

Pengenalan Isyarat Tangan dengan Algoritma Pyramidal Lucas Kanade dan Histogram of Oriented Gradients

Hand Gesture Recognition using Pyramidal Lucas Kanade Algorithm and Histogram of Oriented Gradients

Gustav Nugraha ¹, Hertog Nugroho, Ir., M.Sc., Ph.D ², Tjokorda Agung Budi Wirayuda, ST., MT ³

^{1,3}Prodi S1 Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

²Prodi S2 Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

¹gustav.nugraha@gmail.com, ²hertog@melsa.net.id, ³cokagung2001@gmail.com

Abstrak

Isyarat tangan merupakan salah satu metode interaksi manusia dan komputer yang populer digunakan saat ini. Terdapat isyarat tangan statis yang hanya menggunakan perubahan posture tangan dan dinamis yang secara bersamaan menggunakan perubahan posture tangan dan pola gerakan tangan.

Untuk mengenali isyarat tangan statis dan dinamis maka diperlukan metode yang mampu mendeskripsikan posture tangan dan mendeskripsikan pola gerakan tangan. Dalam penelitian ini, Histogram of Oriented Gradients adalah metode yang digunakan untuk mendeskripsikan posture tangan dan Pyramidal Lucas Kanade dipilih untuk mengestimasi optical flow yang digunakan untuk mendeskripsikan pola pergerakan tangan dalam bentuk sebuah Motion Descriptor. Pendekatan Bag of Features akan digunakan untuk mendapatkan descriptor akhir dari sebuah video untuk HOG, dan klasifikasi akan dilakukan dengan menggunakan Support Vector Machine.

Sistem yang telah dibuat diuji dengan dataset yang terdiri dari 6 isyarat tangan. Pengujian dengan metode k-fold cross validation menunjukkan akurasi rata-rata 95.73% pada dataset yang terdiri dari 240 video yang terdiri dari isyarat tangan dari 2 orang yang berbeda.

Kata kunci : Isyarat Tangan, Histogram of Oriented Gradients, Pyramidal Lucas Kanade, Optical Flow, Bag of Features

Abstract

Hand gesture is one of popular human computer interaction method these day. There are static gesture which rely on hand posture changes and dynamic gestures simultaneously rely on posture change and hand motion.

To recognize static and dynamic hand gestures, we need methods to describe hand posture and describe hand motion. In this final project, Histogram of Oriented Gradients used to describe hand posture and Pyramidal Lucas Kanade algorithm used to estimate optical flow which used to describe hand motion pattern in the form of a Motion Descriptor. Bag of Features used to get a video final descriptor for HOG, and Support Vector Machine used for classification.

System tested using dataset consist of 6 hand gesture. Testing using k-fold cross validation shown average accuracy 95.73% over 240 videos dataset consist of hand gestures performed by 2 different individual.

Keywords : Hand Gesture, Histogram of Oriented Gradients, Pyramidal Lucas Kanade, Optical Flow, Bag of Features

1. Pendahuluan

Pengenalan isyarat tangan oleh komputer dapat membuat interface antara mesin dan manusia yang dapat digunakan untuk berkomunikasi dengan komputer secara lebih natural daripada menggunakan perintah keyboard atau mouse. Aplikasi dari pengenalan isyarat tangan banyak digunakan pada pengenalan bahasa isyarat, robotik, dan permainan [13].

Terdapat isyarat tangan statis yang hanya menggunakan perubahan posture tangan dan isyarat tangan dinamis yang juga melibatkan pola gerakan tangan [2]. Untuk mengenali isyarat tangan yang statis dan dinamis, maka diperlukan sistem yang dapat mendeskripsikan posture tangan yang secara bersamaan mengenali pola gerakan tangan.

