

Pengenalan Isyarat Tangan dan Pengaruh Lingkungan pada Ekstraksi Fitur

Andika Mulyawan Dwi Prapmanto^{#1}, Setia Budi^{*2}

[#]Teknik Informatika, Universitas Kristen Maranatha
Jl. Prof. drg. Surya Sumantri No. 65 Bandung

¹dikaprapmanto@gmail.com

² Setia.Budi@it.maranatha.edu

Abstract — Communication can be done in verbal or nonverbal way. Nonverbal communication means communicate without using voice from the mouth but using letters, pictures, or sign language. In communication with the gestures there will be difficulties if the other person does not understand the meaning of the sign given. This final project propose to develop an application that able to translate alphabet c, i, l, o, u, and v from BISINDO sign language using SVM. For computers able to receive data from the given sign language, computer vision methods is used. To find out the ideal environment used to make detection, exploring combination of background colors and direction of light is done. From the results of obtained evaluation, the black background and front light results in the best evaluation value.

Keywords— BISINDO, computer vision, hand sign, opencv, SVM

I. PENDAHULUAN

Komunikasi dapat dilakukan melalui *verbal* maupun *nonverbal* [1]. Komunikasi *verbal* dilakukan dengan cara berkomunikasi menggunakan suara yang dihasilkan melalui mulut individu dan dapat dimengerti oleh lawan bicara. Sedangkan komunikasi *nonverbal* dilakukan menggunakan perantara seperti tulisan, gambar, atau isyarat tubuh yang dapat dimengerti agar informasi yang ingin disampaikan dapat dipahami oleh lawan bicara.

Isyarat adalah segala bentuk gerakan yang dipakai sebagai tanda dan memiliki informasi yang ingin disampaikan pada lawan bicara[8]. Tanda ini berupa isyarat yang diakui dan dipahami oleh pemakai dan lawan bicaranya. Namun, akan timbul permasalahan bila lawan bicara tidak memahami isyarat yang diberikan. Untuk itu, diperlukan sebuah media (alat) yang dapat membantu menerjemahkan isyarat (tanda) tersebut agar informasi yang tersampaikan meski lawan bicara tidak memahami isyarat yang dipakai.

Dalam masalah ini dapat diterapkan *machine learning* agar mesin (komputer) dapat menjadi media yang membantu menerjemahkan isyarat. Terdapat dua metode agar komputer dapat mengenali isyarat yang diberikan di antaranya menggunakan kamera dengan metode penglihatan komputer [2] [3] dan sensor [4]. Penggunaan sensor yang dipasang pada anggota tubuh dapat menyulitkan pemakainya saat memberikan isyarat. Dalam penglihatan komputer tidak diperlukan pemasangan sensor pada anggota tubuh sehingga dapat memberikan isyarat dengan cara lebih natural.

Oleh karena itu, pada proyek penelitian ini penulis akan fokus pada eksplorasi *machine learning* untuk menerjemahkan isyarat tangan dan menganalisis pengaruh cahaya dan warna latar terhadap *classifier* yang terbentuk. Penulis akan membangun *dataset* sesuai kondisi lingkungan dan isyarat yang dipakai sebagai fondasi *machine learning* membangun *classifier*. Dalam pengerjaannya penulis akan memanfaatkan algoritma *Support Vector Machine* (SVM) dan *library Open Source Computer Vision* (OpenCV) [5] [6] [7].

II. KAJIAN TEORI

A. Isyarat Tangan

Gerak isyarat tangan dikelompokkan menjadi isyarat statis dan dinamis. Pada isyarat statis disampaikan dalam keadaan diam seperti sedang berpose. Berbeda dengan isyarat dinamis yang disampaikan dengan gerakan - gerakan tertentu menggunakan anggota tubuh [8].

Demi keselarasan berkomunikasi menggunakan isyarat, setiap negara memiliki bentuk bahasa isyarat tersendiri. American Sign Language (ASL) di negara Amerika, British Sign Language (BSL) di negara Inggris, dan Indian sign language di

negara India. Untuk Negara Indonesia masih terjadi pertentangan penggunaan bahasa isyarat antara Sistem Isyarat Bahasa Indonesia (SIBI) dan Bahasa Isyarat Indonesia (BISINDO).. [9] [10] [11]

