

# HISTOGRAM OF ORIENTED GRADIENT UNTUK DETEKSI EKSPRESI WAJAH MANUSIA

Marlinda Vasty Overbeek

Program studi Informatika, Fakultas Teknik dan Informatika  
Universitas Multimedia Nusantara  
Tangerang, Banten  
[marlinda.vasty@umn.ac.id](mailto:marlinda.vasty@umn.ac.id)

## ABSTRACT

*This research focuses on the detection of human facial expressions using the Histogram of Oriented Gradient algorithm. Whereas for the classification algorithm, Convolutional Neural Network is used. Image data used in the form of seven different expressions of humans with the extraction of 48x48 pixels. The use of Histogram of Oriented Gradient as a feature extracting algorithm, because Histogram of Oriented Gradient is good to be used in detecting moving objects. Whereas Convolutional Neural Network is used because it is an improvement of the Multi Layer Perceptron algorithm. Of the three epoches done, it produced the best accuracy of 77% re-introduction of human facial expressions. These results are quite convincing because it only uses three epochs.*

**Keywords:** expression detection, Histogram of Oriented Gradient, Convolutional Neural Network

## 1. PENDAHULUAN

Sebagai insan yang memiliki jiwa sosial, manusia bersosialisasi secara verbal dan non-verbal. Salah satu bentuk bahasa non – verbal adalah dengan menggunakan ekspresi wajah [1]. Dengan ekspresi wajah, maka bisa menyampaikan apa saja yang ada didalam pikiran mereka dan dapat mengenai emosi mereka. Tetapi ketika emosi mencapai intensitas tertentu maka seluk beluk ekspresi wajah menjadi hilang.

Orang berpendapat jika bahasa tubuh sama dengan ekspresi wajah, namun hal ini adalah sesuatu yang sangat berbeda jauh. Dengan ekspresi wajah, bisa menyampaikan apa yang ada di pikiran orang, dan dapat dikenali emosi mereka. Ekspresi wajah memiliki perbedaan dengan bahasa tubuh. Bahasa tubuh memiliki akurasi yang lebih baik untuk mendeteksi suasana hati dibandingkan dengan ekspresi wajah. Ekspresi wajah seringkali bersifat lebih ambigu dibandingkan dengan bahasa tubuh, sehingga diperlukan adanya suatu cara untuk memahami ekspresi wajah manusia.

Cara untuk bisa melakukan pendeteksian ekspresi wajah manusia adalah dengan teknologi pengolahan citra. Penelitian di bidang ini semakin banyak dilakukan, beberapa diantaranya melakukan penelitian untuk melakukan pengenalan wajah, tanpa peduli dari suku apa, seberapa sehat, seberapa tua, dan juga mengembangkan penelitian tentang ekspresi wajah, apakah sedih, normal atau sedang tertawa.

Pada studi yang dilakukan oleh [1],[2], didapatkan bahwa ekspresi wajah dapat menyampaikan hingga 55% informasi dalam interaksi, dibandingkan dengan vocal dan verbal keseluruhan yang hanya dapat memberikan informasi hingga sebesar 45%. Ekspresi wajah dapat digunakan untuk mengembangkan interaksi yang natural antara manusia dan computer. Beberapa penelitian tentang pengenalan ekspresi manusia sudah dilakukan, antaranya adalah Lekshmi dan Sasikumar [3] menggunakan Gabor dan Support Vector Machine, penelitian ekspresi manusia juga dilakukan oleh Haghghat *et al* [4] dengan menggunakan *cloud based* sebagai pengenalan biometrik wajah manusia.

Alwakeel dan Shaaban [5] menggunakan Haar Wavelet transform dan *Principal Component Analysis* baru di klasifikasikan dengan Levenberg-Marquardt backpropagation neural network. Sedangkan Goyani dan Patel [6], menggunakan Multi – Level Haar Wavelet dan Logistic Regression untuk melakukan pengenalan ekspresi wajah. Pada penelitian ini dilakukan dengan memfokuskan pada pengenalan ekspresi wajah dari citra digital yang telah diekstraksi beberapa fiturnya.

