

PENGENALAN BAHASA ISYARAT INDONESIA (BISINDO) MENGGUNAKAN ALGORITMA SCALE INVARIANT FEATURE TRANSFORM (SIFT) DAN CONVOLUTIONAL NEURAL NETWORK (CNN)

Ayu Widya Agata, Wahyu S J Saputra, Chrystia Aji Putra

Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur
Jalan Raya Rungkut Madya No. 1, Gunung Anyar, Surabaya, Indonesia
19081010005@student.upnjatim.ac.id

ABSTRAK

Bahasa isyarat berperan penting sebagai media komunikasi bagi tuna rungu dan tuna wicara. Namun, keterbatasan pemahaman masyarakat umum terhadap bahasa isyarat menjadi hambatan dalam interaksi dengan mereka. Untuk mengatasi masalah ini, dirancang suatu sistem pengenalan bahasa isyarat yang dapat membantu masyarakat umum berkomunikasi dengan mudah. Sistem ini mengenali alfabet Bahasa Isyarat Indonesia (BISINDO) dengan menggabungkan algoritma Scale Invariant Feature Transform (SIFT) dan Convolutional Neural Network (CNN). Pendekatan ini dipilih karena SIFT digunakan untuk ekstraksi fitur awal, sementara CNN untuk mengenali pola yang kompleks, meningkatkan kekuatan model terhadap variasi gerakan dan pencahayaan dalam bahasa isyarat. Dengan menggunakan algoritma SIFT dan CNN gerakan tangan Bahasa isyarat Indonesia dapat diperoleh, dicocokkan, dikenali, diautentikasi dan kinerja pencocokannya disimulasikan menggunakan library tensorflow dan OpenCV. Dalam penelitian, beberapa gerakan memerlukan percobaan berulang untuk deteksi yang akurat, mengingat kemiripan dengan gerakan bahasa isyarat lain. Meskipun demikian, akurasi tertinggi tercapai pada model dengan nilai epoch 55, mencapai 99.78%. Dengan demikian, integrasi algoritma SIFT dan CNN dalam sistem pengenalan bahasa isyarat dapat menjadi solusi efektif untuk mengatasi hambatan komunikasi antara tuna rungu, tuna wicara, dan masyarakat umum.

Kata kunci: Bahasa Isyarat Indonesia (BISINDO); Pengenalan Bahasa Isyarat; Scale Invariant Feature Transform (SIFT); Convolutional Neural Network (CNN)

1. PENDAHULUAN

Bahasa isyarat merupakan gerak yang menggunakan tubuh atau fisik untuk menyampaikan informasi [1]. Bahasa isyarat merupakan salah satu media komunikasi bagi para tuna rungu dan tuna wicara dalam berkomunikasi satu sama lain ataupun ke masyarakat umum. Bahasa isyarat sangat berguna untuk berkomunikasi dalam kehidupan sehari-hari karena sistem ini dapat menjadikan lawan bicara memahami apa yang disampaikan dengan mengenali pola alfabet yang dibentuk dengan jari sebagai isyarat dalam menyusun kata. Akan tetapi, tidak semua masyarakat umum memiliki kemampuan untuk mengerti dalam berkomunikasi dengan bahasa isyarat yang mengakibatkan keterbatasan tuna rungu dan tuna wicara dalam berkomunikasi dengan masyarakat.

Masalah tersebut dapat diatasi dengan suatu sistem yang dapat mengenali pola alfabet dalam bahasa isyarat sehingga dapat membantu masyarakat umum dalam berkomunikasi dengan bahasa isyarat dengan mudah. BISINDO adalah salah satu bahasa isyarat Indonesia yang belum memiliki banyak fasilitas untuk diterapkan. Hal tersebut dapat menyebabkan penyandang tunarungu mengalami kesulitan dalam menjalani kehidupan sehari-hari [2]. Oleh karena itu, penelitian ini mencoba menawarkan sistem pengenalan (*Recognition*) atau penerjemahan abjad BISINDO ke dalam sebuah huruf. Sistem ini diharapkan dapat membantu penyandang tuna rungu maupun tuna wicara untuk berkomunikasi dua arah.

