dengan menggunakan metode SOM (self Organizing feature Maps), sehingga pola gerak mulut yang sejenis akan berada dalam cluster yang sama. Hasil eksperimen menunjukkan pengenalan gerak mata dengan menggunakan SOM mampu mengidentifikasi dengan keakuratan 83%.

Keywords : Jaringan Kohonen, SOM, Pengenalan Pola, Segmentasi Mulut.

I. PENDAHULUAN

Analisa gerak mulut merupakan bagian yang sangat penting dalam pengenalan ekspresi emosi manusia. Secara umum literatur literatur psikologi membagi emosi manusia tujuh yaitu, *anger, disgust, fear, happiness, sadness, surprise, contempt* [1,2]. Pengenalan pola gerak mulut atau ekspresi oral menjadi sangat penting, berdasarkan formula FACs (*Facial Action Code*) pengenalan ketujuh jenis emosi selalu melibatkan AUs (*Action Units*) otot-otot di daerah mulut [3]. Dalam analisa kebohongan yang didasarkan *micro gesture* wajah, secara dominan melibatkan daerah wajah bagian bawah, yang dominan berada pada area mulut [4].

Mulut adalah salah satu bagian dari ekspresi wajah yang dapat dikaitkan dengan emosi seseorang, pada orang marah ekspresi mulutnya tentu berbeda dengan orang yang berbahagia. Bahkan pada orang yang benar-benar bahagia dengan orang yang pura-pura bahagia, ekspresi mulutnya akan berbeda. Dalam [5], beberapa ekspresi yang dikaitkan dengan mulut, antara lain: *Jaw Droop* (Rahang menurun) dikaitkan dengan mulut terbuka yang kemungkinan memiliki makna keterkejutan, *Lip Compression* menyatakan kondisi tertekan atau stress, *Lip pout* (menyebil) menyatakan kondisi ketidakpuasan.

Dalam bidang *computer science* beberapa studi dan penelitian yang dikaitkan dengan mulut, pada umumnya dikaitkan dengan masalah deteksi mulut. Diantaranya [6,7,8], penelitian untuk mendeteksi mulut berdasarkan warna bibir. Penelitian dilakukan dalam sistem warna YCbCr, area mulut memiliki karakter kuat pada komponen warna merah dan lemah pada komponen warna biru, dibanding area wajah lainnya. Berdasarkan fakta tersebut dirumuskan sebuah formula untuk mendeteksi mulut yang dikenal dengan *Mouth Map*. Salah satu penelitian yang dikaitkan dengan pengenalan gerak otot-otot pada wajah adalah, penelitian untuk mengukur dan mengidentifikasi mikrogesture dengan menggunakan *3D gradient descriptor* dan kamera yang digunakan memiliki spesifikasi khusus dengan frame rate 200 fps [7]. Kamera kecepatan tinggi diperlukan untuk menangkap *micro gesture* (gesture yang durasi kejadiannya sangat singkat, kurang dari setengah detik).

Jaringan Kohonen SOM adalah salah satu model pengembangan jaringan syaraf tiruan yang proses pembelajarannya tidak terbimbing (*unsupervised*), model ini sangat cocok dipakai untuk pengenalan pola masukan ke dalam beberapa cluster. Pengembangan jaringan Kohonen dikenal dengan SOM (*Self Organizing Feature Maps*) atau Kohonen Map. Keluaran jaringan

adalah cluster dengan anggota anggota yang paling mirip [9,10].

Dalam digunakan pendekatan peta mulut (*mouth map*) sebagai metode segmentasi mulut dan kohonen SOM untuk mengenali pola gerak mulut pada sebuah adegan video. Metode segmentasi mulut dibahas pada bagian II, pada bagian III akan dibahas tentang metode *cropping* area mulut , pada bagian IV dibahas tentang fitur mulut yang digunakan, percobaan, pada bagian V dibahas tentang pengenalan pola dengan SOM dan hasil dibahas pada bagian VI, dan bagian VII tentang kesimpulan.