2. Dasar Teori

2.1. Histograms of Oriented Gradients

Histograms of Oriented Gradients (HOG) adalah metode untuk mendeskripsikan bentuk sebuah objek pada sebuah gambar dengan memanfaatkan intensitas lokal dari gradient gambar [7]. Hal ini dilakukan dengan membagi window dari image menjadi beberapa bagian yang disebut “cell”. Untuk invarian yang lebih baik terhadap pencahayaan, bayangan, dan lain-lain, akan lebih baik jika melakukan normalisasi lokal sebelum menggunakannya, hal ini dapat dilakukan dengan mengumpulkan histogram lokal dari semua cell pada sebuah region yang lebih besar yang disebut “block”. Gradient citra akan dihitung pada setiap pixel dengan menggunakan persamaan (1).

$$\begin{aligned} I_x(x, y) &= I(x + 1, y) - I(x - 1, y) \\ I_y(x, y) &= I(x, y + 1) - I(x, y - 1) \end{aligned} \tag{1}$$

dimana $I_x(x, y)$ dan $I_y(x, y)$ adalah gradient vertikal dan horizontal dan $I(x, y)$ adalah intensitas atau nilai pixel pada posisi (x, y) . Perhitungan magnitude $M(x, y)$ dan orientasi $\Phi(x, y)$ dari gradient dilakukan dengan persamaan (2) dan (3).

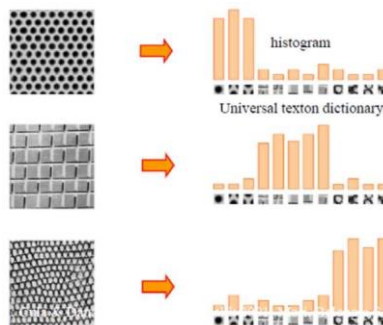
$$M(x, y) = \sqrt{I_x^2(x, y) + I_y^2(x, y)} \tag{2}$$

$$\Phi(x, y) = \tan^{-1} \left(\frac{I_y(x, y)}{I_x(x, y)} \right) \tag{3}$$

Histogram pada setiap cell akan membagi rentang sudut gradient menjadi n bin orientasi. Bin orientasi dibagi secara merata pada rentang 0°- 180°. Setiap pixel di dalam cell akan memberikan vote berbobot untuk bin orientasi tepi dimana orientasi gradient tersebut berpusat [7]. Vote akan diinterpolasi linear di antara pusat bin yang bertetangga. Histogram dari cell-cell yang berada dalam blok yang sama akan digabungkan untuk membentuk histogram pada level blok.

2.2. Bag of Features [5]

Bag of Features (BOF) juga sering disebut Bag of Words merupakan sebuah pendekatan yang memperlakukan ciri dari sebuah gambar sebagai sebuah words. Dengan metode ini, sebuah gambar direpresentasikan ke dalam kumpulan ciri lokal yang tidak beraturan [5]. Di Bag of Words, ciri local yang tidak berurutan akan menjadi kandidat dari elemen dasar yang disebut words, descriptor ciri akan direpresentasikan seperti vector numerik. Dengan clustering, vector numerik akan dikonversi menjadi codeword (pusat cluster) untuk membentuk codebook. Jumlah dari kluster akan menjadi ukuran codebook.



Gambar 1 Contoh Bag of Features pada Klasifikasi Texture.

2.3. Pyramidal Lucas Kanade [3]

Optical flow merupakan metode untuk melakukan perkiraan posisi sebuah titik pada sebuah gambar I yang berada pada titik (x, y) pada sebuah gambar I , sehingga diperoleh nilai $(u, v) = (u_1 + \Delta x, v_1 + \Delta y)$ dimana (u_1, v_1) dan (u, v) bernilai "sama" dan meminimalkan nilai dari persamaan (2.4).

$$E(I) = \sum_{\Delta x, \Delta y} \sum_{\Delta u, \Delta v} |I(x, y) - I(x + \Delta x, y + \Delta y, u + \Delta u, v + \Delta v)| \quad (4)$$

Area $(2\Delta x + 1) \times (2\Delta y + 1)$ yang mengelilingi titik (x, y) disebut integration window. Dengan,

$$I = \sum_{\Delta x, \Delta y} \sum_{\Delta u, \Delta v} |I(x, y) - I(x + \Delta x, y + \Delta y, u + \Delta u, v + \Delta v)| \quad (5)$$

$$E = \sum_{\Delta x, \Delta y} \sum_{\Delta u, \Delta v} |I(x, y) - I(x + \Delta x, y + \Delta y, u + \Delta u, v + \Delta v)| \quad (6)$$

dimana I_x dan I_y merupakan gradient gambar pada sumbu x dan y, Δu perbedaan nilai pixel dari I dan I .

$$v = \Delta u \Delta y \quad (7)$$

Representasi piramida dari gambar merupakan urutan gambar dimana ukuran gambar pada level $\Delta + 1$ merupakan setengah dari gambar pada level Δ . Konstruksi dari piramida gambar ini dilakukan secara rekursif, dimana gambar pada level $\Delta + 1$ didapatkan dengan melakukan subsampling gambar Δ yang telah difilter dengan menggunakan kernel filter $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 4 & 6 & 4 & 1 \\ 1 & 6 & 16 & 6 & 1 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}$.