Dalam penelitian ini akan dilakukan pengenalan alfabet Bahasa Isyarat Indonesia (BISINDO) dengan menggunakan perpaduan dua algoritma, yaitu algoritma *Scale Invariant Feature Transform* (SIFT) dan juga algoritma *Convolutional Neural Network* (CNN). Algoritma SIFT (*Scale Invariant Feature Transform*) merupakan salah satu algoritma yang berperan dalam pengenalan objek yang tahan dan efektif terhadap perubahan rotasi dan penskalaan [3]. Algoritma ini bekerja dengan menerapkan titik kunci SIFT objek yang pertama kali diekstraksi dari satu set gambar referensi dan disimpan dalam database. Selanjutnya, menetapkan beberapa parameter untuk mengoptimalkan dan menghilangkan titik-titik pencocokan yang salah. Dan dari set kecocokan tersebut, gambar baru diidentifikasi untuk menentukan kecocokan yang tepat. Algoritma *Convolutional Neural Network* (CNN) adalah algoritma Deep Learning yang dapat menerima gambar masukan, memberikan pentingnya (berat dan bias yang dapat dipelajari) untuk berbagai aspek/objek dalam gambar, dan mampu membedakan satu dengan yang lain [4]. CNN dapat mengklasifikasi gambar dengan memproses gambar yang diinput, kemudian mengklasifikasikannya pada kategori tertentu.

Dengan memadukan algoritma *Scale Invariant Feature Transform* (SIFT) dan algoritma *Convolutional Neural Network* (CNN) pada penelitian ini, diharapkan dapat mendapatkan hasil pencocokan yang maksimal pada gerakan tangan bahasa isyarat.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Dalam penelitian ini, terdapat beberapa penelitian terdahulu yang dijadikan acuan, acuan pertama yang digunakan adalah Jurnal Indonesian *Sign Language Number Recognition using SIFT Algorithm* oleh Isa Mahfudi, Moechammad Sarosa, Rosa Andrie Asmara, M. Azrino Gustalika. Jurnal ini membahas tentang penggunaan algoritma SIFT untuk pengenalan bahasa isyarat Indonesia dan menghasilkan tingkat pengenalan menjadi 82% dengan penggunaan total 100 sampel dataset citra yang terdiri dari 50 sampel untuk data latih dan 50 sampel citra untuk data pengujian. Jurnal ini juga membahas tentang perubahan nilai threshold yang dapat mempengaruhi hasil pengenalan. Ambang batas nilai terbaik adalah 0,45 dengan tingkat pengenalan 94% [5].

Selanjutnya terdapat Jurnal *Sign Language Recognition System using Convolutional Neural Network and Computer Vision* yang oleh Mehreen Hurroo dan Mohammad Elham Walizad [6]. Pada jurnal ini membahas tentang pengenalan Bahasa Isyarat menggunakan Bahasa Isyarat Amerika. Dalam jurnal ini, pengguna harus dapat menangkap gambar isyarat tangan menggunakan kamera web, dan sistem akan memprediksi dan menampilkan nama gambar yang diambil. Penulis menggunakan algoritma warna HSV untuk mendeteksi gerakan tangan dan menyetel latar belakang menjadi hitam. Fitur yang diekstrak adalah piksel biner dari citra. Penulis menggunakan *Convolutional Neural Network* (CNN) untuk melatih dan mengklasifikasikan gambar. Hasil yang didapatkan yaitu dapat mengenali sepuluh huruf gerakan Tanda Amerika dengan akurasi tinggi. Model yang penulis gunakan telah mencapai akurasi yang luar biasa yaitu di atas 90%.

Selain kedua jurnal diatas, pada tahun 2021, terdapat penelitian yang dilakukan oleh Mohammad Idham Fachrurrozi [7] tentang Implementasi Objek *Localization* dengan Metode CNN untuk Deteksi *Indonesian Sign Language* BISINDO. Dalam penelitian tersebut Fachrurrozi melakukan pengembangan sistem pengenalan dengan menggunakan metode *Convolutional Neural Network* yang merupakan salah satu jenis algoritma Jaringan Syaraf Tiruan dalam melakukan klasifikasi sedangkan pada deteksi citra, Fachrurrozi menggunakan *Regional-CNN* yang merupakan pengembangan CNN. Data masukan dalam penelitian ini merupakan dataset dari ISLBISINDO1 [8] yang berupa citra dan mempunyai data sebanyak 2080 citra dengan *abjad* bahasa isyarat yang mempunyai 26 kelas. Pada penelitian ini, Fachrurrozi berhasil mencapai tingkat akurasi tertinggi sebesar 97% dalam beberapa percobaan, menunjukkan bahwa jumlah epoch yang signifikan selama pelatihan memiliki dampak besar pada peningkatan kinerja akurasi. Perbedaan akurasi antara uji coba, yaitu 97% untuk epoch 50, 83% untuk epoch 30, dan 72% untuk epoch 15.

