

Pengamatan Ekspresi Wajah Secara Adaptif dengan Presentasi Pemutaran Musik

Agam Pamungkas Lumintan¹, Liliana², Henry Novianus Palit³
Program Studi Teknik Informatika, Fakultas Teknologi Industri, Universitas Kristen Petra
Jl. Siwalankerto 121-131 Surabaya 60236
Telp. (031) – 2983455, Fax. (031) - 8417658
E-mail: agamp94@gmail.com¹, lilian@petra.ac.id², hnpalit@petra.ac.id³

ABSTRAK

Tuntutan kehidupan yang tinggi di jaman sekarang ini membuat seorang individu mengalami perubahan kondisi emosional dengan mudah. Musik adalah salah satu media yang digunakan untuk menstabilkan kondisi emosional. Akan tetapi penggunaan musik yang salah dapat memperburuk kondisi emosional seseorang, seperti yang banyak terjadi di jaman ini.

Untuk mencegah hal tersebut, dibuatlah sebuah aplikasi yang mampu mendeteksi ekspresi pada wajah untuk menstabilkan kondisi emosional *user* melalui musik. Ekspresi pada wajah ditentukan melalui posisi fitur alis, mata, mulut, dan kerutan pada bagian tertentu. Posisi fitur yang terdeteksi akan dibandingkan perubahannya dengan posisi fitur pada wajah netral yang telah dikalibrasi sebelumnya. Perbandingan kedua posisi fitur akan dimasukkan ke dalam *neural network (backpropagation)* yang telah di-*training* sebelumnya untuk menentukan ekspresi pada wajah yang terdeteksi.

Melalui hasil pengujian, telah didapatkan akurasi pendeteksian fitur wajah dan ekspresi wajah yang terdeteksi. Ekspresi yang terdeteksi akan memicu pemutaran musik untuk menstabilisasi kondisi emosional *user*. Kelemahan dari aplikasi ini adalah *library* musik yang harus diatur secara manual oleh *user*.

Kata Kunci: Ekspresi wajah, psikologi, pendeteksian emosi, pemutar musik, dan *backpropagation*

ABSTRACT

High life demands nowadays make the individual emotional state change easily. Music is one of many media that is used to stabilize a person's emotional condition. However, wrong use of music can worsen a person's emotional condition, as in the case of today's society.

To prevent that, an application that can detect the user's facial expression to stabilize the user's emotional state through music is developed. The expression on the face is determined by the position of eyebrows, eyes, mouth, and wrinkles features. The positions of detected features are compared with those in a neutral face that has been previously calibrated. The difference of features' positions will be used as input of the trained neural network to determine the expression on the detected face.

Through some experiments, the accuracy of detection of facial features and facial expressions are known. The detected expression will trigger the playback of music to stabilize the user's emotional condition. The weakness of this application is that the music library must be set manually by the user.

Keywords: Facial expression, psychology, emotion detection, music player, and *backpropagation*

1. PENDAHULUAN

Wajah menjadi salah satu aspek komunikasi yang penting antar manusia. Wajah manusia dapat menyampaikan banyak informasi, salah satunya untuk mengekspresikan emosi[6]. Terkadang emosi yang dirasakan tidak diekspresikan secara verbal, disengaja atau tidak. Tetapi manusia pada umumnya selalu menunjukkan emosi yang sedang dirasakan melalui wajah, dengan atau tanpa disadari. Terkadang seseorang akan mencoba untuk menyembunyikan ekspresi yang tersampaikan di wajahnya, tetapi kebocoran ekspresi yang sedang dialami akan selalu ada[1][8]. Sudah banyak penelitian yang dilakukan tentang hubungan ekspresi wajah manusia dan ekspresi. Dalam penelitian tentang ekspresi yang sudah berkembang sampai saat ini, ditemukan bahwa setiap manusia memiliki kesamaan ciri pada bagian wajah tertentu dalam menunjukkan suatu ekspresi tanpa mempedulikan berbagai faktor, salah satunya budaya[1].

Dewasa ini, musik sudah menjadi salah satu aspek yang tidak dapat dilepaskan dari kehidupan sehari-hari. Musik merupakan karya seni yang memiliki nilai emosional tertentu di setiap liriknya. Dengan perkembangan teknologi saat ini, musik sangat mudah untuk didapatkan dan dimainkan melalui teknologi mumpuni yang ada saat ini. Menurut beberapa penelitian tentang relasi musik terhadap emosi, musik bisa mempengaruhi kondisi emosional pendengarnya melalui berbagai macam hal yang berhubungan dengan mekanisme psikologi[3][9][10].