Pada setiap level piramida Δ akan terdapat vector Δ yang merupakan hasil perhitungan perpindahan pada level $\Delta + 1$. Untuk level paling atas Δ_0 vector $\Delta_0 = (0, 0)$. Sehingga harus dicari vector Δ meminimalkan nilai fungsi:

$$E(\Delta) = \sum_{\Delta x, \Delta y} \sum_{\Delta u, \Delta v} |I(x, y) - I(x + \Delta x, y + \Delta y, u + \Delta u, v + \Delta v)| \quad (8)$$

Jika, $I_x(x, y) = I_x(x, y)$, $I_y(x, y) = I_y(x, y)$ dan $I(x, y) = I(x + \Delta x, y + \Delta y)$, maka persamaan di atas dapat ditulis sebagai berikut.

$$E(\Delta) = \sum_{\Delta x, \Delta y} \sum_{\Delta u, \Delta v} |I(x, y) - I(x + \Delta x, y + \Delta y, u + \Delta u, v + \Delta v)| \quad (9)$$

sehingga, $\Delta = 2(\Delta_1 + \Delta_2)$. Optical flow akhir yang akan didapatkan adalah Δ .

2.4. Velocity Orientation Histogram [12]

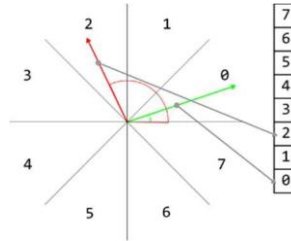
Histogram ini akan mengkalkulasi distribusi orientasi dari pergerakan. Untuk vector optical flow dengan magnitude bukan nol, maka $\theta(I)$ merupakan orientasi dari vector I . Sama seperti HOG, orientasi dari optical flow pada masing-masing frame akan diberi bobot sesuai magnitude vector, sehingga nilai dari setiap bin histogram akan memenuhi persamaan berikut.

$$H(I) =$$

$$\frac{\sum_{i=1}^n \|\varphi_i\|^2}{\sum_{i=1}^n \|\varphi_i\|}$$

(10) $\|\varphi_i\|$

dimana $\theta_1, \theta_2, \dots, \theta_n$, n adalah jumlah dari orientasi. Voting histogram akan dilakukan sebagai berikut.



Gambar 2 Voting Histogram 8 Bin Optical Flow.

Perhitungan orientasi atau arah θ dan magnitude ρ dari sebuah vector $\psi = \psi_x, \psi_y$ dilakukan dengan menggunakan persamaan berikut.

$$\theta = \tan^{-1} \frac{\psi_y}{\psi_x} \tag{11}$$

$$\rho = \sqrt{\psi_x^2 + \psi_y^2} \tag{12}$$

2.5. Motion Descriptor [12]

Descriptor untuk sebuah video dengan n frame akan berisi statistik temporal yang didapatkan dari histogram bin semua frame dengan memenuhi persamaan berikut.

1. Maximum : $\mu(\theta) = \max_{\theta} \{h(\theta)\}$
2. Mean : $\mu(\theta) = \frac{\sum_{\theta} \theta \cdot h(\theta)}{\sum_{\theta} h(\theta)}$
3. Standard deviation : $\sigma(\theta) = \sqrt{\sum_{\theta} h(\theta) \cdot \theta^2 - (\mu(\theta))^2}$

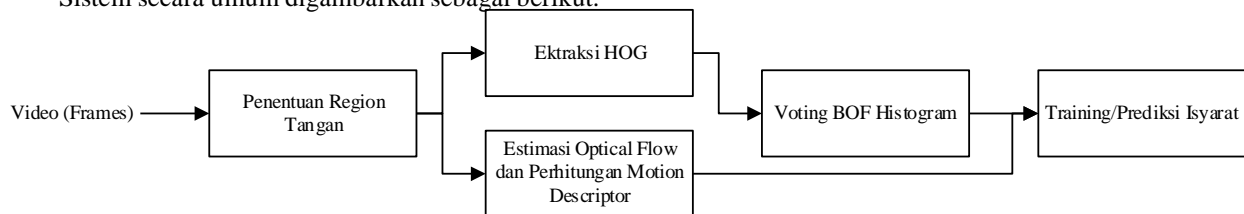
Deret frame juga akan dibagi menjadi 3 interval dengan durasi yang sama dari ketiga interval akan dihitung rata-rata histogramnya dengan persamaan sebagai berikut.