II. SEGMENTASI AREA MULUT.

Bibir memiliki karakter warna yang khas yaitu warna merah kuat sedang warna biru rendah, dibandingkan dengan area lain pada wajah. Berdasarkan fakta tersebut diturunkan formula untuk mendeteksi mulut, yang dikenal dengan *mouth map* [11]. Rumusan peta mulut dinyatakan dalam persamaan 1.

$$MouthMap = Cr^2 \cdot (Cr^2 - \eta \cdot Cr/Cb)^2 \quad (1)$$

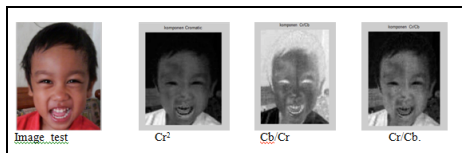
$$\eta = 0.95 \cdot x \cdot ((1/N) \sum Cr^2) / ((1/N) \sum (Cr/Cb)) \quad (2)$$

N = ukuran spatial dari image wajah.

Proses yang dilakukan adalah analisa dilakukan pada sistem warna YCbCr , pada umumnya sistem warna yang digunakan adalah RGB, sehingga dilakukan konversi dari RGB ke YCbCr. Hubungan system warna RGB dengan YCbCr adalah :

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3)$$

Proses normalisasi dilakukan pada komponen Cr^2 dan Cr/Cb , sehingga nilai komponen warna tetap berkisar antara 0 – 255. Gambar 1 menunjukkan image dalam komponen RGB, Cr^2 , Cr/Cb , dan Cb/Cr .



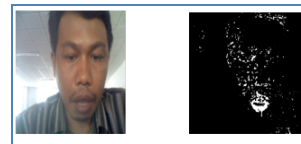
Gambar. 1. Image dalam komponen YCbCr

Komponen Cr^2 dan Cb/Cr digunakan dalam (1), untuk mendeteksi letak mulut. Hasil deteksi sangat dipengaruhi data inputannya, kemungkinan adanya noise akan sangat mempengaruhi hasil segmentasi mulut. Gambar 2 menunjukkan hasil uji coba deteksi mulut dengan (1).



Gambar .2. Segmentasi mulut dengan peta mulut




Pre processing sangat diperlukan untuk menghilangkan adanya noise, yang diakibatkan proses *capture* video yang masih kurang sempurna. Dalam penelitian ini ditemukan noise pada hasil segmentasi dengan peta mulut, ditunjukkan dalam Gambar 3.



Gambar 3. Noise pada hasil segmentasi mulut .

Dalam percobaan ini dicoba untuk memisahkan antara noise dengan area mulut dengan proses morfologi dilatasi dengan ukuran kernel (7,7). Ukuran kernel (7,7) merupakan hasil uji coba yang paling ideal, titik noise bisa tereduksi namun titik putih area mulut masih dominan. Hasil proses morfologi secara umum noise cukup tereduksi, namun tetap masih ada . Proses kompresi dilakukan dengan mengecilkan ukuran image frame, dan hasilnya secara signifikan dapat mengurangi noise. Hasil yang ideal didapatkan pada ukuran 250 x 250, pada kondisi ini noise sudah bisa dihilangkan. Tabel I menunjukkan hasil ujicoba perubahan ukuran image frame dan pengaruh noise.

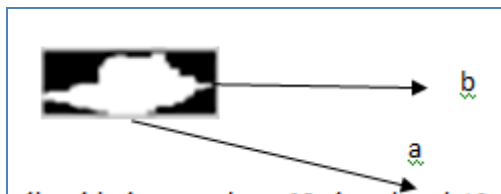
TABEL I. UKURAN KOMPRESI IMAGE DAN PENGURANGAN NOISE

No	Ukuran Kompresi	Hasil Segmentasi mulut
1	1024 x 1280	
2	900 x 900	
3	700 x 700	

5	400 x 400	
---	-----------	---

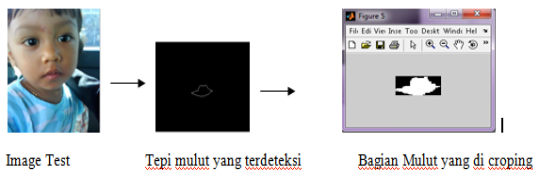
III. CROPPING AREA MULUT.