Penelitian ini menggunakan metode Histogram of Oriented Gradient (HOG). Hal ini dikarenakan HOG merupakan bagian dari computer vision yang cepat untuk mendeteksi perubahan bentuk secara cepat. Misalnya gambar yang berasal dari kamera pengawasan atau CCTV sehingga ekspresi manusia yang cepat berubah bisa dengan ditangkap oleh algoritma HOG. Contohnya adalah HOG digunakan pada daerah bencana. HOG mampu mendeteksi posisi tubuh manusia, apakah sedang berjalan dengan posisi tegak atau berada pada posisi di bawah rotasi planar yang berbeda dan karena itu diperlukan rotasi-invariant [7]. Selain itu kombinasi antara HOG dan Support Vector Machine juga baik untuk mendeteksi objek manusia. Segmentasi yang dilakukan sudah cukup baik dan menghasilkan akurasi sebesar 89.59% [8].

Berdasarkan kelebihan tersebut, maka pada penelitian tersebut digunakan HOG sebagai pendeteksi ekspresi manusia sebagai teknik pemecahan masalah. Teknik yang dilakukan adalah melakukan praproses citra dengan membuat citra menjadi hitam putih lalu mendeteksi bagian tepi dari citra. Setelah mendapatkan citra hasil praproses tersebut,, maka citra akan diklasifikasikan dengan Convolutional Neural Network (CNN).

## 2. TINJAUAN PUSTAKA

### Histogram of Oriented Gradient (HOG)

HOG merupakan salah satu bagian dari computer vision yang cocok digunakan sebagai descriptor untuk pencarian gambar berbasis sketsa [9]. HOG adalah descriptor berbasis window yang mendeteksi pada titik interest. Metode ini menghitung nilai gradient dalam daerah tertentu pada suatu citra. Setiap citra memiliki karakteristik yang ditunjukkan oleh

distribusi gradient yang diperoleh dengan membagi citra ke dalam daerah kecil yang disebut dengan *cell*. Tiap *cell* tersebut disusun dari sebuah histogram dari sebuah gradient. Kombinasi dari histogram ini yang dijadikan sebagai descriptor yang mewakili sebuah objek [10]. Langkah kerja pada HOG adalah sebagai berikut [10].

1. Lakukan *input* gambar yang hendak digunakan
2. Hitung gradien citra  
Penghitungan gradien citra menggunakan persamaan (1) :

$$|G| = \sqrt{I_x^2 + I_y^2} \dots\dots(1)$$

Dimana  $I$  adalah citra *graylevel*.  $I_x$  merupakan matriks terhadap sumbu  $x$  dan  $I_y$  merupakan matriks terhadap sumbu  $y$ .  $I_x$  dan  $I_y$  dapat dihitung dengan persamaan (2)

$$I_x = I * D_x, I_y = I * D_y \dots\dots(2)$$

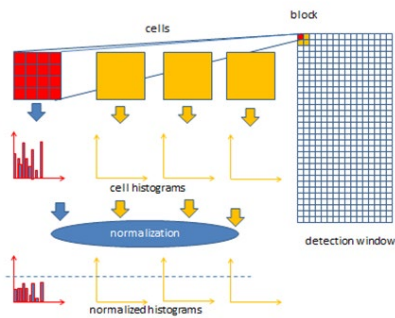
$D_x$  adalah mask  $[-1 \ 0 \ 1]$ , sedangkan

$$D_y \text{ adalah mask } \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Masing-masing dihitung dengan cara konvolusi. Kemudian gradien ditransformasikan kedalam koordinat sumbu dengan sudut di antara  $0$  sampai  $180^0$  yang disebut sebagai orientasi gradien. Orientasi gradien ( $\theta$ ) dihitung dengan persamaan (3).

$$\theta = \arctan \left( \frac{I_x}{I_y} \right) \dots\dots(3)$$

3. Hitung histogram dari *orient* gradien tiap *cell*  
Setiap piksel dalam sebuah *cell* mempunyai nilai histogram sendiri-sendiri berdasarkan nilai yang dihasilkan dalam perhitungan gradien yang kemudian dilakukan normalisasi pada setiap blok *cell*. *Cell* memiliki ukuran  $8 \times 8$  piksel pada sebuah citra. sedangkan blok memiliki ukuran  $2 \times 2$  *cell*. Gambaran *cell* dan blok ditampilkan pada Gambar 1.