B. Computer Vision (CV)

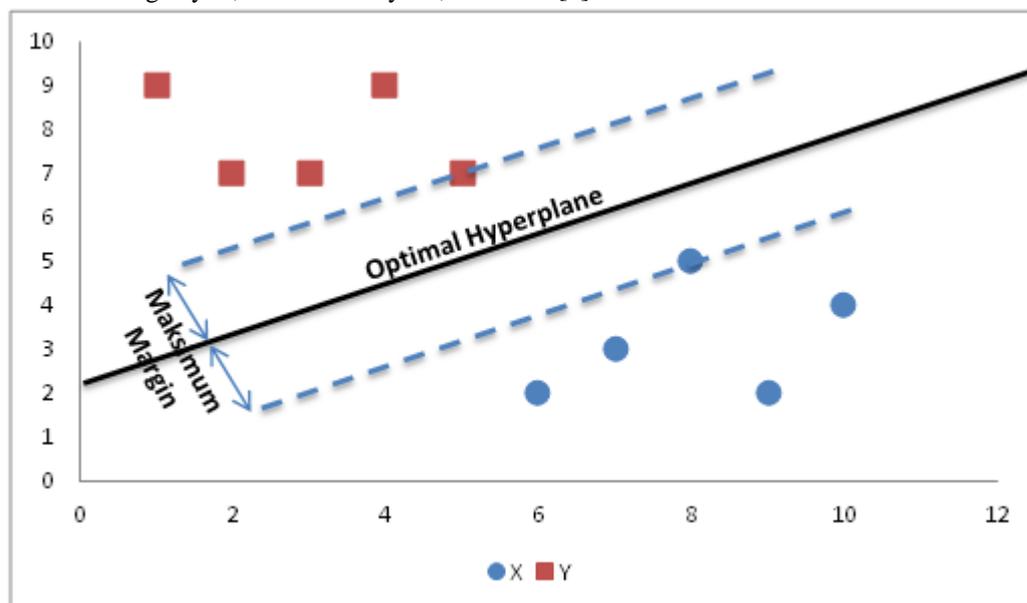
Computer Vision (CV) merupakan salah satu cabang ilmu pengolahan citra (image processing) dengan tujuan menghasilkan informasi melalui pengolahan gambar atau video. Tentu komputer tidak akan menghasilkan informasi hanya melalui gambar, perlu dilakukan pemrosesan pada gambar terlebih dahulu dan ditambahkan data lain berupa kata – kata (label atau *class*) sebagai penanda isi gambar [3] [12] [5].

Proses CV bekerja merupakan peniruan manusia dalam melihat, mempelajari, dan mengenali benda di sekitarnya. Manusia melalui matanya sedangkan CV melalui kamera atau gambar pada perangkat penyimpanan komputer. Terkadang orang keliru dengan metode ini melihat betapa mudahnya manusia dalam mengenali dan mempelajari apa yang dilihat namun, berbeda dengan CV. CV pada komputer melihat gambar dalam bentuk kumpulan angka dengan nilai yang bervariasi. Komputer perlu memproses gambar yang diterima dan mendapatkan data yang dianggap penting dalam hal ini CV melakukan proses mempelajari gambar yang diterima. Data yang telah diperoleh diolah sehingga didapatkan informasi yang diinginkan. Informasi dari gambar yang diperoleh melalui CV dapat berupa nama benda, jumlah benda, atau kondisi benda yang ada pada gambar pada tahapan ini CV telah melakukan proses mengenali gambar. Dengan informasi yang tersedia diharapkan mesin dapat mengenali benda pada gambar dan mengerjakan aksi yang telah ditentukan [5].

C. Support Vector Machine (SVM)

SVM adalah salah satu metode pembelajaran mesin yang dapat melakukan klasifikasi maupun regresi. Menurut Boser SVM sebagai metode klasifikasi melakukan pengelompokan data menjadi dua dan mencari nilai margin maksimal yang memisahkan kedua data. Jadi dalam klasifikasi sederhana SVM memisahkan data menggunakan bidang pemisah (*hyperplane*) dengan mencari nilai margin dengan jarak antara objek terluar dengan *hyperplane* bernilai maksimal seperti pada Gambar 1. Bentuk *hyperplane* akan bergantung pada jumlah fitur dari objek, bila terdapat dua fitur maka bentuk *hyperplane* adalah garis, bila terdapat tiga fitur maka bentuk *hyperplane* adalah bidang, dengan perhitungan dimensi *hyperplane* (n-1). Setiap objek yang akan diklasifikasi akan memiliki label dan fitur masing – masing. Fitur ini akan menjadi tanda yang membedakan kelompok dalam klasifikasi. [5] [6].

Gambar 1 menunjukkan contoh hasil klasifikasi yang terbentuk dari kasus pemisahan dua kelas atau *binary classification* dengan pemisahan sempurna antar kelasnya. Pada gambar ini setiap kelas memiliki dua buah fitur sehingga membentuk bidang pemisah berupa garis dengan nilai $y=0, w.x+b=0$. Di mana nilai b merupakan bias atau *offset* dari *hyperplane*, w merupakan *weight* dari *hyperplane* dan yang menentukan orientasinya dan x merupakan nilai titik yang ada pada *hyperplane*. *Hyperplane* akan memiliki nilai margin yang membentuk batasan terluar dengan masing – masing kelas. Nilai margin ini dirumuskan dengan $y=1, w.x+b=1$ dan $y=-1, w.x+b=-1$ [7].



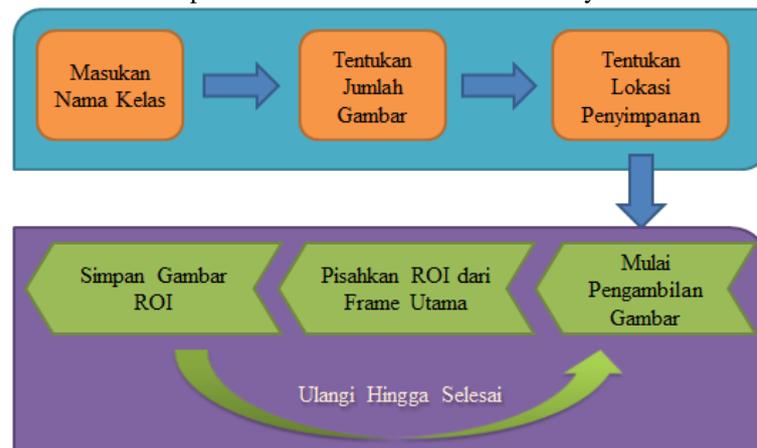
Gambar 1. Contoh klasifikasi SVM

III. METODOLOGI

A. Dataset

Dalam pengerjaan proyek penelitian ini akan dilakukan eksplorasi pengaruh kondisi lingkungan dalam pengenalan isyarat tangan. Untuk itu, dibutuhkan *dataset* gambar isyarat tangan dengan jumlah memadai pada masing - masing kondisi lingkungan.