4. Mean Begin : $\mu_b(\theta) = \frac{\sum_{\theta} \theta \cdot h(\theta)}{\sum_{\theta} h(\theta)}$
5. Mean Middle : $\mu_m(\theta) = \frac{\sum_{\theta} \theta \cdot h(\theta)}{\sum_{\theta} h(\theta)}$
6. Mean End : $\mu_e(\theta) = \frac{\sum_{\theta} \theta \cdot h(\theta)}{\sum_{\theta} h(\theta)}$

Descriptor yang didapatkan akan berukuran $6 \times$ jumlah bin orientasi.

3. Pembahasan

Sistem secara umum digambarkan sebagai berikut.



Gambar 3 Gambaran Umum Sistem.

3.1. Penentuan Region Tangan

Untuk menemukan posisi tangan pada sebuah frame video dilakukan tahap-tahap berikut.



Gambar 4 Proses Penentuan Region Tangan.

1. Konversi warna dari RGB ke YCrCb.
2. Lakukan proses thresholding untuk menemukan area dengan warna yang sesuai dengan distribusi warna kulit dengan menggunakan nilai threshold berikut.

$$\begin{cases} 133 \leq Y \leq 173 \\ 77 \leq Cr \leq 127 \end{cases} \quad (13)$$

3. Dari citra binary atau mask yang dihasilkan melalui langkah 2, akan dilakukan pencarian area dengan luas terbesar, dan area tersebut dijadikan sebagai Region of Interest (ROI) tangan.
4. Lebar atau panjang ROI akan disamakan (mengikuti yang lebih besar) untuk membentuk persegi dan akan ditambahkan 2 pixel padding.

3.2. Ekstraksi HOG

Dari ROI tangan yang telah ditemukan, maka ekstraksi dapat dilakukan, melalui beberapa tahap. Konversi citra dari representasi RGB ke grayscale dan lakukan proses cropping area tangan pada representasi grayscale tersebut. Ekstraksi HOG diperlukan ukuran yang tetap, maka dilakukan proses scaling pada citra tangan menjadi ukuran 42×42 [18].

Dari citra hasil scaling maka dilakukan kalkulasi gradient gambar yang akan menghasilkan 2 buah matrix gradient G_x dan G_y yang berukuran 42×42. Orientasi dan magnitude gradient dikalkulasi juga menghasilkan matrix berukuran 42×42 untuk masing-masing orientasi dan magnitude gradient.

Masing-masing matrix orientasi dan magnitude akan dibagi menjadi beberapa cell berukuran 21×21. Pada setiap cell akan dilakukan voting untuk histogram orientasi berukuran 9 bin yang membagi rata rentang orientasi 0°- 180°, dan setiap elemen matrix dalam sebuah cell akan memberikan vote orientasi sebesar magnitude-nya dan vote akan diinterpolasi secara linear pada 2 pusat bin terdekat.

Normalisasi akan dilakukan per blok yang berukuran 2×2 (dalam unit cell) dan setiap blok pada gambar, histogram berukuran 9 bin pada setiap cell yang berada pada blok yang sama akan digabungkan.

3.3. Bag of Features untuk HOG

Berikut adalah tahapan dilakukan untuk memperoleh histogram BOF untuk HOG untuk membentuk vector ciri akhir yang akan mendeskripsikan posture tangan dalam sebuah video (sebuah video hanya berisikan sebuah isyarat tangan).

Seluruh vector ciri HOG dari seluruh video (yang digunakan untuk training) akan dikumpulkan sebagai menjadi satu bag. Dilakukan clustering dengan menggunakan k-means clustering untuk memperoleh 4 buah words (pusat kluster yang berupa histogram) yang akan membentuk codebook. Nilai 4 dipilih karena memberikan akurasi maksimal pada percobaan.