Ketiga penelitian yang disebutkan memiliki fokus yang serupa, yaitu pengenalan Bahasa Isyarat. Mereka menggunakan teknik pengolahan citra dan pengenalan pola sebagai fondasi utama untuk mencapai tujuan pengenalan isyarat. Namun, perbedaan mendasar terletak pada metode yang digunakan. Isa menggunakan algoritma SIFT, suatu pendekatan yang lebih tradisional dalam pengolahan citra, sementara Maureen mengadopsi pendekatan yang lebih modern dengan penggunaan *Convolutional Neural Network* (CNN), memanfaatkan deteksi gerakan tangan dan piksel biner. Sementara itu, Fachrurrozi melibatkan penggunaan CNN dan *Regional-CNN* untuk mendeteksi dan mengklasifikasikan isyarat.

Perbedaan juga terdapat pada dataset dan jumlah kelas yang digunakan dalam setiap penelitian, yang memiliki variasi tergantung pada konteks spesifik dari penelitian tersebut. Terakhir, fokus utama dari masing-masing penelitian juga berbeda, mulai dari penyesuaian nilai *threshold* pada penelitian Isa, implementasi kamera web pada Maureen, hingga penekanan pada pengaruh jumlah *epoch* pada penelitian Fachrurrozi. Dengan demikian, ketiga penelitian ini menunjukkan pendekatan yang beragam dalam mengatasi tantangan pengenalan Bahasa Isyarat, dari metode klasik hingga teknologi jaringan saraf tiruan modern.

2.2. Pengenalan

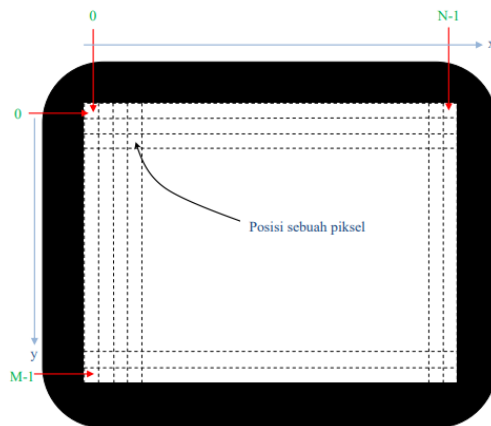
Pengenalan atau *recognition* memiliki tujuan bukan hanya mendapatkan citra dengan kualitas tertentu, melainkan juga mengklasifikasikan berbagai jenis citra. Dalam proses ini, sejumlah citra diolah sehingga citra dengan ciri yang serupa dikelompokkan ke dalam kategori tertentu. Interpretasi melibatkan penekanan pada pemahaman objek yang teridentifikasi. Pengenalan pola mencakup sejumlah teknik dan metode untuk mengenali struktur atau karakteristik tertentu dalam data [9].

Dalam perkembangan lebih lanjut, pengenalan atau *recognition*, *image processing* dan *computer vision* telah berfungsi sebagai mata manusia. Perangkat input *image capture*, seperti kamera dan scanner, dianggap sebagai mata, sementara mesin komputer dengan program komputasinya dianggap sebagai otak yang bertanggung jawab untuk memproses informasi. Hal ini menghasilkan berbagai bidang penting dalam *computer vision*, termasuk *pattern recognition* (pengenalan pola), *biometric* (pengenalan identifikasi manusia berdasarkan ciri-ciri biologis pada tubuh manusia), *content-based image and video retrieval* (pengambilan kembali citra atau video dengan informasi tertentu), *video editing*, dan berbagai bidang lainnya.