Guna memperoleh *dataset* yang memadai penulis akan membangun aplikasi untuk membantu mengumpulkan *dataset* yang diperlukan. Proses yang akan dilakukan dalam menghasilkan *dataset* gambar dapat dilihat dalam diagram Gambar 2. Tahap pertama yang perlu dilakukan adalah menentukan nama kelas (label), jumlah data gambar yang akan diperoleh, dan lokasi direktori penyimpanan data gambar yang telah diperoleh. Selanjutnya memulai proses merekam dan menampilkannya pada *frame* utama. Posisi isyarat tangan harus berada di dalam kotak merah atau ROI yang disediakan pada *frame* utama agar perolehan data fokus dan bebas dari objek yang tidak diperlukan. ROI gambar isyarat tangan kemudian disimpan ke dalam direktori komputer dan diberi nama sesuai kelas isyarat dan urutan data gambar diambil.



Gambar 2. Diagram alir proses perolehan *dataset*

Guna memfokuskan bentuk isyarat yang akan digunakan, dalam proyek ini penulis memutuskan menggunakan enam bentuk isyarat alfabet dari BISINDO seperti pada Gambar 3. Bentuk isyarat tersebut di antaranya:

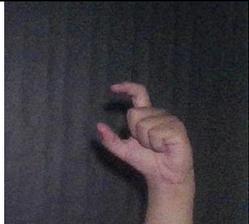
- Alfabet huruf C (Gambar 3 kiri atas);
- Alfabet huruf I (Gambar 3 tengah atas);
- Alfabet huruf L (Gambar 3 kanan atas);
- Alfabet huruf O (Gambar 3 kiri bawah);
- Alfabet huruf U (Gambar 3 tengah bawah);
- Alfabet huruf V (Gambar 3 kanan bawah);



Gambar 3. Gambar isyarat tangan yang digunakan

Kondisi lingkungan yang akan dieksplorasi dalam proyek penelitian ini adalah pengaruh arah datang cahaya dan warna latar. Arah cahaya yang akan dieksplor di antaranya arah cahaya depan, atas, dan samping kanan dengan jarak antara lampu dan latar masing – masing arah cahaya adalah 103cm, 60cm, dan 60cm menggunakan lampu LED dengan daya tiga watt dan terakhir menggunakan cahaya ruangan. Pada latar penulis akan menggunakan kertas karton dengan warna hitam dan putih. Contoh data gambar yang diambil sesuai kondisi lingkungan dapat dilihat pada TABEL I.

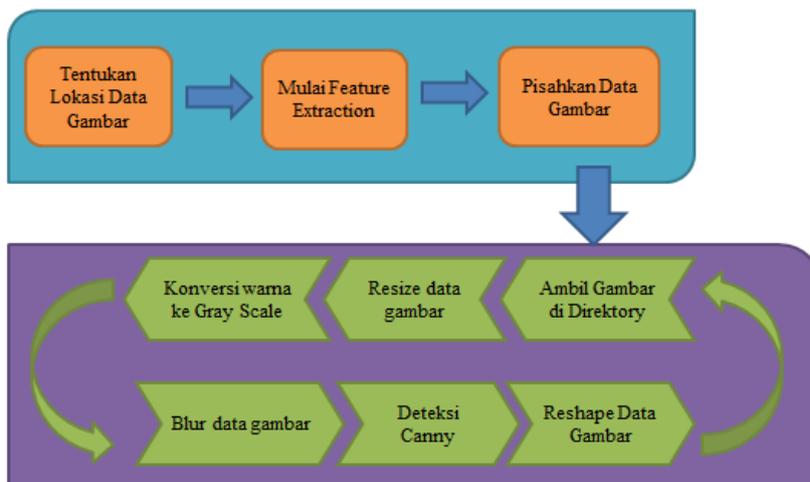
TABEL I
 CONTOH DATA KOMBINASI ARAH CAHAYA DAN LATAR

| | | Latar Belakang | |
|--------|---------------|---|---|
| | | Hitam | Putih |
| Cahaya | Depan |  |  |
| | Atas |  |  |
| | Samping kanan |  |  |
| | Ruangan |  |  |

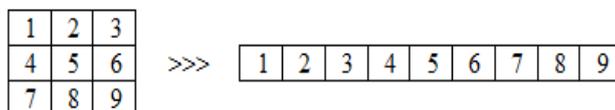
B. Training Classifier

Terdapat beberapa tahap dalam proses *training* dan *testing* pada *computer vision* (CV). Tahapan tersebut di antaranya proses melihat, mempelajari, dan mengenali. Dalam proses melihat, CV perlu memperoleh data yang spesifik berupa fitur dari objek yang akan dikenali. Mempelajari dalam CV dapat berupa proses klasifikasi dari fitur yang telah diperoleh dan menghasilkan suatu model klasifikasi (*classifier*). Dari model klasifikasi yang diperoleh, CV baru dapat mengenali objek yang telah dipelajari.