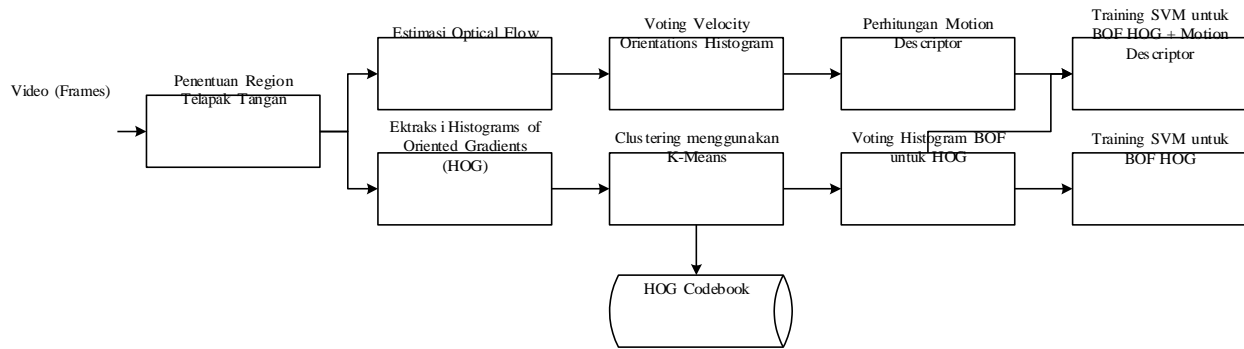
Histogram BOF untuk setiap video akan didapatkan mengelompokkan setiap histogram HOG berdasarkan word (pusat kluster) paling dekat. Proses ini akan menghasilkan histogram akhir dari sebuah video dengan ukuran N .

3.4. Ekstraksi Motion Descriptor

Dari ROI yang telah ditemukan, maka 20×20 titik akan disebar secara rata pada ROI pada frame untuk diestimasi optical flow-nya pada frame selanjutnya (20×20 titik akan disebar ulang pada setiap frame video). Titik yang akan diestimasi optical flownya hanya titik yang berada di daerah tangan saja. Estimasi dilakukan dengan menggunakan algoritma Pyramidal Lucas Kanade dengan bantuan fungsi `calcOpticalFlowPyrLK` pada library OpenCV. Vector perpindahan akan didapatkan dengan mengurangkan posisi pada titik pada dengan posisi titik pada . Proses ini akan menghasilkan vector perpindahan pada sumbu x dan y, sebut saja dan .

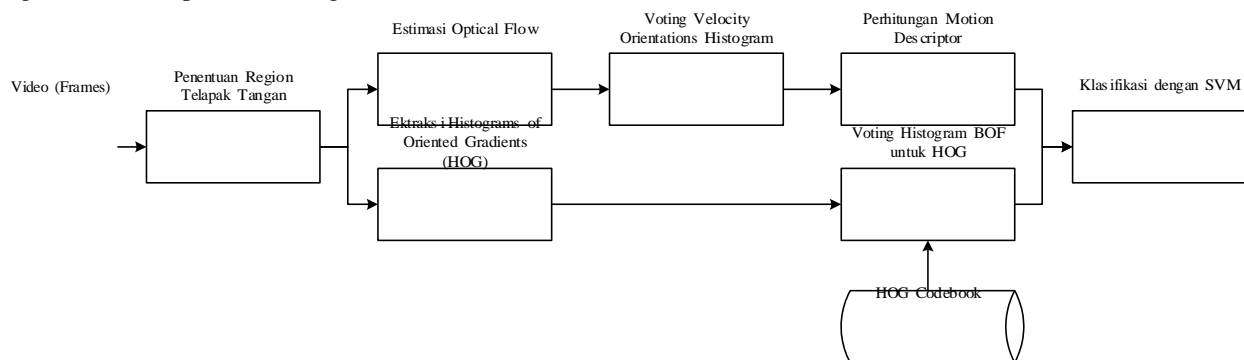
Voting histogram dilakukan untuk membentuk Velocity Orientations Histogram seperti yang dijelaskan pada bagian 2.4. Pada penelitian ini, jumlah bin histogram yang digunakan adalah 8 yang akan membagi rata rentang $0^\circ - 360^\circ$. Motion Descriptor didapatkan sesuai dengan yang dijelaskan pada bagian 2.5.

3.5. Training dan Klasifikasi



Gambar 5 Block Diagram Training Sistem.