Dari hasil deteksi mulut yang telah dilakukan, kemudian dilakukan deteksi tepi mulut dengan tujuan untuk mendapatkan tepi tepi mulut, tepi atas bawah, tepi kiri kanan. Titik titik tersebut dijadikan acuan untuk melakukan *cropping* area mulut, ditunjukkan Gambar 4. Proses tersebut berikut hasil *cropping* ditunjukkan Gambar 5.



Gambar 4. Titik acuan untuk *cropping* mulut

Titik a dan b akan dijadikan acuan untuk *cropping* mulut pada setiap *frame*.













Gambar 5. Proses *cropping* mulut dan hasil *cropping*.

Hasil *cropping* area mulut pada setiap *frame* video akan berbeda beda tergantung pada titik acuannya, sehingga ukuran (*size*) citra akan berbeda beda mulut. Sehingga untuk proses perhitungan perlu dilakukan normalisasi, penyamaan ukuran citra. Pada penelitian ini setelah dinormalisasi ukuran hasil *cropping* pada setiap *frame* video adalah 41 x 66 *pixels*.

IV. FITUR MULUT.

Kontur tepi sering digunakan untuk mengidentifikasi sebuah image, misalnya dalam mengenali sidik jari. Sama seperti jari area mulut juga memiliki guratan guratan tertentu, setiap gerak mulut tertentu akan memiliki bentuk guratan atau kontur tepi tersendiri. Pada beberapa beberapa gerak mulut yang telah disegmentasi dilakukan deteksi tepi dengan operator canny, didapatkan hasil bahwa untuk setiap gerak mulut tertentu memiliki pola guratan tepi tertentu. Tabel I menunjukkan berbagai pola gerak mulut dan guratan tepinya.

TABEL II. POLA GERAK MULUT DAN BENTUK GURATAN TEPI

No	Image pola gerak mulut	Deteksi tepi	
1			
2			
3			
4			
5			

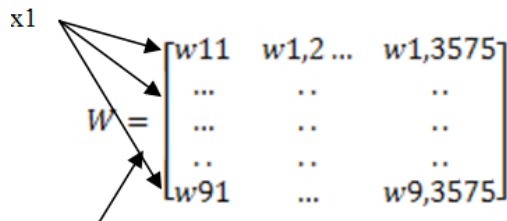
V. PENGENALAN POLA DENGAN SOM

SOM (*self Organizing Map*) adalah sebuah metode pengenalan pola yang tidak terbimbing. Tujuan dari metode ini adalah membentuk *cluster* sesuai dengan kedekatan ciri atau *features* dari masing masing item. Ciri yang digunakan dalam penelitian ini adalah bentuk guratan tepi, masing masing state gerak mulut memiliki bentuk guratan tepi tertentu (ditunjukkan dalam Tabel II).

Pengenalan pola dengan SOM, terdiri atas dua fase yaitu fase training dan fase pencocokan dijelaskan sebagai berikut

A. Fase training

1. Untuk setiap image hasil *cropping* area mulut yang berukuran $j \times k$, dijadikan sebuah flat vector $1 \times m$, dimana $m = j \times k$. Dalam penelitian ini ukuran (*size*) adalah 41 x 66 *pixels*.
2. Mula mula inisialisasi bobot W , dibangkitkan secara acak dengan nilai elemen matrik antara 0 – 255 dan ukurannya adalah $n \times m$, dimana n adalah jumlah *cluster* yang diinginkan dan m adalah ukuran *flat vector* citra.
3. Dilakukan q kali iterasi (q adalah jumlah citra inputan yang diuji untuk pembentukan *cluster*), dalam setiap iterasi dilakukan pengecekan/ perhitungan untuk setiap *vector* inputan citra (*vector* inputan adalah *flat vector* citra, pada langkah 1), temukan jarak yang paling minimum antara *vector* inputan dengan dengan setiap baris dari matrik bobot W (ilustrasi dalam Gambar 6)