Tahap dalam proses *feature extraction* (melihat) dapat di lihat pada diagram Gambar 4. Tahap pertama adalah *me-load* kembali data gambar yang telah diperoleh untuk diproses. berikutnya data yang telah di-*load* akan dipisahkan menjadi 30% sebagai data *testing* dan 70% sebagai data *training*. kemudian gambar akan diubah resolusinya dengan perbandingan lebar 48 *pixel* dan diperoleh fitur gambar berupa garis tepi. Hasil garis tepi gambar ini akan berbentuk data martriiks dua dimensi dan akan diubah menjadi satu baris data *array* seperti diilustrasikan Gambar 5. Satu baris data *array* ini adalah fitur data gambar yang akan digunakan sebagai data *training* dan *testing*.

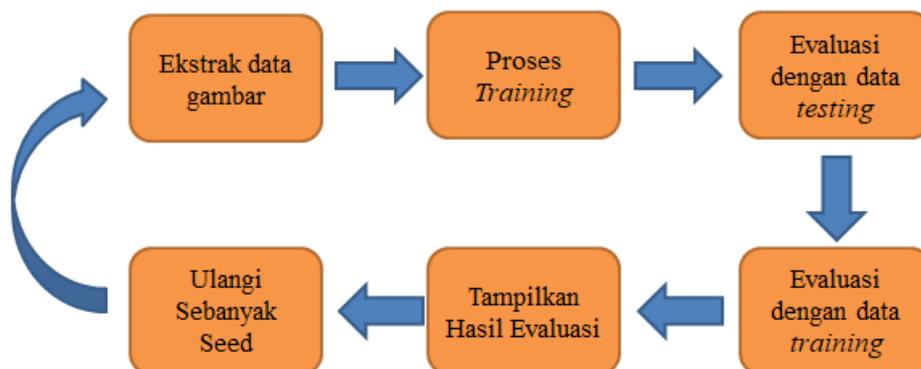


Gambar 4. Diagram alir proses deskripsi gambar



Gambar 5. Reshaping matriks menjadi 1 baris data array

Proses yang dilakukan setelah memperoleh ekstrak data gambar adalah proses *training* dan *testing classifier* SVM seperti diagram Gambar 6. Proses *training* akan diikuti oleh *testing* hasil klasifikasi yang diperoleh. Hasil dari data *testing* akan digunakan untuk evaluasi akurasi hasil klasifikasi.



Gambar 6 Diagram alir proses *training* SVM

C. Evaluasi

Proses klasifikasi dalam *machine learning* dengan metode *supervised learning* terdiri dari proses *training* dan proses *testing*. Proses *training* dilakukan untuk memperoleh model klasifikasi dari data yang telah diberi label dan proses *testing* bertujuan menguji hasil klasifikasi menggunakan data tanpa diberi label. Dari proses *testing* ini dapat dievaluasi nilai akurasi yang dihasilkan dari sebuah pengklasifikasi (*classifier*). Guna mengetahui nilai akurasi dari *classifier* maka diperlukan metode untuk melakukan evaluasi. Dalam proyek ini penulis menerapkan metode *confusion matrix* dalam menemukan nilai akurasi untuk menilai keberhasilan klasifikasi.

IV. HASIL DAN DISKUSI

A. Perbandingan Evaluasi

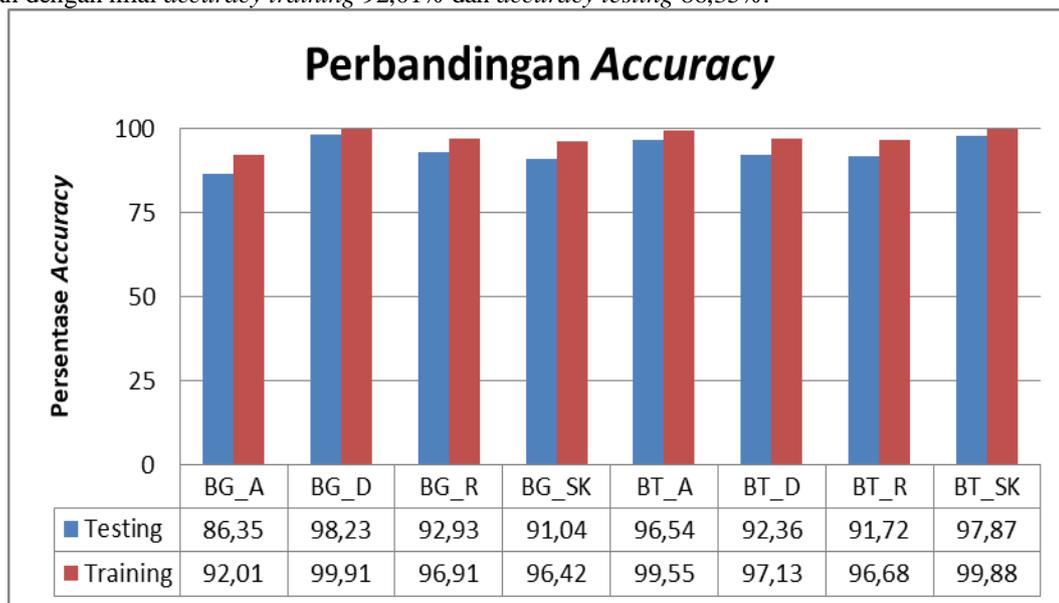
Pada proyek ini penulis memanfaatkan *confusion matrix* untuk melakukan evaluasi. Evaluasi dilakukan untuk setiap kombinasi lingkungan (arah datang cahaya dan warna latar). Kombinasi tersebut di antaranya :

- Latar putih dan cahaya atas (BT_A).
- Latar putih dan cahaya depan (BT_D).
- Latar putih dan cahaya ruangan (BT_R).
- Latar putih dan cahaya samping kanan (BT_SK).
- Latar hitam dan cahaya atas (BG_A).
- Latar hitam dan cahaya depan (BG_D).
- Latar hitam dan cahaya ruangan (BG_R).
- Latar hitam dan cahaya samping kanan (BG_SK).