Gambar 1 : *Cell* yang menyusun sebuah blok

- Lakukan normalisasi pada fitur blok  
Nilai normalisasi fitur blok didapat dari persamaan (4). Fitur blok dinormalisasikan untuk mengurangi efek perubahan kecerahan objek pada satu blok. Variabel  $b$  merupakan nilai blok fitur dan variabel  $e$  merupakan bilangan positif yang bernilai kecil untuk mencegah pembagian dengan nol.

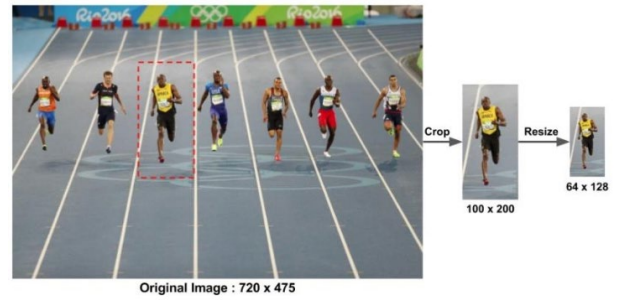
$$b = \frac{b}{\sqrt{b^2 + e}} \dots(4)$$

- Lakukan normalisasi pada fitur HOG  
Nilai dari normalisasi tiap blok ini digabungkan menjadi 1 vektor fitur HOG. Kemudian fitur vector HOG dilakukan normalisasi. Normalisasi dilakukan menggunakan persamaan (5). Variabel  $h$  merupakan fitur HOG dan variabel  $e$  merupakan bilangan positif yang bernilai kecil untuk mencegah pembagian dengan nol.

$$h = \frac{h}{\sqrt{\|h\|^2 + e}} \dots(5)$$

### Fitur Deskriptor

Fitur deskriptor adalah representasi dari gambar atau gambar *patch* yang menyederhanakan gambar dengan mengekstraksi informasi yang berguna dan membuang informasi asing. Biasanya, deskripsi fitur mengkonversi gambar ukuran lebar x tinggi x 3 *channel* ke saluran vektor fitur atau larik panjang  $n$ . Misalnya dalam kasus deskripsi fitur HOG, gambar masukkan jika berukuran 64 x 128 x 3, maka fitur keluaran adalah panjang 3780. Contoh dari fitur descriptor dengan HOG digambarkan pada Gambar 2.



Gambar 2 : Fitur descriptor dengan HOG

### Convolution Neural Network (CNN)

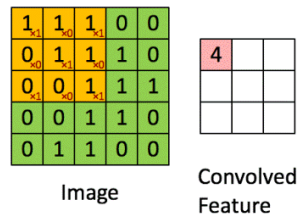
CNN adalah pengembangan dari Multilayer Perceptron (MLP) yang didesain untuk mengolah data 2 dimensi. CNN termasuk dalam jenis Deep Neural Network karena kedalaman jaringan yang tinggi dan banyak diaplikasikan pada data citra. pada kasus klasifikasi citra, MLP kurang sesuai untuk digunakan karena tidak menyimpan informasi spasial dari data citra dan menganggap setiap piksel adalah fitur yang independen sehingga menghasilkan hasil yang kurang baik[11].