2.3. Citra Digital

Citra digital merujuk pada representasi visual yang terdiri dari piksel-piksel diskrit, di mana setiap piksel memiliki nilai intensitas yang

merepresentasikan warna atau tingkat kecerahan pada posisi tertentu. Menurut "Computer Vision: Algorithms and Applications" oleh Richard Szeliski [10], konsep ini terlibat dalam pemrosesan dan analisis citra untuk aplikasi penglihatan komputer. Setiap piksel digambarkan sebagai satu kotak kecil. Setiap piksel mempunyai koordinat posisi. Sistem koordinat yang dipakai untuk menyatakan citra digital ditunjukkan pada Gambar 1.



Gambar 1. Sistem koordinat citra berukuran $M \times N$ yaitu baris x kolom [11]

Berdasarkan gambar 1, dengan sistem koordinat yang mengikuti asas pemindaian pada layar TV standar itu, sebuah piksel mempunyai koordinat berupa (x, y)

Dalam hal ini,

- x menyatakan posisi kolom
- y menyatakan posisi baris
- piksel pojok kiri-atas mempunyai koordinat $(0, 0)$ dan piksel pada pojok
- kanan-bawah mempunyai koordinat $(N-1, M-1)$

2.4. Bahasa Isyarat Indonesia (BISINDO)

Bahasa isyarat adalah satu kaedah komunikasi yang menggunakan symbol-simbol tanpa menggunakan suara atau yang sering dikenal sebagai 'non-verbal communication' [12]. Isyarat-isyarat yang digunakan dapat berupa gerakan tangan, bagian tubuh lainnya, ekspresi wajah, gambar, atau simbol-simbol yang memiliki makna khusus dan dapat dimengerti oleh kedua belah pihak, yaitu penutur dan penerima.

Dalam Kamus Besar Bahasa Indonesia (KBBI), istilah "bahasa isyarat" merujuk pada suatu bentuk komunikasi yang tidak melibatkan penggunaan kata-kata atau tulisan, melainkan menggunakan sistem simbol. Setiap daerah di Indonesia memiliki bahasa isyaratnya sendiri, dan perbedaan signifikan dapat ditemukan antara bahasa isyarat dari satu tempat ke tempat lain. Terdapat kendala dalam pemahaman bahasa isyarat yang berbeda, sehingga muncul ide untuk menciptakan sebuah sistem bahasa isyarat yang seragam untuk Indonesia, disingkat sebagai BISINDO.



Gambar 2. Kumpulan abjad BISINDO [13]

Pada gambar 2 memperlihatkan bahwa bahasa isyarat Indonesia BISINDO menggunakan kombinasi kedua tangan untuk mengekspresikan isyaratnya. Diharapkan bahwa Bahasa Isyarat Indonesia (BISINDO) akan menjadi bahasa isyarat nasional dan juga digunakan secara luas dalam bahasa Indonesia. BISINDO telah memiliki struktur tata bahasa khusus yang dapat dipelajari saat ini melalui berbagai media atau dari individu yang memiliki pengalaman dalam Bahasa Isyarat Indonesia (BISINDO).

2.5. Scale Invariant Feature Transform

Scale Invariant Feature Transform (SIFT) adalah algoritma dalam bidang visi komputer yang digunakan untuk menemukan dan menggambarkan fitur lokal pada gambar. Algoritma ini dikembangkan oleh David Lowe pada tahun 1999. Dengan menerapkan SIFT, sebuah gambar dapat diubah menjadi vektor fitur lokal yang nantinya digunakan sebagai pendekatan untuk mendeteksi dan mengenali objek tertentu melalui titik-titik atau keypoint. Keypoint ini berfungsi sebagai representasi fitur dari gambar target, terutama dalam konteks *augmented reality*.

Metode Scale-Invariant Feature Transform (SIFT) menjadi pilihan unggul dalam ekstraksi fitur untuk pengenalan objek. Keunggulan utamanya melibatkan kemampuan menjaga hasil ekstraksi fitur stabil terhadap perubahan ukuran, translasi, dan rotasi 2D.

SIFT juga mampu melakukan banyak ekstraksi fitur pada citra dengan ciri khusus, menjadikannya fleksibel dan andal. Hasil ekstraksi fitur SIFT yang benar-benar mencirikan secara khusus menjamin keakuratan dan keunikan, membuatnya sangat efektif dalam tugas pengenalan objek. Keseluruhan, SIFT merupakan metode yang handal dan efisien dalam mengatasi variasi visual dalam pengolahan citra.