Hitung selisih, W dan X1. Dengan menggunakan Euclian distance
 Misal pada W baris 1,
 $D = (W1(1) - X1(1))^2 + (W1(2) - X1(2))^2 + \dots + (W1(3575) - X1(3575))^2$

Gambar . 6. Penghitungan Bobot Untuk Setiap Inputan

4. Dari langkah 2, didapatkan 9 jarak, tentukan yang paling minimum pada baris ke berapa. Pada baris dengan jarak minimum lakukan modifikasi Bobot W, dengan aturan

$$W_{(i,j)}^{baru} = W_{lama(i,j)} + \alpha (0) (X_i - W_{lama(i,j)}) \quad (4)$$

5. Lakukan langkah 2,3 sampai vector inputan X9 (semua vector sudah melakukan modifikasi bobot).
6. Setelah langkah 2,3,4 dan 5 dilakukan berarti satu iterasi selesai, lakukan iterasi berikutnya dengan langkah yang sama seperti 2,3,4. Dengan menyesuaikan perubahan laju pemahaman ,
 $0.5 \alpha (t) = \alpha (t+1) \quad (5)$

Iterasi dianggap cukup , bila perubahan matrik bobot sangat kecil, sudah dianggap konvergen.

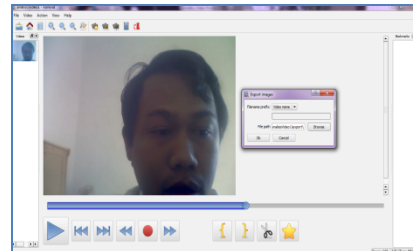
B. Fase Pencocokan

Pencocokan dapat dilakukan, setelah didapat bobot yang ideal. Citra inputan (dalam bentuk *flat vector*) diuji , mencari jarak mana yang paling minimum pada Matrik bobot idealnya. Misal minimum pada baris 2, berarti polanya sesuai dengan pola pada baris no 2.

VI. HASIL PERCOBAAN DAN ANALISA

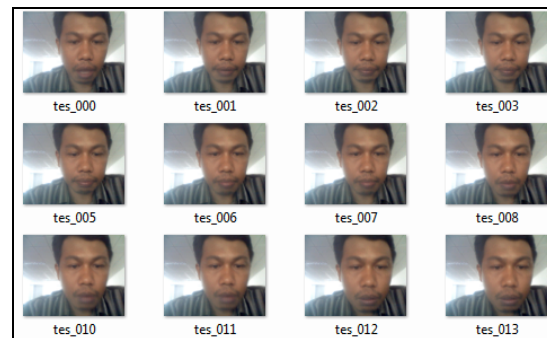
Pada tahap awal dilakukan perekaman video dan ekstraksi image frame. Proses pengambilan video menggunakan web camera, Lenovo Easy Camera, (*Device* ini merupakan bawaan pada Lenovo, diproduksi Vimicro), dengan resolusi maksimal 1280 x 1024. Jenis video yang digunakan berupa *. Wmv. Proses ekstraksi rekaman video menjadi image frame penyusunnya, menggunakan software *Forevid*(open source).Objek yang diambil dalam rekaman

video adalah tubuh bagian atas, bagian muka. Gambar 7 menunjukkan tampilan *interface* dari *foreVid*



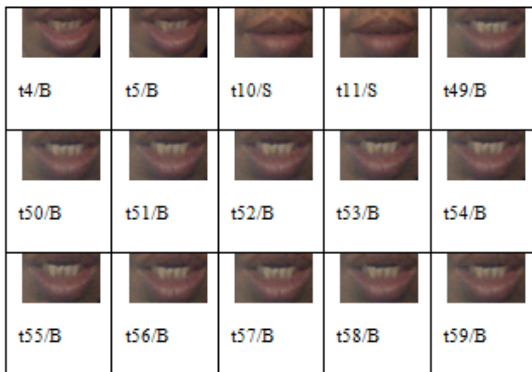
Gambar . 7. Tampilan Interface Forevid

Rekaman video tersebut selanjutnya diekstraksi menjadi *image frame*, proses ekstraksi dari video menjadi image frame menggunakan fitur fitur yang sudah tersedia pada *forevid*. Gambar 8 menunjukkan hasil *image frame* yang telah di ekstraksi.