Pada skema training, seluruh ciri BOF HOG dari seluruh dataset (2 isyarat statis dan 4 isyarat dinamis) digunakan sebagai data training untuk SVM yang akan digunakan untuk mengklasifikasi isyarat tangan dari posturanya saja. U vector ciri BOF HOG sama dengan jumlah kluster (4) yang dipilih. Kemudian, vector ciri dari BOF HOG untuk dataset 4 isyarat dinamis digabungkan secara horizontal dengan vector motion descriptor yang juga berasal dari dataset 4 isyarat dinamis dan digunakan sebagai data training SVM untuk melakukan klasifikasi terhadap isyarat tangan yang dinamis. Ukuran vector ciri motion descriptor adalah $6 \times$ jumlah bin velocity orientation histogram. Metode yang digunakan untuk proses training dan klasifikasi adalah SVM linear.



Gambar 6 Skema Klasifikasi Sistem.

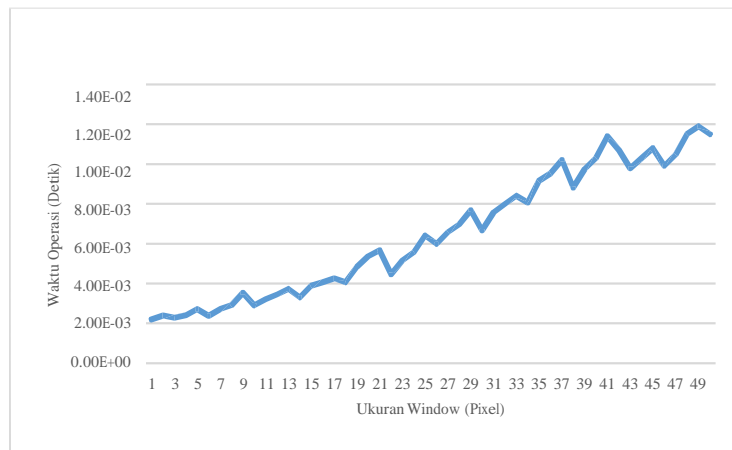
Proses klasifikasi dilakukan pertama hanya menggunakan fitur BOF HOG yang telah dihasilkan dengan menggunakan SVM yang telah dilatih dengan vector ciri BOF HOG saja, jika hasil klasifikasi menunjukkan bahwa isyarat tangan tersebut merupakan isyarat tangan yang statis (Isyarat 1 dan 2), maka motion descriptor tidak akan digunakan, namun jika hasil klasifikasi menunjukkan bahwa isyarat tangan merupakan isyarat tangan yang dinamis

(Isyarat 3,4,5,6) maka proses klasifikasi akan dilanjutkan dengan menambahkan motion descriptor pada vector ciri video, kemudian klasifikasi ulang akan dilakukan dengan menggunakan SVM yang telah dilatih dengan isyarat tangan yang dinamis (Vector ciri BOF HOG dan motion descriptor).

4. Pengujian

Pengujian dilakukan pada dataset 240 video yang terbagi menjadi 6 isyarat tangan. K-fold cross validation digunakan untuk pengujian dengan rotasi data training dan data testing, akurasi rata-rata sebesar 95.73%. Pengujian juga dilakukan dengan merotasi 120 video isyarat tangan individu 1 dan 120 video isyarat tangan individu 2 sebagai data training dan data testing, akurasi yang didapatkan rata-rata 97.08%.

Pengujian juga dilakukan pada kompleksitas algoritma Pyramidal Lucas Kanade terhadap ukuran integration window yang digunakan, karena untuk dapat melakukan estimasi optical flow pada objek yang bergerak dengan cepat algoritma ini sangat bergantung dengan ukuran integration window yang digunakan. Kompleksitas algoritma Pyramidal Lucas Kanade terhadap perubahan ukuran integration window adalah $O(n^2)$ yang artinya peningkatan ukuran integration window akan meningkatkan cost komputasi yang mengikuti pertumbuhan fungsi eksponensial (n^2) .



Gambar 7 Pengukuran Waktu Operasi Terhadap Perubahan Ukuran Integration Window.