Setiap kombinasi menghasilkan 10 *classifier* dengan data *training* yang diacak dan akan dicari nilai *confusion matrix*-nya. Dari *confusion matrix* yang diperoleh akan dicari nilai rata – rata *accuracy*, *recall*, *precision*, *f1 score*, dan *false positive rate* (FPR) dari keseluruhan kelas dalam satu pengulangan. Dari hasil rata – rata satu pengulangan kelima nilai (*accuracy*, *recall*, *precision*, *f1 score*, dan FPR) akan dihitung nilai rata – rata dari keseluruhan pengulangan *training*. Kelima nilai ini akan digunakan sebagai pembandingan untuk menentukan lingkungan mana yang baik digunakan untuk mendeteksi isyarat.

B. Perbandingan Accuracy

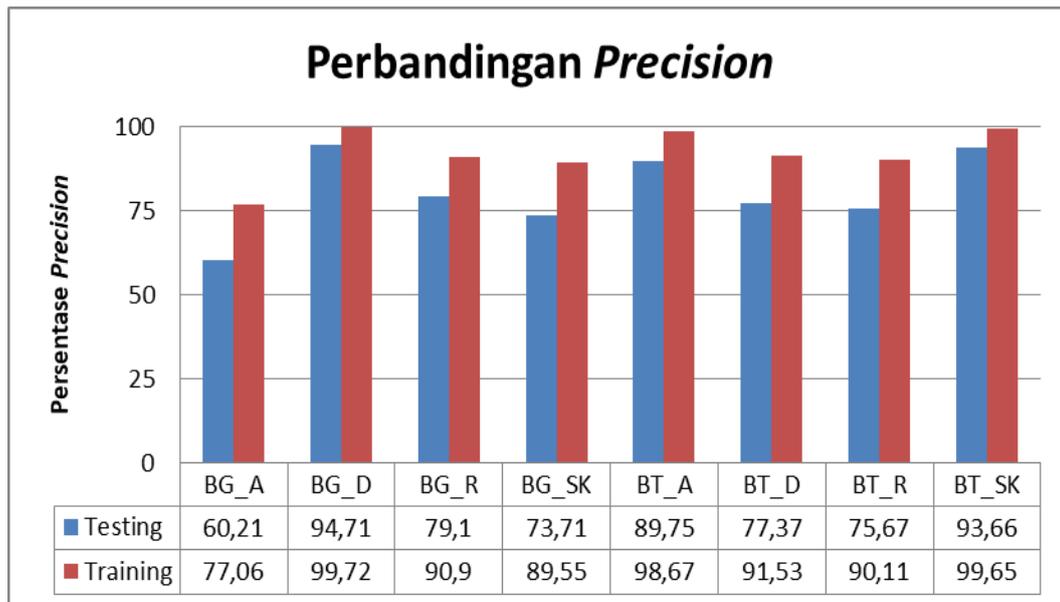
Gambar 7 menunjukkan hasil rata – rata *accuracy* di tiap pengulangan kombinasi. Dapat dilihat lingkungan BG_D memiliki nilai rata – rata *accuracy* yang paling baik dibandingkan lingkungan yang lain dengan nilai *accuracy training* 99,91% dan *accuracy testing* 98,23%. Di samping itu, kondisi lingkungan BG_A menghasilkan nilai rata – rata *accuracy* paling rendah dengan nilai *accuracy training* 92,01% dan *accuracy testing* 86,35%.



Gambar 7. Diagram perbandingan *accuracy*

C. Perbandingan Precision

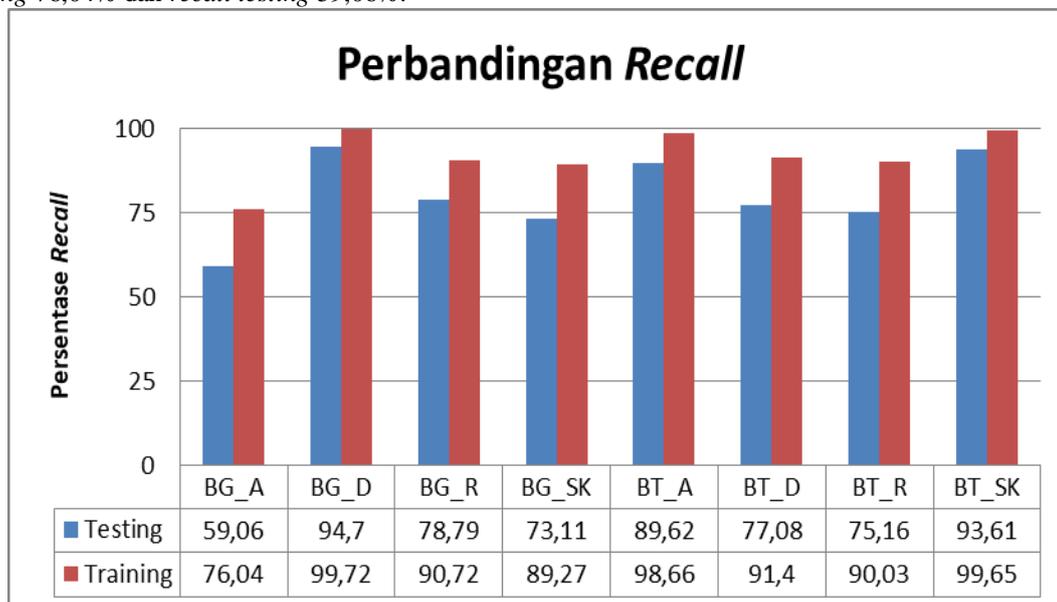
Gambar 8 menunjukkan hasil rata – rata *precision* di tiap pengulangan kombinasi. Dapat dilihat lingkungan BG_D memiliki nilai rata – rata *precision* yang paling baik dibandingkan lingkungan yang lain dengan nilai *precision training* 99,72% dan *precision testing* 94,71%. Di samping itu, kondisi lingkungan BG_A menghasilkan nilai rata – rata *precision* paling rendah dengan nilai *precision training* 77,06% dan *precision testing* 60,21%.