Maka dari itu dibuatlah sebuah aplikasi yang mengumpulkan data berdasarkan pengamatan kondisi emosional pengguna. Kondisi emosional ditentukan berdasarkan pengumpulan ekspresi yang muncul pada wajah pengguna dalam rentang waktu tertentu. Ekspresi wajah pengguna akan ditentukan berdasarkan ciri tertentu pada beberapa bagian wajah, sesuai dengan ilmu psikologi dalam bidang ekspresi wajah. Data yang terkumpul akan menentukan pemilihan daftar musik yang akan diputar. Pemilihan musik yang sesuai akan meningkatkan emosi positif, sehingga kondisi emosional pendengarnya bisa stabil dan tingkat emosi yang bersifat negatif bisa diminimalisir.

2. TEORI PENUNJANG

Metode segmentasi bibir berdasarkan *Local Information* ini merupakan pengembangan dari metode Lip-Map[4]. Metode ini disebut dengan EnLip-Map. Metode ini lebih akurat karena menggunakan komponen saturasi dari HSI *color space* untuk bisa lebih membedakan antara piksel kulit dengan piksel bibir. Metode ini juga lebih cepat karena menghilangkan bagian atas dari gambar *input*.

Perhitungan EnLip-Map menggunakan persamaan sebagai berikut:

$$EnLip - Map = S \times \left\{ Cr^2 \times (Cr^2 - \eta \left(\frac{Cr}{Cb} \right)^2) \right\} \quad (1)$$

$$\eta = 0.95 \frac{\sum Cr^2}{\sum Cr/Cb} \quad (2)$$

Variabel Cr^2 dan Cr/Cb dinormalisasi pada jarak [0, 255]. Hasil dari persamaan Lip-Map juga dinormalisasi pada jarak [0, 255]. Parameter η didefinisikan sebagai rasio dari rata-rata Cr^2 dengan rata-rata Cr/Cb . Variabel S adalah komponen saturasi pada *color space* HSI.

Mata merupakan salah satu fitur yang sangat penting dalam mendeteksi emosi pada wajah, karena mata merupakan fitur permanen yang memiliki perpindahan posisi yang relatif kecil. Maka dari itu seringkali posisi mata pada wajah digunakan untuk mencari posisi fitur lain yang ada di wajah, sehingga posisi mata yang akurat merupakan hal yang krusial. Metode ini dipilih karena mampu memberikan hasil yang baik, dalam spektrum warna RGB maupun *grayscale*. Untuk mengurangi kesalahan pendeteksian dan lama penghitungan, hanya daerah wajah tertentu saja yang akan dideteksi. Penghitungan ROI (*Region of interest*) daerah mata dilakukan dengan metode *anthropometric*, dengan memperkirakan posisi mata pada wajah manusia pada umumnya[7]. Persamaan yang digunakan untuk menghitung ROI dapat dilihat sebagai berikut:

$$R_l = \begin{pmatrix} I(x_c - w_e, y_c - h_e) & \dots & I(x_c, y_c - h_e) \\ \vdots & \ddots & \vdots \\ I(x_c - w_e, y_c) & \dots & I(x_c, y_c) \end{pmatrix} \quad (3)$$

$$w_e = \alpha w_f, \quad h_e = \beta h_f \quad (4)$$

Nilai α yang digunakan pada metode ini adalah 2/5 dan β adalah 1/6, nilai ini dipilih karena merupakan nilai paling minimum untuk menampilkan daerah mata pada *face database* percobaan. Variabel x_c dan y_c merupakan titik tengah dari wajah yang terdeteksi, sedangkan h_f dan w_f merupakan tinggi dan lebar dari wajah yang terdeteksi. Untuk memperjelas perbedaan intensitas antar piksel, dilakukan proses eksponensial menggunakan Persamaan 5. Metode deteksi alis yang digunakan sama dengan metode pendeteksian mata, perbedaan terdapat pada nilai α yang digunakan pada metode ini adalah 3/7 dan $\beta = 1/5$. Perbedaan juga terdapat pada perhitungan ROI dimana y_c diganti menjadi y'_c yang nilainya didapatkan melalui Persamaan 6[7].