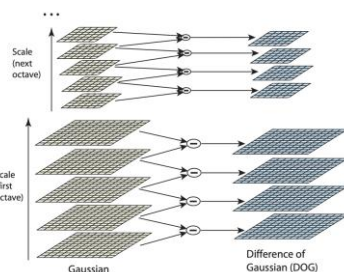
Menurut David Lowe [14], Algoritma Scale Invariant Feature Transform (SIFT) terbagi menjadi 4 (empat) tahapan, yaitu:

- Scale-space Extrema Detection (Pencarian Nilai Ekstrem pada Skala Ruang)
- Keypoint Localization (Penentuan Keypoint)

- c. Orientation Assignment (Penentuan Orientasi)
- d. Keypoint Descriptor (Deskriptor Keypoint)

Tahapan SIFT (*Scale-Invariant Feature Transform*) melibatkan empat langkah utama yang saling terkait untuk ekstraksi fitur invarian skala pada citra. Pertama, dalam *Scale-space Extrema Detection*, citra diolah pada berbagai tingkat skala dengan menggunakan piramida *Gaussian* untuk mendeteksi nilai ekstrim, yang merupakan potensi titik kunci yang invarian terhadap perubahan skala.

Selanjutnya, dilakukan perhitungan *Difference of Gaussian (DoG)* dengan mengurangkan tingkat-tingkat berturut-turut dari piramida tersebut. Lokal maksima dan minima dalam gambar DoG kemudian dianggap sebagai titik kunci potensial, yang mewakili fitur-fitur yang khas baik dari segi posisi maupun skala.

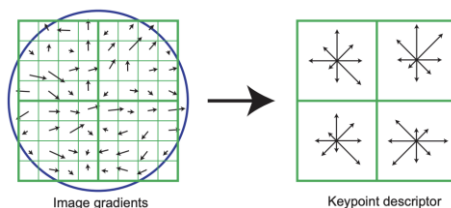


Gambar 3. Gaussian dan DOG [14]

Berdasarkan ilustrasi pada gambar 3, keypoint diambil dari maksimal atau minimal dari DoG (*Difference of Gaussians*). Dimana $L(x, y, \sigma)$ adalah konvolusi dari citra $I(x, y)$ dengan *Gaussian filter* $G(x, y, \sigma)$ sehingga mendapatkan hasil persamaan:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (1)$$

Langkah kedua, *Keypoint Localization*, melibatkan penentuan akurasi posisi titik kunci yang dihasilkan pada langkah sebelumnya dengan menghitung nilai *Taylor* dari citra pada lokasi tersebut. Setelahnya, langkah ketiga, *Orientation Assignment*, bertujuan untuk menentukan orientasi dari setiap titik kunci dengan memperhitungkan gradien *local* pada lingkungan sekitarnya, sehingga fitur menjadi invarian terhadap rotasi. Terakhir, pada langkah *Keypoint Descriptor*, dilakukan pembentukan deskriptor yang menggambarkan lingkungan piksel di sekitar titik kunci.



Gambar 4. Keypoint descriptor [14]

Sesuai pada gambar 4, deskriptor ini mencakup informasi tentang gradien dan arah gradien pada setiap piksel dalam area sekitar titik kunci, sehingga memungkinkan untuk pencocokan fitur antara citra.

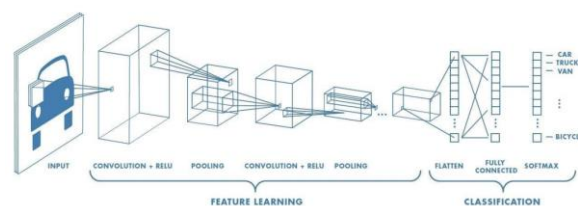
2.6. Convolutional Neural Network

Algoritma *Convolutional Neural Network (CNN)* adalah algoritma Deep Learning yang dapat menerima gambar masukan, memberikan pentingnya (berat dan bias yang dapat dipelajari) untuk berbagai aspek atau objek dalam gambar, dan mampu membedakan satu dengan yang lain [4]. Metode CNN merupakan suatu kelas dalam jaringan saraf yang secara khusus dirancang untuk memproses data dengan topologi seperti grid, seperti halnya pada gambar. CNN dapat diterapkan dalam berbagai bidang, seperti pengenalan wajah, analisis dokumen, klasifikasi gambar, klasifikasi video, dan lain sebagainya.