Arsitektur jaringan pada CNN juga terdiri dari beberapa lapisan, tidak seperti pada jaringan syaraf tiruan pada dasarnya yang hanya memiliki layer dan beberapa neuron pada masing-masing layer. Layer pada CNN tersebut adalah sebagai berikut [11]:

1. Convolution layer

Layer ini melakukan operasi konvolusi pada keluaran dari layer sebelumnya. Layer tersebut adalah proses utama yang mendasari sebuah CNN. Konvolusi adalah suatu istilah matematika yang berarti mengaplikasikan sebuah fungsi pada keluaran fungsi lain secara berulang. Dalam pengolahan citra, konvolusi berarti mengaplikasikan sebuah kernel (ditunjukkan oleh kotak kuning) pada citra disemua offset yang memungkinkan seperti yang ditunjukkan pada Gambar 3. Kotak hijau secara keseluruhan adalah citra yang akan di konvolusi. Kernel bergerak dari sudut kiri atas ke kanan bawah. Sehingga hasil konvolusi dari citra dapat dilihat pada gambar sebelah kanan. Tujuan dilakukan hal ini adalah mengekstraksi fitur dari citra masukkan. Konvolusi akan

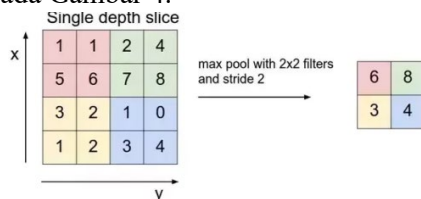
menghasilkan transformasi linear dari data masukkan sesuai dengan informasi spasial pada data. Bobot pada layer tersebut menspesifikasikan kernel konvolusi yang digunakan sehingga kernel konvolusi dapat dilatih berdasarkan masukkan pada CNN.



Gambar 3 : Konvolusi dengan citra 5x5x1 dengan kernel 3x3x1

## 2. Subsampling layer

Subsampling adalah proses mereduksi ukuran suatu data citra. Dalam pengolahan citra, subsampling juga bertujuan untuk meningkatkan invariansi posisi dari fitur. Dalam sebgaaian CNN, metode subsampling dilakukan dengan max pooling. Max pooling membagi keluaran dari convolution layer menjadi beberapa grid yang kecil lalu mengambil nilai maksimal dari setiap grid untuk menyusun matriks citra yang telah direduksi seperti yang ditunjukkan pada Gambar 4.



Gambar 4 : Operasi max pooling

## 3. Fully connected layer

Layer tersebut adalah layer yang biasanya digunakan pada penerapan MLP dan bertujuan untuk melakukan transformasi pada dimensi data agar bertujuan untuk melakukan transformasi pada dimensi data agar data dapat diklasifikasikan secara linear. Setiap neuron pada *convolution layer* perlu ditransformasikan menjadi satu dimensi terlebih dahulu sebelum dimasukkan dalam suatu *fully connected layer*. Karena hal tersebut menyebabkan data kehilangan informasi spasialnya dan tidak

*reversible, fully connected layer* hanya dapat diimplementasikan di akhir jaringan.

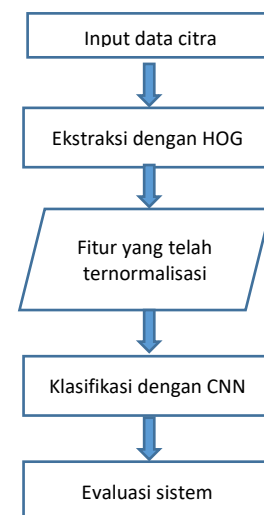
## OpenCV

OpenCV (*Open Source Computer Vision*) adalah library dari fungsi pemrograman untuk realtime visi komputer [12]. OpenCV menggunakan lisensi BSD dan bersifat gratis baik untuk penggunaan akademis maupun komersial. OpenCV dapat digunakan pada bahasa pemrograman C, C++, Python, Java, dan sebagainya [13]. OpenCV dapat digunakan pada sistem operasi Windows, Linux, Android, iOS, dan Mac OS. OpenCv memiliki lebih dari 250 algoritma yang telah dioptimalkan.