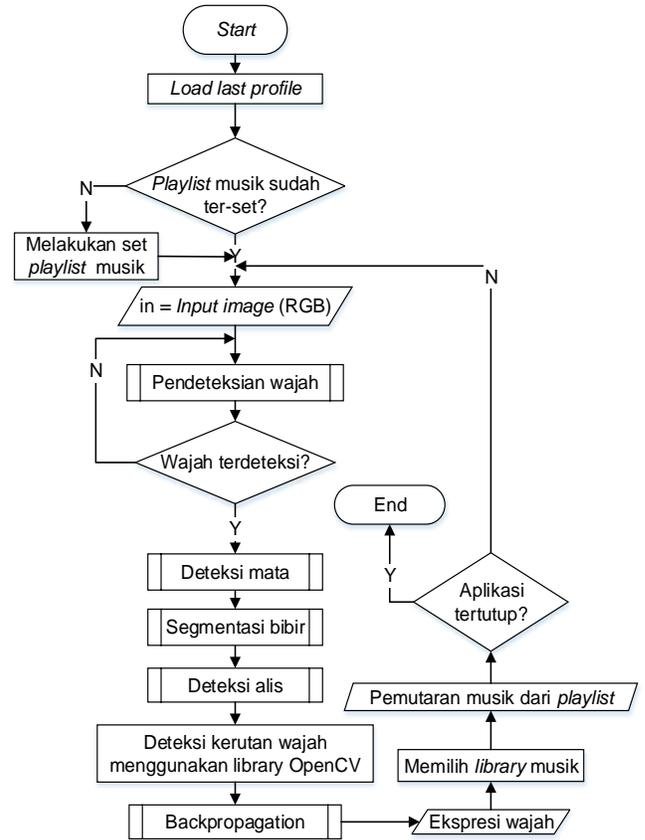
$$I'_{i,j} = \exp[I_{i,j} \cdot k] \quad (5)$$

$$y'_c = y_c + \frac{h_e}{2} \quad (6)$$

Nilai fitur yang didapatkan akan dibandingkan dengan nilai fitur pada wajah netral[12]. Selisih dari nilai kedua fitur ini akan digunakan sebagai *input* pada *neural network backpropagation*[2].

3. DESAIN SISTEM

Dalam sistem perangkat lunak ini memiliki beberapa tahapan langkah-langkah yang harus dilakukan untuk mendeteksi ekspresi wajah pada gambar. Langkah-langkah tersebut diantaranya adalah deteksi mata, bibir, alis, kerutan, dan ekspresi. Seperti yang ditunjukkan pada Gambar 1.