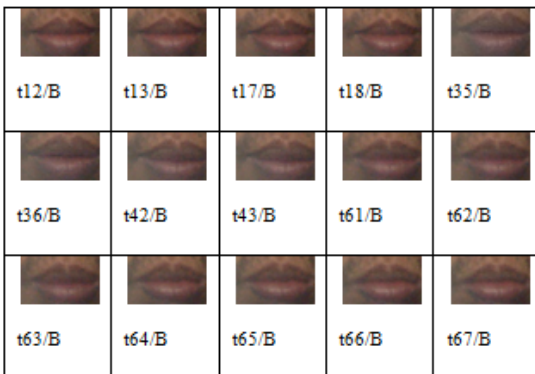
Gambar . 8 . Image frame hasil ekstraksi dari rekaman video

Proses pengujian pengenalan pola gerak mulut dilakukan pada sejumlah sampel (75 image frame). Fitur mulut dalam bentuk guratan guratan tepi mulut diekstraksi dengan ukuran 66 x 41 pixel, kemudian dilakukan proses pengelompokan dengan menggunakan SOM atau kohenen map. Proses validasi dilakukan dengan membandingkan pengamatan dan penilaian secara visual, dalam sebuah *cluster* dibandingkan apabila terdapat image yang berbeda dengan image yang dominan dalam *cluster* tersebut, maka image tersebut dianggap salah dan sebaliknya dikatakan benar. Gambar 9,10,11,12,13 dan 14 menunjukkan hasil pengujian *cluster*, indek tn/ status, menunjukkan number *image frame* yang diuji dan statusnya(S: Salah,B: Benar) misal t10/S : *image frame* 10, status Salah. Dari hasil pengujian 75 *image frame* didapatkan 6 *cluster*.



Gambar. 9. Hasil Pengujian Cluster 1

Karakteristik *cluster 1* : Mulut Terbuka penuh ,Bagian Gigi Kelihatan Secara Jelas. Anggota *cluster* : 15, Tidak sesuai secara visual 2 buah yaitu *frame* 10 dan 11.



Gambar. 10. Hasil Pengujian Cluster 2

Karakteristik *cluster 2* adalahmulut tertutup normal , Anggota *cluster* 15, secara visual semua anggota *cluster* validasinya benar.



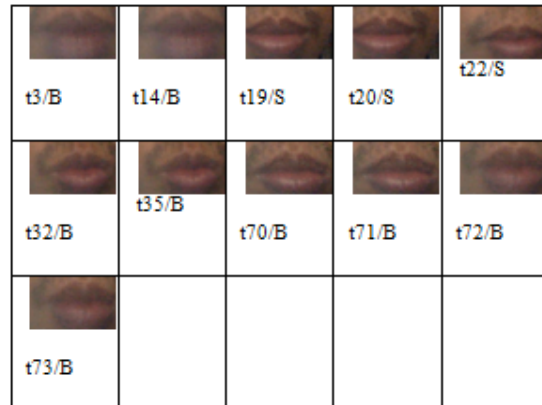
Gambar. 11. Hasil Pengujian Cluster 3

Karakteristik*cluster 2* menunjukkan mulut tertutup Normal, anggota *cluster* 5, validasi semua anggota *cluster* benar.