Metode CNN melakukan klasifikasi gambar dengan cara memproses gambar yang diinput dan kemudian mengkategorikannya ke dalam kelas tertentu, misalnya mengidentifikasi elemen-elemen seperti wajah, mata, bibir, hidung, tangan, dan sebagainya dalam citra manusia.

Proses ini melibatkan transformasi gambar menjadi array yang berisi nilai pada setiap piksel, dengan format resolusi tinggi x panjang x dimensi yang disebut sebagai channel. Channel ini umumnya terdiri dari tiga lapisan, yang mewakili warna *Red*, *Green*, dan *Blue* pada citra RGB, atau hanya satu lapisan untuk gambar *grayscale*. Namun, jumlah lapisan juga bisa melebihi tiga, bahkan mencapai ratusan, yang merepresentasikan berbagai warna dalam arsitektur RGB.

CNN terbentuk dari beberapa lapisan tersembunyi (*hidden layer*), melibatkan lapisan konvolusi, fungsi aktivasi (ReLU), lapisan *pooling*, dan lapisan *fully connected*. Pada Gambar 4, terlihat gambaran keseluruhan mengenai cara *Convolutional Neural Network* bekerja dalam menjalankan tugas klasifikasi citra.



Gambar 5. Cara kerja CNN [4]

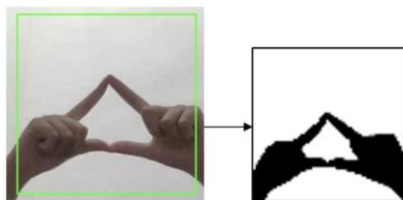
Berdasarkan Gambar 5, *Feature Learning* merupakan bagian awal dari *Convolutional Neural Network (CNN)*, yang bertugas mengubah input menjadi fitur dengan merinci ciri-ciri dari input tersebut dalam bentuk vektor berupa angka. Bagian ini terdiri dari *convolution layer* dan *pooling layer* dalam struktur CNN.

3. METODE PENELITIAN

3.1. Data Preparation

Mengikuti pendekatan 80:20 [15], dataset yang ada akan dialokasikan sekitar 80% atau 10.400 data sebagai subset latih dan 20% atau 2.600 data sebagai subset uji. Proses ini sangat penting agar kedua subset dapat mewakili dengan baik keseluruhan dataset. Penting untuk melaksanakan pembagian secara acak guna memastikan setiap subset memiliki karakteristik yang serupa dengan dataset utuh.

Selanjutnya, dilakukan *preprocessing data* dimana terbagi menjadi beberapa tahapan diantaranya adalah mengatur ukuran citra, mengubah satuan warna menjadi satuan HSV (*Hue, Saturation, Value*) dan langkah berikutnya akan melibatkan proses pengumpulan data, di mana penulis akan mengubah data mentah ke dalam format yang sesuai dan konsisten. Langkah ini diambil karena data mentah seringkali tidak lengkap dan memiliki format yang tidak konsisten. Upaya ini diperlukan karena kualitas data secara langsung memengaruhi keberhasilan setiap proyek yang melibatkan analisis data.



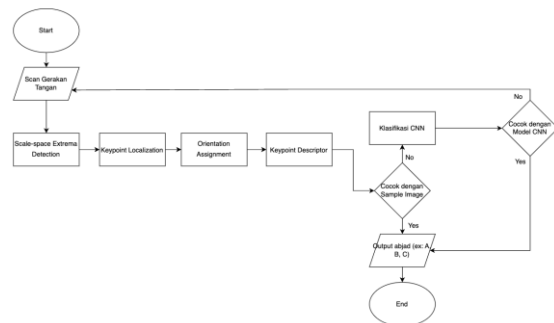
Gambar 6. Sample gambar sebelum dan sesudah dilakukan *preprocessing*