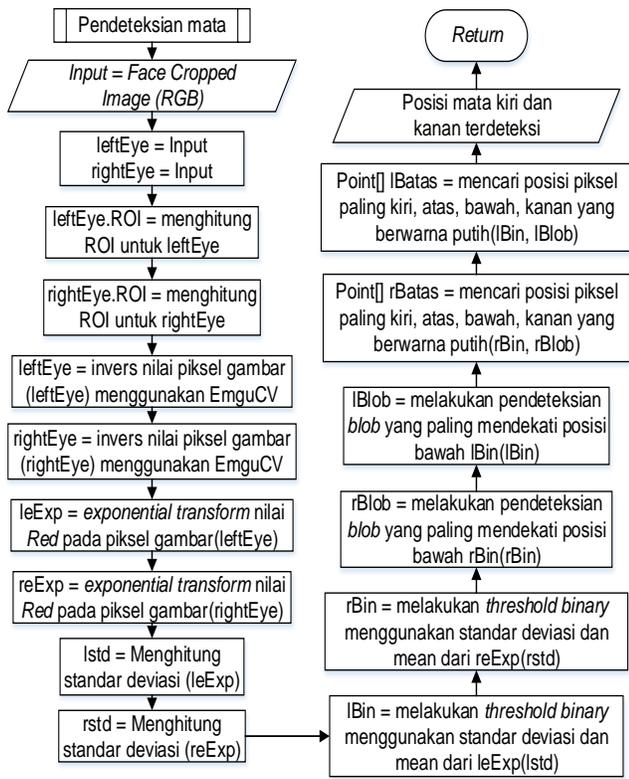


Gambar 1. Flowchart Sistem Secara Garis Besar

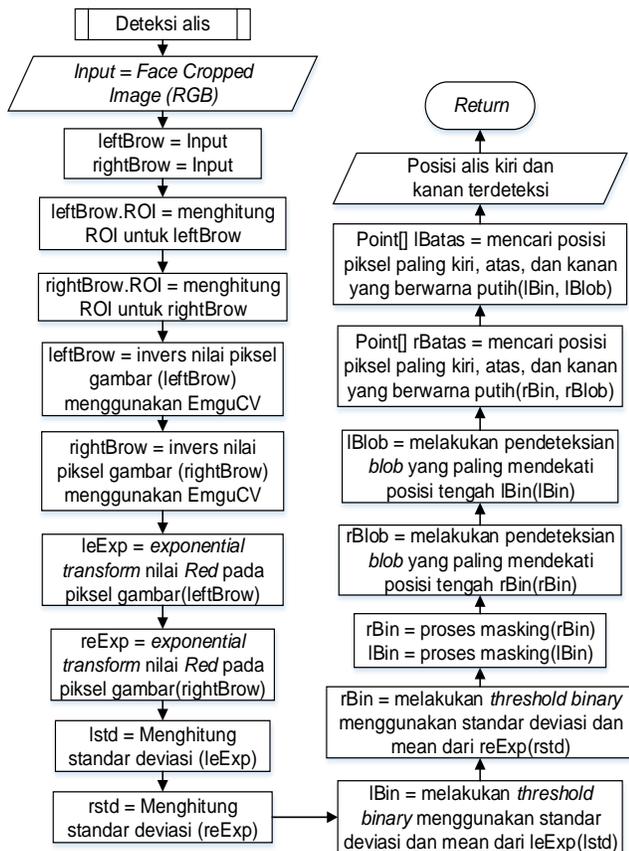
Sebelum melakukan proses *backpropagation*, diperlukan proses ekstraksi fitur sebagai *input* pada *backpropagation*. Fitur yang diekstraksi yaitu fitur mata, alis, dan bibir. *Input* dari proses ekstraksi fitur mata, bibir dan alis adalah wajah yang telah terdeteksi menggunakan metode Viola-Jones dengan *skin pixel classifier*[11]. Flowchart dari setiap proses ekstraksi dapat dilihat pada Gambar 2 sampai Gambar 4.

Variabel *input* pada Gambar 2 berisi gambar wajah yang telah di-*crop* sebelumnya. Perhitungan ROI disesuaikan dengan teknik *anthropometric* untuk menentukan posisi daerah mata pada wajah. Proses invers pada nilai piksel dilakukan agar nilai spektrum warna merah pada piksel mata yang rendah menjadi tinggi. Pencarian posisi piksel paling atas dan bawah pada mata menggunakan persamaan gradien garis.

Variabel *input* pada Gambar 3 berisi gambar wajah yang telah di-*crop* sebelumnya. Perhitungan ROI disesuaikan dengan teknik *anthropometric* untuk menentukan posisi daerah alis pada wajah. Proses invers pada nilai piksel dilakukan agar nilai spektrum warna merah pada piksel mata yang rendah menjadi tinggi. Pencarian posisi piksel paling atas pada mata menggunakan persamaan gradien garis. Proses *masking* dari metode alis adalah menghilangkan bagian mata, posisi mata didapatkan dari metode ekstraksi fitur mata.