5. Kesimpulan

Dari implementasi, pengujian dan analisis yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut :

1. Sistem mampu mengenali isyarat tangan statis dan dinamis dengan baik, terlihat dengan akurasi rata-rata 95.73% pada pengujian dengan menggunakan k-fold cross validation.
2. Sistem dapat mengenali isyarat meskipun data testing yang digunakan belum pernah digunakan pada training. Hal ini terlihat pada rata-rata akurasi sistem 97.08% ketika 120 dataset dari individu 1 dan 120 dataset dari individu 2 digunakan sebagai data training dan data testing secara bergantian.
3. Agar sistem mampu menangani pergerakan yang cepat akan diperlukan integration window yang besar pada algoritma Pyramidal Lucas Kanade, namun cost komputasinya akan meningkat.

Daftar Pustaka :

- [1] Alpaydin, Ethem, 2010, Introduction to Machine Learning, The MIT Press.
- [2] Baraldi, Lorenzo., Paci, Francesco., Serra Giuseppe., Benini, Luca., Cucchiara, Rita., 2014, Gesture Recognition in Ego-Centric Videos using Dense Trajectories and Hand Segmentation, IEEE Conference on Computer Vision and Pattern Recognition.
- [3] Bouguet, Jean-Yves., 2001, Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the Algorithm, Intel Corporation.

- [4] Bradski, Gray., Kaehler, Adrian., 2008, Learning OpenCV, O'Reilly Media.
- [5] Cerna, L. R., Camara-Chavez, G., Menotti, D., 2013, Face Detection: Histogram of Oriented Gradients and Bag of Feature Method, International Conference on Image Processing, Computer Vision, and Pattern Recognition.
- [6] Chaudhry, Rizwan., Ravichandran, Avinash., Hager, Gregory., Vidal, Rene., 2009, Histograms of Oriented Optical Flow and Binet-Cauchy Kernels on Nonlinear Dynamical Systems for the Recognition of Human Actions, IEEE Conference on Computer Vision and Pattern Recognition.
- [7] Dalal, Navneet., Tiggs, Bill., 2005, Histograms of Oriented Gradients for Human Detection, IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
- [8] Dalal, Navneet., Tiggs, Bill., Schmid, Cordelia., 2006, Human Detection Using Oriented Histogram of Flow and Appearance, European Conference on Computer Vision 2006.
- [9] Freeman, William T., Roth, Michal., 1994, Orientation Histograms for Hand Gesture Recognition, Mitsubishi Research Laboratory.
- [10] Kalal, Zdenek., Mikolajczyk, Krystian., Matas, Jiri., 2011, Tracking Learning Detection, Pattern Analysis and Machine Intelligence.
- [11] Konecny, Jakub., Hagara, Michal., 2014, One-Shot-Learning Gesture Recognition using HOG-HOF Features, Journal of Machine Learning Research 15.
- [12] Martinez, Fabio., Manzanera, Antoine., Romero, Eduardo., 2012, A Motion Descriptor Based on Statistics of Optical Flow Orientations for Action Classification in Video-Surveillance, Springer.
- [13] Misra, Arindam., Takashi, Abe., Okatani, Takayuki., Deguchi, Koichiro., 2011, Hand Gesture Recognition using Histogram of Oriented Gradients and Partial Least Squares Regression, IAPR Conference on Machine Vision Applications, Nara, Japan.
- [14] Muchidiarti, Fitri., 2013, Perancangan dan Implementasi Sistem Penerjemah Bahasa Isyarat Penderita Tunawicara dan Tunarungu Dengan Jaringan Syaraf Tiruan, Institut Teknologi Telkom, Bandung.
- [15] Syahrir, Taufiq., 2013, Penelusuran dan Klasifikasi Isyarat Tangan Menggunakan Skeleton Graph Secara Real Time, Institut Teknologi Telkom, Bandung.
- [16] Wang, Heng., Klaser, Alexander., Schmid, Cordelia., Liu, Cheng-Lin., 2011, Action Recognition by Dense Trajectories, IEEE Conference on Computer Vision and Pattern Recognition.
- [17] Yahya Urwate, Shulha., 2012, Deteksi dan Klasifikasi Isyarat Tangan Menjadi Remote Control Komputer dengan Metode Haar Cascade dan Contour Flexibility, Institut Teknologi Telkom, Bandung.
- [18] Zhao, Yafei., Wang, Weidong., Wang, Yuehai., 2011, A Real-time Hand Gesture Recognition Method, IEEE International Conference on Electronics, Communication and Control.