## 3. HASIL DAN PEMBAHASAN

Dataset yang digunakan adalah dataset yang disediakan pada *website* kaggle (<https://www.kaggle.com/>) dengan dataset yang digunakan adalah Face Expression Recognition Dataset. Dataset direkayasa menggunakan Python 3.7.3 dan dimanipulasi dengan Jupyter Notebook.

Untuk prosedur penelitian dapat dilihat pada Gambar 5 sebagai berikut :



Gambar 5 Prosedur penelitian

Pengembangan sistem ini dilakukan dengan menitikberatkan pada citra ekspresi wajah, dimulai dari akuisisi citra, selanjutnya dilakukan ekstraksi fitur dengan HOG. Setelah itu dilakukan pengklasifikasian hasil ekstraksi dengan CNN, dimana data akan dikelas-kelaskan dalam 7 kelas yang berbeda.

**4. HASIL DAN PEMBAHASAN**

Dari hasil pengerjaan sistem, didapatkan dataset yang digunakan terdiri dari 28 821 citra yang terdiri dari 7 kelas ekspresi wajah yang berbeda. Citra awal akan dilakukan praproses menjadi ukuran 48x48 piksel menggunakan Python 3.7.3, dan praproses dikerjakan menggunakan library Keras Python versi 2.3.1. Gambar 6 menampilkan citra yang digunakan dalam penelitian ini.



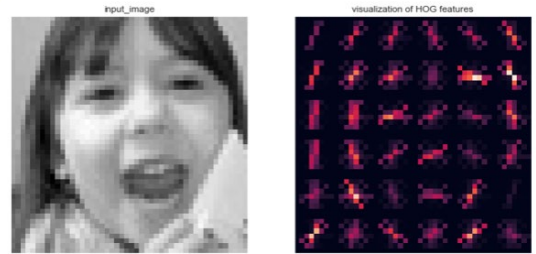
Gambar 6 : Dataset ukuran 48x48 piksel  
 Adapun dataset yang digunakan adalah sebagai berikut ditampilkan pada Tabel 1.

Tabel 1 : Komposisi dataset ekspresi wajah yang digunakan

Ekspresi	Jumlah
Angry	3993
Disgust	436
Fear	4103
Happy	7164
Neutral	4982
Sad	4938
Surprise	3205

Setelah dibaca dataset yang digunakan pada sistem, selanjutnya dilakukan ekstraksi fitur dengan menggunakan HOG. Ekstraksi fitur di-generate dengan menggunakan library Python skimage. Perintah yang digunakan adalah : `hog_vec, hog_vis = feature.hog(image, visualize = True)`. Untuk hasil yang diberikan ditunjukkan pada Gambar 6.

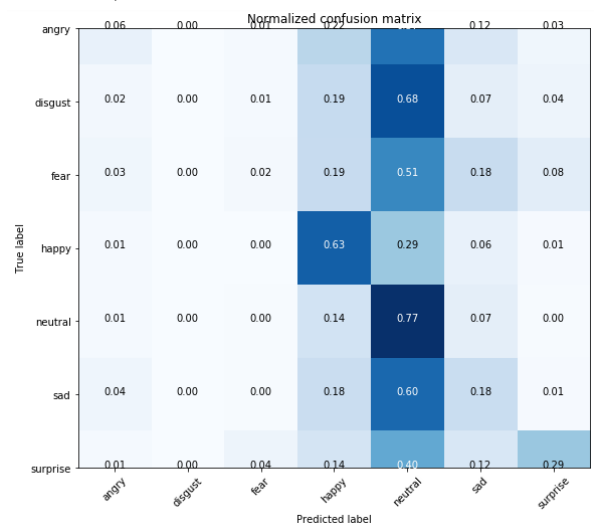
Out[5]: Text(0.5, 1.0, 'visualization of HOG features')



Gambar 6 Hasil visualisasi ekstraksi fitur dengan HOG

Setelah melakukan ekstraksi fitur, langkah selanjutnya adalah melakukan klasifikasi dengan deep learning CNN. Dari parameter yang digunakan yaitu 3 epochs, 4 convolution layers, 2 fully connected layer, laju belajar = 0.0001, dan menggunakan fungsi aktivasi adalah ReLU (Rectified linear unit), maka lama pembelajaran yang dilakukan oleh sistem adalah selama 52 menit dan 22 detik, maka hasil loss = 1.7192 dan akurasi dari epoch ketiga adalah 0.3358.