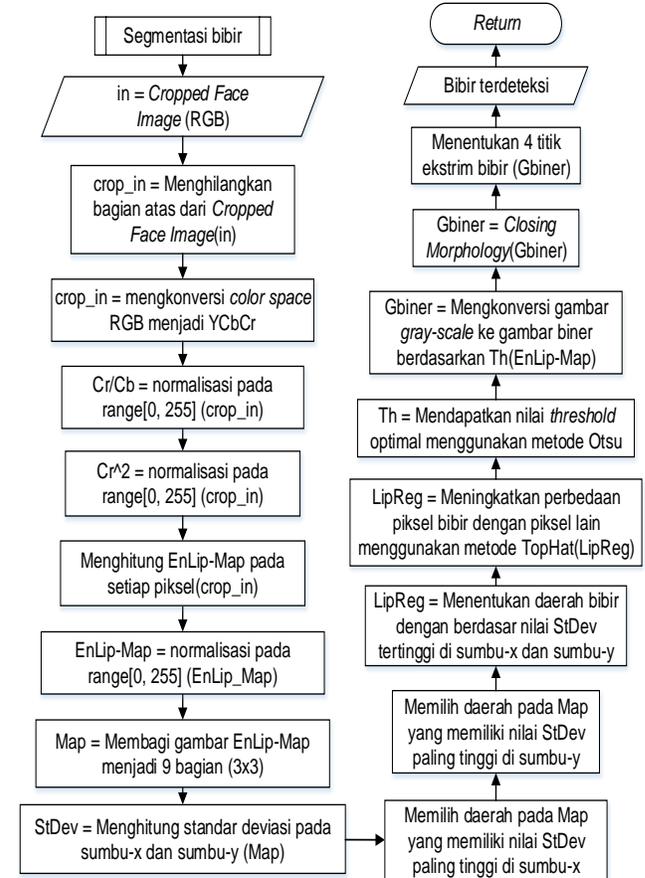
Gambar 2. Flowchart metode ekstraksi fitur mata



Gambar 3. Flowchart metode ekstraksi fitur alis

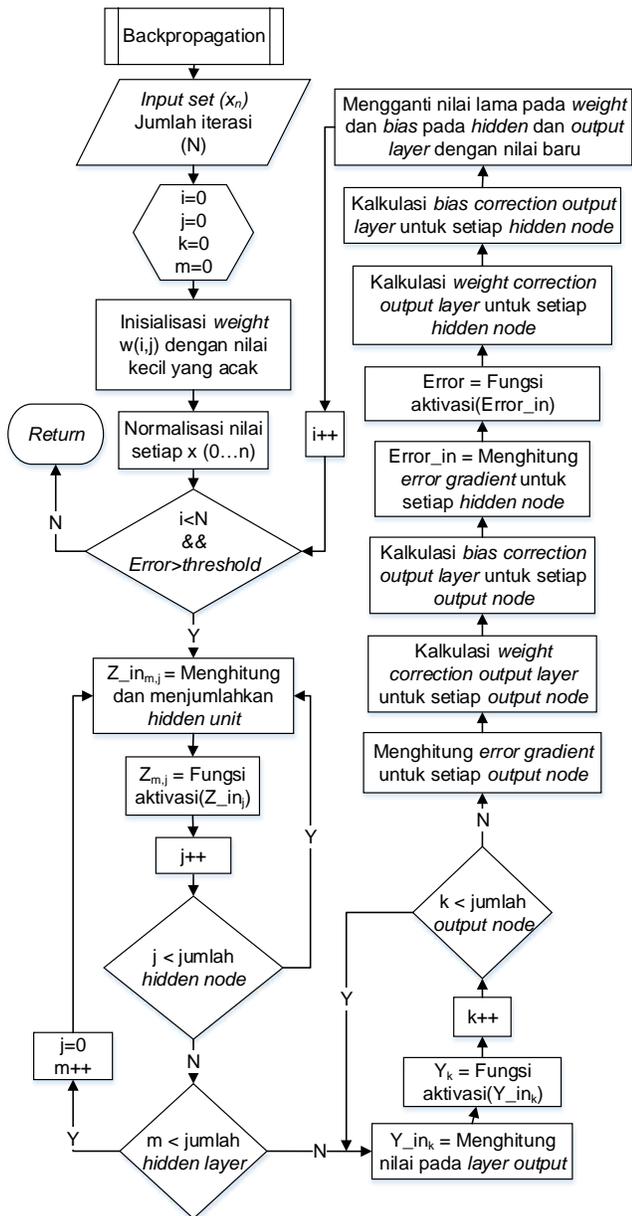
Pada Gambar 4, menghilangkan bagian atas pada gambar *input* dari metode ekstraksi fitur bibir dilakukan untuk menghemat waktu, karena bibir pasti berada pada bagian bawah wajah. Pembagian gambar wajah menjadi 9 bagian ditujukan untuk memetakan bibir ke ROI yang lebih kecil. Bagian yang terpilih akan diperbesar luasnya dengan mengambil setengah daerah bagian yang berada di sekitarnya.