Sesuai pada gambar 6, setiap citra gerakan akan diproses menjadi satuan warna HSV dengan ukuran gambar 64×64 pixel dan akan tersimpan dalam format file .PNG. Perubahan skema warna menjadi skema warna HSV atau konversi citra ke citra hitam-putih dilakukan dengan tujuan tertentu sesuai dengan kebutuhan analisis atau pengolahan citra. Penggunaan skema warna HSV, dengan fokus pada saluran Hue yang menggambarkan warna, memungkinkan analisis yang lebih spesifik terhadap karakteristik warna tertentu dalam citra. Selain itu, konversi ke citra hitam-putih dapat bermanfaat untuk mengurangi dimensi warna, yang dapat meningkatkan efisiensi dan menghemat sumber daya komputasi, terutama ketika informasi warna tidak menjadi fokus utama.

Penggunaan citra hitam-putih juga berguna dalam menekankan struktur dan tekstur objek tanpa terpengaruh oleh variasi warna, sehingga dapat digunakan dalam konteks deteksi tepi atau pengenalan pola. Selain itu, konversi ini memudahkan pemrosesan lebih lanjut atau implementasi algoritma tertentu yang lebih efisien pada citra dengan representasi warna yang lebih sederhana. Keputusan untuk melakukan perubahan skema warna atau konversi ke citra hitam-putih seringkali didasarkan pada keperluan aplikasi atau tujuan analisis spesifik yang ingin dicapai.

3.2. SIFT Recognition

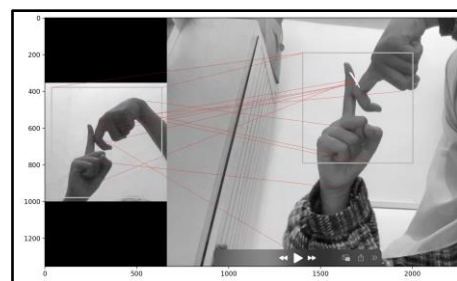
Pada proses pengenalan ini, hasil model klasifikasi menggunakan CNN akan diproses. Scan gerakan tangan kemudian akan dilakukan pencocokan dengan menerapkan titik kunci SIFT objek yang pertama kali diekstraksi dari satu set gambar referensi dan disimpan dalam database. Selanjutnya, menetapkan beberapa parameter untuk mengoptimalkan dan menghilangkan titik-titik pencocokan yang salah. Dan dari set kecocokan tersebut, gambar baru diidentifikasi untuk menentukan kecocokan yang tepat.



Gambar 7. Diagram alur SIFT

Berdasarkan diagram pada gambar 7, dalam pengenalan ini akan diimplementasikan metode SIFT yang mengkoordinasikan seluruh proses mendeteksi dan mengidentifikasi keypoint dalam gambar. Ini melakukan ini dengan mengonversi gambar masukan ke ruang skala *Gaussian*, mendeteksi keypoint dalam ruang skala, menghitung orientasi keypoint, menghasilkan deskriptor untuk setiap keypoint, menghapus keypoint duplikat, dan mengonversi keypoint ke ukuran gambar masukan. Akhirnya, fungsi ini mengembalikan daftar keypoint dan daftar deskriptor yang terkait. Melakukan prediksi gestur berdasarkan citra yang ditangkap kamera.

Pertama, citra dari kamera diolah dan disesuaikan ukurannya untuk dikirim sebagai input ke model jaringan saraf yang telah dilatih sebelumnya (CNN dari Keras). Selanjutnya, dilakukan perbandingan fitur SIFT antara citra yang ditangkap dan setiap citra gestur dalam direktori *sample-gestures*.



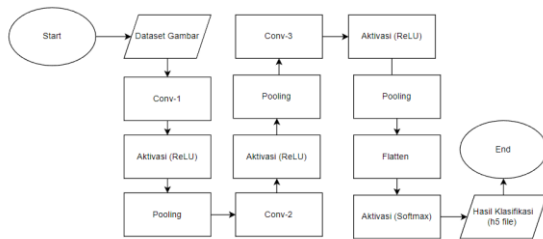
Gambar 8. Keypoint descriptor pada sample input

Berdasarkan kecocokan gambar 8 jika gambar sesuai maka akan muncul keypoint yang cocok antar gambar. Jika perbandingan fitur SIFT menghasilkan kecocokan yang signifikan, maka gestur yang dikenali

akan sesuai dengan citra tersebut. Jika tidak ada kecocokan yang signifikan dengan citra gestur, program akan menggunakan hasil prediksi dari model CNN.