Selanjutnya, dihitung hasil dari kebenaran 7 kelas dari sistem. Tabel confusion matrix yang menampilkan kebenaran kelas target (True label) berbanding terhadap kelas model (Predicted label) ditunjukkan pada Gambar 7.



Gambar 7 Confusion matrix untuk deteksi ekspresi wajah dengan HOG dan CNN

Dari hasil confsion matrix, didapatkan bahwa sistem bisa mengenali untuk ekspresi wajah manusia dengan akurasi tertinggi sebesar 77%. Akurasi tersebut didapatkan dari hasil pengenalan kembali untuk wajah netral. Sedangkan untuk ekspresi wajah yang lain masih dikenali sebagai ekspresi wajah yang

lain. Hal ini disebabkan oleh epoch yang tidak terlalu banyak sehingga menyebabkan hasil dari sistem yang kurang baik untuk pengenalan ekspresi yang lain.

## 5. KESIMPULAN

Dari hasil sistem, didapatkan untuk pengenalan atau deteksi ekspresi wajah pada manusia untuk 7 ekspresi wajah yang berbeda sudah bisa dilakukan dengan HOG dan CNN. Hasil yang diberikan adalah sebesar 77% untuk akurasi tertinggi. Namun masih ada kelas yang belum memberikan hasil yang akurat, sehingga sistem masih menempatkan gambar pada kelas yang salah. Karena itu disarankan untuk meningkatkan epoch diatas 50 sehingga menjadi sistem yang lebih akurat.

## DAFTAR PUSTAKA

- [1] A. Mehrabian "Communication without words", *Psychology Today*, vol.2, pp.52-55,1968
- [2] T.Kanade, J.F Cohn and Y.Tian "Comprehensive database for facial expression analysis", in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, Grenoble, France, 2000
- [3] M. Haghighat, S. Zonouz and M. Abdel-Mottaleb, "CloudID : Trustworthy cloud-based and cross-enterprise biometric identification," *Expert Systems with Applications*, vol.42 no.21,pp.7905-7916,2015
- ..
- [4] M.Alwakeel and Z. Shaaban, "Face recognition based on Haar wavelet transform and principal component analysis via Levenberg-Marquardt backpropagation neural network," *European Journal of Scientific Research*, vol.42, no.1, pp.25-31,2010
- [5] M. Goyani and N. Patel, "Multi-Level Haar Wavelet based Facial Expression Recognition using Logistic Regression," *Indian Journal of Science and Technology*, vol.10, no.9,2017
- [6] J. Zhang et al., "Therapeutic detoxification of quercetin against carbon tetrachloride-induced acute liver injury in mice and its mechanism," *J. Zhejiang Univ. Sci. B*, vol. 15, no.12, pp.1039-1047, 2014
- [7] K, Seemanthini and S.S. Manjunath, "Human Detection and Tracking using HOG for Action Recognition," *Procedia COMpt. Sci*, vol. 132, no. Iccids, pp.1317-1326,2018
- [8] R. Hu, M. Barnard, and J.P. Collomosse, "Gradient field descriptor for sketch based retrieval and localization," in *ICIP*, 2010, pp.1025-1028.
- [9] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," dalam *one-Alps*, 655 avenue de 'Europe, Montbonnot 38334, France, 2005
- [10] A. Karpathy, "CS231 in Convolutional Neural Network for Visual Recognition," *Stanford University*, [Online]. Available: <http://cs231n.github.io/neural-networks-1/>
- [11] I. Corporation, "OpenCV," Itseez, 2019. [Online]. Available : <https://opencv.org/>