Setelah melakukan proses ekstraksi fitur, nilai dari fitur-fitur yang terdeteksi dijadikan *input* pada *backpropagation*[4]. Nilai-nilai dari fitur akan diproses dengan membandingkan wajah netral seseorang dengan wajah yang terdeteksi saat itu juga. Proses pendeteksian ekspresi dilakukan pada *background* aplikasi. Diagram alir dari *backpropagation* dapat dilihat pada Gambar 5.



Gambar 4. Flowchart metode ekstraksi fitur bibir

Gambar 5 menunjukkan alur kerja dari metode *backpropagation*, metode ini digunakan untuk memprediksi ekspresi dari gambar yang telah diekstraksi fiturnya. Tetapi sebelum bisa melakukan prediksi terhadap wajah, perlu dilakukan proses *training* terlebih dahulu. Proses *training* ini dimaksudkan agar *backpropagation* dapat belajar berdasarkan data *training* yang telah disiapkan terlebih dahulu. Untuk menentukan jumlah *hidden layer* dan *hidden node* diperlukan percobaan untuk mencari kombinasi *layer* dan *node* yang memberikan hasil terbaik.



Gambar 5. Flowchart backpropagation

4. PENGUJIAN

Pengujian terhadap aplikasi dibagi menjadi beberapa bagian, gambar yang digunakan dalam pengujian diambil dari database wajah KDEP[5]. Pengujian yang dilakukan diantaranya adalah sebagai berikut:

- Pengujian metode deteksi mata untuk melihat tingkat kesuksesan dan mencoba meningkatkan tingkat kesuksesan dari metode yang digunakan. Pengujian dilakukan pada 485 gambar.

Tabel 1. Hasil pengujian deteksi mata terhadap nilai β

Nilai β	Jumlah kegagalan menentukan ROI	Tingkat kesuksesan
1/6	68 gambar	269 gambar
11/60	19 gambar	229 gambar
1/5	7 gambar	222 gambar
1/4	1 gambar	214 gambar

Seperti yang dapat dilihat pada Tabel 1, nilai β berpengaruh terhadap tinggi ROI mata. Semakin besar nilai β , maka semakin sedikit mata yang terpotong saat menentukan ROI. Konsekuensi dari ROI yang semakin tinggi adalah semakin besar kemungkinan benda selain mata yang berada dalam ROI mata. Hal ini mempengaruhi nilai *threshold* pada proses binerisasi. Sehingga berdasarkan pengujian, didapatkan bahwa nilai β yang memberikan hasil terbaik adalah 1/6. Dari pengujian pada 485 gambar, didapat tingkat akurasi metode pendeteksian mata yang digunakan adalah 55.46%. Contoh hasil pendeteksian mata yang sukses dapat dilihat pada Gambar 6.



Gambar 6. Hasil pendeteksian mata yang berhasil

- Pengujian metode deteksi alis untuk melihat tingkat kesuksesan alis. Pengujian dilakukan pada 485 gambar.

Tabel 2. Hasil pengujian deteksi alis

Gambar input	Jumlah	Sukses	Gagal
Sukses deteksi mata	269 gambar	170 gambar	99 gambar
Gagal deteksi mata	216 gambar	108 gambar	108 gambar

Pengujian ini dipisah menjadi 2 segmen dikarenakan pada deteksi alis, terdapat proses *masking* yang memerlukan posisi mata. Sehingga posisi mata yang tidak sesuai akan mempengaruhi pendeteksian alis. Contoh hasil pendeteksian alis dapat dilihat pada Gambar 7.



Gambar 7. Hasil pendeteksian alis yang berhasil

- Pengujian metode deteksi bibir untuk melihat tingkat kesuksesan dan mencoba meningkatkan tingkat kesuksesan dari metode yang digunakan. Pengujian dilakukan pada 485 gambar. Pengujian yang dilakukan yaitu mengganti metode pemilihan kontur bibir dengan memilih 2 kontur terlebar. Metode yang digunakan sebelumnya adalah metode *thresholding* untuk menghilangkan *noise* yang ada pada ROI bibir sehingga hanya kontur bibir saja yang tersisa, namun cara ini tidak efektif karena *noise* tidak dapat tersaring dengan baik. Dengan menggunakan metode yang baru, tingkat kesuksesan metode ini naik menjadi 47.21% yang sebelumnya hanya 28.24%. Dilakukan percobaan untuk menambahkan proses morfologi *closing* karena saat pengkonversian ke gambar biner, bibir atas cenderung terdeteksi menjadi beberapa bagian-bagian kecil. Sehingga mengurangi lebar dari kontur bibir atas, memungkinkan terpilihnya *noise* yang memiliki lebar lebih dari bibir atas. Adapun hasil pengujian dapat dilihat pada Tabel 3.