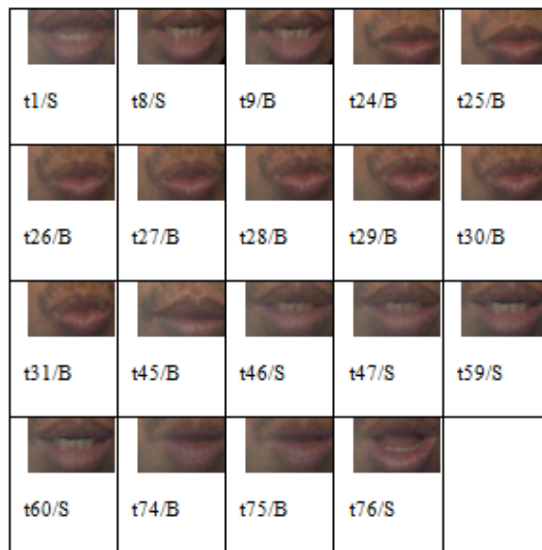
Gambar. 12. Hasil Pengujian Cluster 4

Karakteristik *cluster 4* adalah mulut tertutup normal, secara visual semua anggota *cluster* validasinya benar.



Gambar. 13. Hasil Pengujian Cluster 5

Karakteristik *cluster 5* adalah mulut dengan bibir atas tertarik kedalam dan bibir bawah keluar (nyebil) anggota *cluster* 11 . Validasi ditemukan 3 *image frame* yang tidak sesuai, yaitu *frame* 19,20,22 . *image frame* tersebut lebih sesuai dengan mulut yang tertutup normal.



Gambar. 14. Hasil Pengujian Cluster 6

Karakteristik *cluster 6* adalah mulut cenderung maju, anggota *cluster* : 19 , validasi salah :7 pada *frame* 1,8,46,47,59,60,76

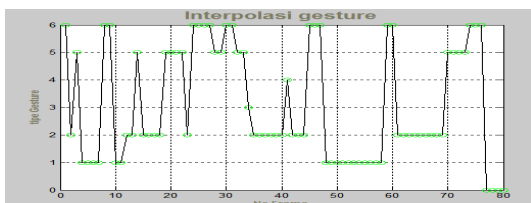
Dari 75 data *image frame* yang diambil didapatkan 13 *frame* yang tidak tepat letak klusternya.

Presentase kesalahan : $13/75 * 100\% = 17\%$

Presentase kebenaran :83 %

Berdasarkan hasil pengujian sebenarnya, *cluster 2,3,4* sebenarnya menunjukkan bentuk *gesture* mulut yang sama, mulut tertutup normal. Penyebab terjadinya *cluster* yang berbeda adalah karena potongan *crooping* yang berbeda, *cluster 2*

hasil area mulut yang *dicropping* menunjukkan hasil yang ideal, sedang pada *cluster* 3 ada daerah samping mulut yang terpotong lebih (hasil *cropping* cenderung ke kanan), dan pada *cluster* 4 area samping kiri cenderung lebih. Hal ini diakibatkan hasil deteksi tepi mulut yang tidak maksimal. Disekitar mulut terdapat titik titik *noise* yang terdeteksi sebagai tepi mulut, dan titik titik tersebut dijadikan acuan melakukan *cropping*. Sehingga hasil *cropping* pada beberapa *image frame* menjadi kurang maksimal. Permasalahan *noise* dalam deteksi tepi adalah memunculkan tepi salah, yaitu *noise* yang dikenali sebagai tepi [12]. Hasil visualisasi pola gerak bibir untuk tiap *image frame* ditunjukkan dalam grafik gambar 15.



Gambar 15. Visualisasi Gerak Bibir tiap Image Frame, Tipe Gesture 1: terbuka penuh, 2: normal (3,4 : normal, dalam visualisasi grafik dinyatakan dengan 2), 5 : nyebil , 6 : Mulut mulut dari setengah terbuka, sampai cenderung maju.

VII. SIMPULAN

Dari penelitian yang telah dilakukan dapat disimpulkan beberapa hal, pertama penggunaan SOM untuk pengenalan gerak mulut sangat ditentukan oleh keakuratan hasil segmentasinya. Seperti dalam percobaan ini karena ada pengaruh *noise* yang tersisa dalam segmentasi citra, akan mengakibatkan kesalahan pada proses pengklusteran oleh SOM. *Noise* tersebut menyebabkan timbulnya tepi salah, yaitu *noise* yang terdeteksi sebagai tepi. Hal ini akan mempengaruhi fitur guratan tepi yang dihasilkan, sehingga mempengaruhi hasil *cluster*. Secara umum pengenalan pola *gesture* (pola gerak/ mimik mulut) dengan menggunakan SOM dapat dikatakan berhasil, dari 75 sampel *image frame* yang diuji presentase keakuratannya berkisar 83%

Kedua kelemahan SOM dalam pengenalan pola adalah kurang *adaptif* terhadap perubahan jumlah sampel. Pada saat penambahan sejumlah sampel data baru maka SOM melakukan perhitungan dari awal kembali, seolah olah ini adalah ini adalah data baru. Sehingga secara komputasi kurang baik.