Gambar 8. Diagram perbandingan *precision*

D. Perbandingan Recall

Gambar 9 menunjukkan hasil rata – rata *recall* di tiap pengulangan kombinasi. Dapat dilihat lingkungan BG_D memiliki nilai rata – rata *recall* yang paling baik dibandingkan lingkungan yang lain dengan nilai *recall training* 99,72% dan *recall testing* 94,7%. Di samping itu, kondisi lingkungan BG_A menghasilkan nilai rata – rata *recall* paling rendah dengan nilai *recall training* 76,04% dan *recall testing* 59,06%.

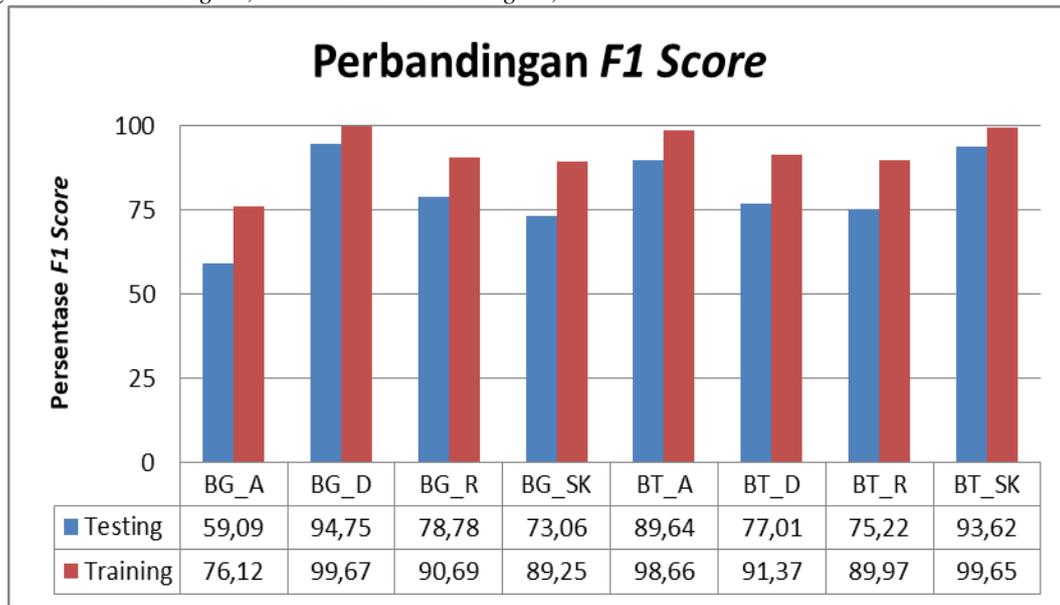


Gambar 9. Diagram perbandingan *recall*

E. Perbandingan F1 Score

Gambar 10 menunjukkan hasil rata – rata *F1 Score* di tiap pengulangan kombinasi. Dapat dilihat lingkungan BG_D memiliki nilai rata – rata *F1 Score* yang paling baik dibandingkan lingkungan yang lain dengan nilai *F1 training* 99,67%

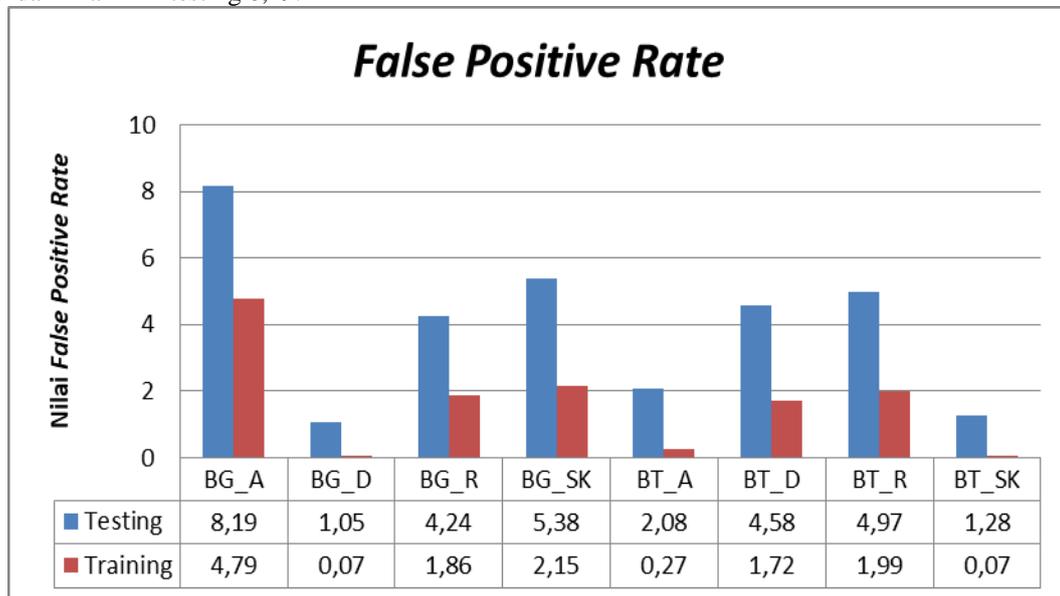
dan nilai *F1 testing* 94,75%. Di samping itu, kondisi lingkungan BG_A menghasilkan nilai rata – rata *F1 Score* paling rendah dengan nilai *F1 training* 76,12% dan nilai *F1 testing* 59,09%.