Tabel 3. Hasil pengujian deteksi bibir terhadap proses morfologi *closing*

Jumlah iterasi	Jumlah sukses
0	229
1	229
2	208
3	194

Berdasarkan hasil pengujian pada Tabel 3, didapati bahwa proses *closing morphology* sebanyak 1 iterasi memiliki tingkat kegagalan yang setara dengan 0 iterasi. Untuk mengantisipasi pecahan pada bibir yang dideteksi, diputuskan untuk melakukan proses *closing morphology* sebanyak 1 kali. Dapat dilihat dengan jumlah iterasi yang lebih besar dari 1 iterasi, jumlah kegagalan deteksi bibir semakin bertambah, salah satu penyebabnya adalah bibir atas dan bibir bawah yang menjadi satu kontur.

- Pengujian metode *backpropagation* untuk mencari kombinasi jumlah *hidden layer* dan *hidden node* yang memberikan hasil terbaik. Gambar wajah yang digunakan sebagai *input training* pada *backpropagation* memiliki kriteria tertentu, yaitu:

1. Gambar yang dipilih minimal harus lulus kriteria 2 metode ekstraksi fitur
2. 1 ekspresi dari gambar wajah diperbolehkan untuk lulus kriteria 1 metode ekstraksi fitur saja.

3. Gambar wajah memiliki minimal 6 ekspresi yang telah lulus kriteria dengan syarat yang disebutkan di poin pertama (Contoh: Wajah A terdiri dari 7 ekspresi wajah, 6 diantaranya memiliki hasil ekstraksi fitur yang lulus 2 metode ekstraksi fitur atau lebih).

Terdapat 269 gambar yang memenuhi kriteria diatas namun hanya 91 gambar yang terpilih sebagai *input backpropagation*, dikarenakan kriteria yang ditetapkan pada *input backpropagation* pada poin ke-3. Pengujian *backpropagation* dilakukan terhadap 12 kombinasi jumlah *layer* dan jumlah *node* yang bervariasi, yang memiliki nilai *error rate* yang berbeda-beda. *Error rate* digunakan sebagai tolak ukur hasil belajar daripada *backpropagation*, semakin rendah nilai *error rate* semakin baik pula hasil belajar *backpropagation*. Hasil pengujian dapat dilihat pada Tabel 4.

Tabel 4. Hasil pengujian *backpropagation* terhadap jumlah *hidden layer* dan *node* yang variatif

<i>Error Rate</i>	Jumlah <i>Layer</i>		
Jumlah <i>Node</i>	2	4	6
20	3.00009	5.00014	78.00045
30	2.00007	1.00030	78.00045
50	4.00005	1.00027	78.00045
70	5.00004	2.00023	78.00045

Dari hasil pengujian, dipilih 6 kombinasi jumlah *layer* dan jumlah *node* yang memiliki *error rate* yang paling kecil dibandingkan kombinasi yang lain. Kemudian 6 kombinasi yang terpilih diujicoba terhadap 91 gambar wajah yang digunakan dalam proses *training* dan 117 gambar wajah yang tidak digunakan dalam proses *training* yang berasal dari kandidat *input backpropagation* yang tidak terpilih. Pengujian ini dilakukan untuk mengukur tingkat akurasi *backpropagation* yang digunakan. Nilai learning rate(α) yang digunakan adalah 0.4 dengan maksimum 200000 iterasi. Adapun hasil pengujian dapat dilihat pada Tabel 5.

Tabel 5. Hasil Pengujian Akurasi *Backpropagation* Terhadap Jumlah *Layer* dan *Node* yang Variatif

Jumlah <i>Layer</i>	Jumlah <i>Node</i>	<i>Error Rate</i>	Akurasi	
			Data <i>training</i>	Data bukan <i>training</i>
2	20	3.00009	93.41%	58.97%
2	30	2.00007	79.12%	54.70%
2	50	4.00005	80.22%	32.48%
4	30	1.0003	20.88%	17.95%
4	50	1.00027	48.35%	23.93%
4	70	2.00023	72.53%	52.99%

Dari hasil pengujian yang dilihat pada Tabel 5, yang memberikan tingkat akurasi emosi yang tertinggi adalah *backpropagation* dengan jumlah *layer* 2 dan jumlah *node* 20. Maka dari itu berdasarkan data yang telah didapatkan melalui pengujian, aplikasi ini menggunakan *backpropagation* dengan 2 *hidden layer* dan 20 *node*.