Ketiga berkaitan dengan proses pengenalan polanya, fitur pola guratan tepi pada area mulut dapat dijadikan salah satu alternatif untuk mengenali pola gerak/ mimik mulut. Pada prinsipnya fitur ini adalah fitur tepi, fitur ini cukup

sederhana dan mudah proses ekstraksinya maupun dalam komputasi pengenalan pola dibanding dengan fitur panjang/ lebar dan tinggi mulut (fitur geometri mulut).

UCAPAN TERIMA KASIH

Terima kasih yang sebesar-besarnya penulis sampaikan kepada semua pihak yang telah membantu penulisan jurnal ini, terutama kepada seluruh staf pengajar di program doctoral ilmu computer UGM, staf administrasi ELINS UGM, rekan rekan di FMIPA UNDIKSHA Singaraja Bali dan kepada teman teman di program S3 ilmu computer UGM yang selalu saling memberi semangat.

DAFTAR PUSTAKA

- [1] Gunadi Aris IG, Harjoko Agus, "Telaah Metode metode Pendeteksi kebohongan ", Indonesian Journal of Computing And Cybernetic Sistesms, Vol 6.No 2, pp 35 - 46, 2012.
- [2] L Catherine, Harris Caldwell, Ayase Aycicegi D, "Emotion and Lying in Non Native Language" , International Journal of Psychophysiology Vol 71, pp 193 -204 , September 2008.
- [3] Matsumoto D, Hwang H S, Skinner L, Frangk M G, "Evaluating Truthfulness and Detecting Deception" , FBI Law Enforcement Bulletin , vol 80, pp 1-25, 2011
- [4] Stephen Porter, Leanne Ten Brinke, 2008, "Reading Between The Lies: Identifying Concealed and Falsified in Universal Facial Expressions", Journal of Psychological Science, Vol 19 No 5, pp 508-514 , 2008.
- [5] Navaro Joe, Karlin Marvin, What Every Body is Saying , Harver Collin Publishe, Pymble Australia , 2008.
- [6] Oravec Milos, Kristof Branislav, Kolarik Michal, Pavlovicova J, "Extraction of Facial Features from Color Images", RadioEngineering, Vol 17, Hal 115 -120 , 2008.
- [7] Adipranata Rudy, G Cherry, Ongkodjojo Ronald P , "Fast Method for Multiple Human Face Segmentation in Color Image", Second International Conference on Future Generation Communication and Networking, 2008
- [8] .R Agustina Dewi, Karmilasari, Tanjung Fahmi, "Extraction Application of Face Feature/ Component ", 2nd Information and Communication Tecnology Seminar ,2006.
- [9] Jong Sek Siang, Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan Matlab , EdisiI, Andy Offset, Yogyakarta, 2005
- [10] Vesanto Juha, Johan Himberg, Alhoniemi Esa, Juha Parhankangas , " Selft Organizing Map in Matlab : The SOM Toolbox ", Proceedings of The Matlab DSP Confrence , Filandia, pp 35 -40 , November 1999.
- [11] Christopher B, Boehnen, B.S A Multi-Modal Approach to Frontal and Non Frontal Facial Feature, University of Notre Dame in Computer Science and Engineering Indiana, 2005.
- [12] Gunadi Aris IG, Wardoyo Retantyo, " Ekstraksi Tepi dengan Menggunakan Fuzzy Spatial Filtering dan Slicing Intensity ", Seminar Nasional Teknologi Informasi dan Aplikasinya, Denpasar Bali, pp 14-22, Oktober 2012.