Gambar 10. Diagram perbandingan *f1 score*

F. Perbandingan False Positive Rate (FPR)

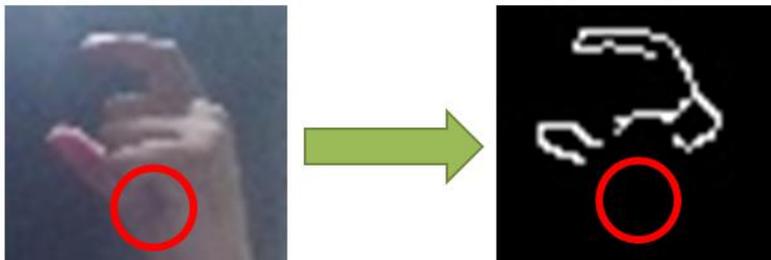
Gambar 11 menunjukkan hasil rata – rata FPR di tiap pengulangan kombinasi. Dapat dilihat lingkungan BG_D memiliki nilai rata – rata FPR yang paling rendah dibandingkan lingkungan yang lain dengan nilai FPR training 0,07 dan nilai FPR testing 1,05. Di samping itu, kondisi lingkungan BG_A menghasilkan nilai rata – rata FPR paling tinggi dengan nilai FPR training 4,79 dan nilai FPR testing 8,19.



Gambar 11. Diagram perbandingan *false positive rate*

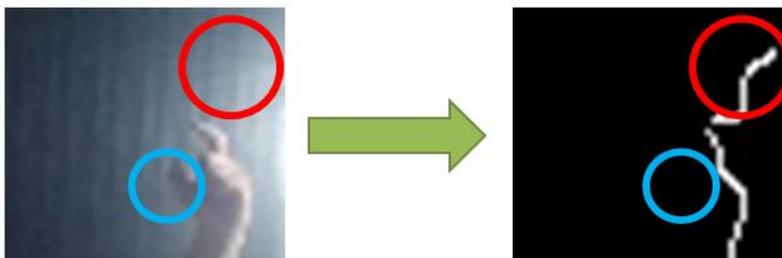
G. Analisis

Dari diagram Gambar 12 kondisi BG_A memiliki nilai akurasi paling minimal. Hal ini dipengaruhi oleh warna latar yang gelap dan arah datang cahaya dari atas. Pada objek (isyarat tangan) di lingkungan ini memiliki sisi gelap yang menyebabkan *Canny* tidak dapat mendeteksi garis tepi dengan sempurna seperti bagian yang dilingkari merah pada ilustrasi Gambar 12. Dengan garis tepi yang tidak sempurna mempengaruhi proses *training* dan *testing* karena data yang akan digunakan tidak mewakili objek yang diinginkan.



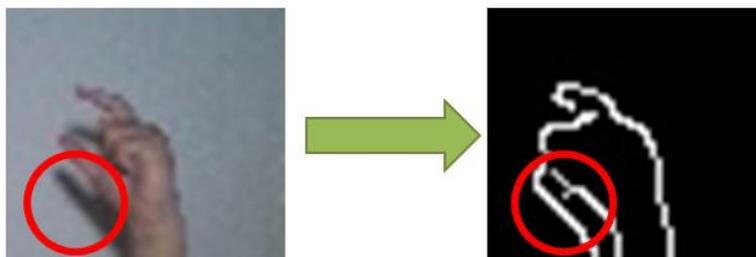
Gambar 12. Hasil *canny* BG_A

Sama halnya pada kondisi BG_SK yang menyebabkan objek memiliki sisi gelap di satu sisi dan terang di sisi lain (sisi terkena cahaya) seperti yang dilingkari biru pada ilustrasi Gambar 13. Selain itu, kondisi permukaan latar yang seolah memantulkan cahaya seperti bagian yang dilingkari merah pada Gambar 13 menyebabkan *Canny* memasang garis tepi.

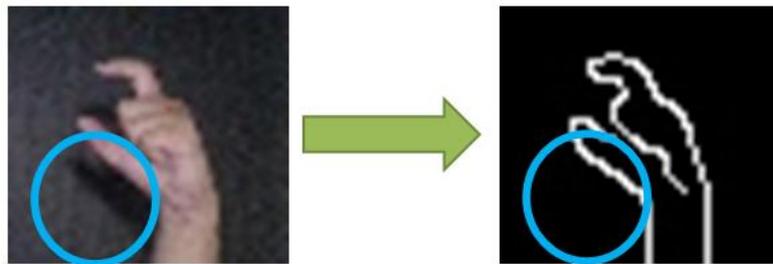


Gambar 13. Hasil *canny* BG_SK

Begitu pula hasil pada kondisi BT_D yang dipengaruhi oleh warna latar yang terang dan arah datang cahaya dari depan. Pada kombinasi ini menyebabkan bayangan muncul pada latar gambar dan bayangan ini ikut terdeteksi oleh *Canny* seperti bagian yang dilingkari merah pada ilustrasi Gambar 14. Berbeda pada hasil kondisi BG_D yang memiliki warna latar gelap dan arah datang cahaya dari depan. Pada kombinasi ini bayangan tetap muncul namun tidak terdeteksi oleh *Canny* seperti yang dilingkari biru pada Gambar 15. Oleh sebab itu, bayangan yang ikut terdeteksi oleh *Canny* menyebabkan perubahan bentuk dari objek yang diinginkan.