5. KESIMPULAN

Pada jurnal ini telah dibahas metode ekstraksi fitur yang digunakan. Pada metode deteksi mata, ditemukan *blob* kelopak mata dan *iris* yang terpisah, dan luas ROI yang tidak dapat mencakupi keseluruhan bagian mata. Maka dilakukan proses morfologi *closing* untuk menyatukan kedua *blob* yang terpisah dan penggantian nilai konstanta β . Tetapi kedua proses tersebut tidak digunakan karena dengan penambahan kedua proses tersebut tingkat kesuksesan deteksi mata berkurang. Pengamatan terhadap hasil deteksi bibir menunjukkan ciri kontur bibir yang lebih lebar dibandingkan dengan kontur lain yang terdapat pada ROI, maka dari itu dilakukan penggantian metode pemilihan kontur dengan memilih 2 kontur terlebar. Dengan penggantian metode pemilihan kontur, terbukti akurasi dari deteksi alis dari 28.24% meningkat menjadi 47.21%.

Backpropagation pada sistem ini menggunakan 2 *hidden layer* dan 20 *hidden node* karena memiliki hasil belajar yang paling baik diantara beberapa percobaan kombinasi *hidden layer* dan *hidden node* yang telah dibahas pada jurnal ini. Hasil belajar diukur melalui tingkat akurasi pendeteksian terhadap data yang digunakan dan data yang tidak digunakan saat proses *training*.

6. REFERENSI

- [1] Ekman, P., & Friesen, V. W. (2003). *Unmasking The Face : A guide to recognizing emotions from facial expressions*. San Fransisco: Malor Books.
- [2] Fausett, L. (1994). *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Prentice-Hall.
- [3] Juslin, P. N., & Vastfjall, D. (2008). Behavioral and Brain Sciences 31. *Emotional Responses to Music: The Need to Consider Underlying Mechanism*, 559-621.
- [4] Kalbkhani, H., & Amirani, M. C. (2012). An Efficient Algorithm for Lip Segmentation in Color Face Images Based on Local Information. In JWEET, *Journal of World's Electrical Engineering and Technology 1(1): 12-16-2012* (pp. 12-15). Urmia, Iran: Scienceline Publication.
- [5] Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.
- [6] Martinez, A., & Du, S. (2012). Journal of Machine Learning Research 13. *A Model of the Perception of Facial Expressions of Emotion by Humans: Research Overview and Perspectives*, 1589-1608.
- [7] Moreira, J. L., Braun, A., & Musse, R. S. (2010). Graphics, Patterns and Images. *SIBGRAPI '10 Proceedings of the 2010 23rd SIBGRAPI Conference* (pp. 17-24). Washington: IEEE Computer Society.
- [8] Porter, S., Brinke, L. t., & Wallace, B. (2012). J Nonverbal Behav. *Secrets and Lies: Involuntary Leakage in Deceptive Facial Expressions as a Function of Emotional Intensity*, 23-37.
- [9] Rea, C., MacDonald, P., & Carnes, G. (2010). Emporia State Research Studies, vol. 46, no. 1. *Listening to classical, pop, and metal music: An investigation of mood*, 1-3.
- [10] Scherer, K. R., & Zentner, M. (2001). Emotional Effects of Music: Production Rules. In P. Juslin, & J. Sloboda, *Music and emotion: theory and research* (pp. 366-371). New York: Oxford University Press.
- [11] Tabataie, Z. S., Rahmat, R. W., Udzir, N. I., & Kheirkhah, E. (2009). IJCSNS Internation Journal of Computer Science and Network Security, Vol. 9 No. 5. *A Hybrid Face Detection System using Combination of Appearance-based and Feature-based methods*, 181-183.
- [12] Tian, Y.-l., Kanade, T., & Cohn, J. F. (2001). Recognizing Action Units for Facial Expression Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 2*, 11-12.