Gambar 14. Hasil *canny* BT_D



Gambar 15. Hasil *canny* BG_D

V. KESIMPULAN

Sistem yang dibangun dapat menerjemahkan bahasa isyarat secara *real-time* dari isyarat C, I, L, O, U, dan V menggunakan isyarat BISINDO. Aplikasi yang dibangun juga dapat melakukan pengumpulan data dan *training* classifier dari data yang telah terbentuk. Guna mengetahui evaluasi dari *classifier* yang terbentuk aplikasi yang dibangun juga dapat melakukan evaluasi yang akan menghasilkan nilai *accuracy*, *precision*, *recall* *f1 score*, dan *false positive rate* setiap melakukan *training*.

Pada hasil eksplorasi diperoleh kondisi lingkungan yang digunakan berpengaruh pada hasil klasifikasi yang diperoleh. Latar yang memiliki kecerahan warna lebih rendah dibandingkan objek atau sebaliknya, dapat menghasilkan proses ekstraksi yang lebih baik dengan *Canny*. Pada arah datang cahaya akan mempengaruhi perubahan sisi gelap dan terang dari objek. Selain itu, arah datang cahaya dapat memberikan bayangan pada latar yang mengakibatkan bayangan ikut terdeteksi oleh *Canny*.

Dari kedelapan kombinasi kondisi lingkungan yang dieksplorasi, diperoleh kondisi lingkungan latar hitam dan arah cahaya depan (BG_D) memiliki nilai akurasi yang paling baik dengan nilai rata – rata akurasi data *testing* 98,2% dan *training* 99,9%. Hal ini dikarenakan gambar hasil proses *feature extraction* memiliki *feature* yang memiliki bentuk (mewakili) isyarat tangan (objek) yang dideteksi. Selain itu, bayangan yang muncul pada kondisi ini tidak berdampak pada hasil *feature extraction* yang diperoleh.

DAFTAR PUSTAKA

- [1] I. Hussain, A. K. Talukdar & K. Ku, "Hand Gesture Recognition System with Real-Time Palm Tracking," dalam 2014 Annual IEEE India Conference (INDICON), Pune, 2014.
- [2] T. . A. B. Wirayuda, H. A. Adhi, D. H. Kuswanto dan R. . N. Dayawati, "Real-Time Hand-Tracking on Video Image Based on Palm Geometry," dalam 2013 International Conference of Information and Communication Technology (ICoICT), Bandung, 2013.
- [3] A. Dhawan dan V. Honrao, "Implementation of Hand Detection based Techniques for Human Computer Interaction," International Journal of Computer Applications (0975 – 8887), vol. LXXII, no. 17, pp. 6-13, 2013.
- [4] J. S. Artal-Sevil dan J. L. Montañés, "Development of a robotic arm and implementation of a control strategy for gesture recognition through Leap Motion device," dalam 2016 Technologies Applied to Electronics Teaching (TAEE), Seville, 2016.
- [5] G. Bradski dan A. Kaehler, Learning OpenCV, Sebastopol: O'Reilly Media, Inc, 2016.
- [6] A. Yilmaz, O. Javed dan M. Shah, "Object Tracking: A Survey," ACM Computing Surveys, vol. XXXVIII, no. 4, pp. 1-45, 2006.
- [7] C. Campbell dan Y. Ying, Learning with Support Vector Machines, Morgan & Claypool Publishers, 2011.
- [8] R. . B. Dan dan P. . S. Mohod, "Survey on Hand Gesture Recognition Approaches," International Journal of Computer Science and Information Technologies (IJCSIT), vol. V, no. 2, pp. 2050-2052, 2014.
- [9] C. J. Enns dan . R. C. Herman, "Adapting the Assessing British Sign Language Development: Receptive Skills Test Into American Sign Language," Journal of Deaf Studies and Deaf Education, vol. XVI, no. 3, pp. 362-374, 2011.
- [10] P. A. Nanivadekar dan D. V. Kulkarni, "Indian Sign Language Recognition: Database Creation, Hand Tracking and Segmentation," dalam International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA), Mumbai, 2014.
- [11] R. A. Mursita, "Respon Tunarungu Terhadap Penggunaan Sistem Bahasa Isyarat Indonesia (SIBI) dan Bahasa Isyarat Indonesia (BISINDO) dalam Komunikasi," INKLUSI, vol. II, no. 2, pp. 221-232, 2015.
- [12] W. S. Pambudi dan B. M. N. Simorangkir, "Facetracker Menggunakan Metode Haar Like Feature dan Pid pada Model Simulasi," Jurnal Teknologi dan Informatika (TEKNOMATIKA), vol. II, no. 2, pp. 142-154, 2012.
- [13] I. Hussain, A. K. Talukdar & K. Ku, "Hand Gesture Recognition System with Real-Time," dalam India Conference (INDICON), India, 2014.