3.3. CNN Classification

Kelas klasifikasi ini terdiri dari 26 kategori yang merupakan kumpulan abjad A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, dan Z. Hasil model klasifikasi disimpan dalam file berekstensi h5. Pada tahapan klasifikasi dilakukan implementasi menggunakan algoritma Convolutional Neural Network (CNN) dengan jumlah epoch 15, 25, 35, 45, 55, dan 65.

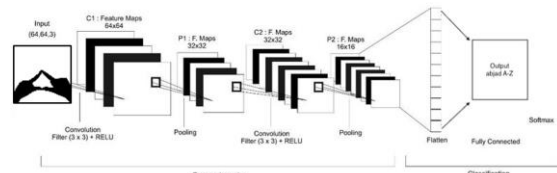


Gambar 9. Diagram alur CNN

Berdasarkan gambar 9, langkah pertama dalam pembuatan model klasifikasi ini adalah dengan melakukan penambahan lapisan konvolusi menggunakan Convolution2D dengan 32 filter berukuran 3x3, aktivasi ReLU, dan langkah (stride) 2x2 pada gambar input berukuran 64x64 piksel dengan 3 channel (RGB).

Setelahnya, lapisan MaxPooling2D diterapkan dengan ukuran pooling 2x2 untuk mengurangi dimensi fitur. Dua lapisan konvolusi berikutnya ditambahkan dengan masing-masing 64 filter 3x3 dan aktivasi ReLU, diikuti oleh lapisan MaxPooling2D. Lalu, hasilnya diubah menjadi vektor satu dimensi menggunakan Flatten untuk dapat dihubungkan ke lapisan tersembunyi Dense dengan 256 neuron dan aktivasi ReLU, diikuti oleh lapisan Dropout untuk mengurangi overfitting.

Akhirnya, lapisan output dengan 26 neuron (diasumsikan untuk klasifikasi 26 kelas) menggunakan aktivasi softmax. Penggunaan softmax di akhir bertujuan untuk menghasilkan distribusi probabilitas melalui beberapa kelas dan membantu dalam menginterpretasikan keluaran model sebagai probabilitas kelas tertentu. Selanjutnya, output probabilitas kelas. Model dikompilasi dengan pengoptimalan Stochastic Gradient Descent (SGD) dengan laju pembelajaran 0.01, menggunakan fungsi loss categorical_crossentropy untuk klasifikasi multi-kelas, dan metrik akurasi untuk evaluasi model. Konfigurasi ini merupakan langkah umum dalam membangun CNN untuk tugas klasifikasi dengan Keras dalam Python. Adapun ilustrasi klasifikasi menggunakan CNN akan terlihat pada gambar 10 di bawah ini.

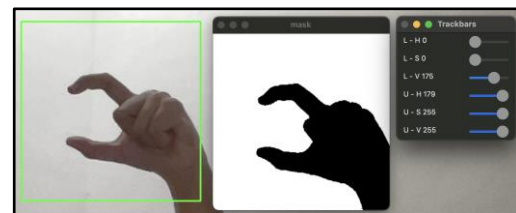


Gambar 10. Ilustrasi klasifikasi dengan CNN

4. HASIL DAN PEMBAHASAN

Proses pengenalan dimulai dengan citra dari kamera diolah dan disesuaikan ukurannya untuk dikirim sebagai input ke model jaringan saraf yang telah dilatih sebelumnya (CNN dari Keras). Selanjutnya, dilakukan perbandingan fitur SIFT antara citra yang ditangkap dan setiap citra gestur dalam direktori sample-gestures. Jika perbandingan fitur SIFT menghasilkan kecocokan yang signifikan, maka gestur yang dikenali akan sesuai dengan citra tersebut. Jika tidak ada kecocokan yang signifikan dengan citra gestur, program akan menggunakan hasil prediksi dari model CNN.

Selain itu terdapat penyesuaian warna dengan menggunakan trackbars yang digunakan untuk menentukan ambang batas warna yang diinginkan untuk mendeteksi gestur pada objek yang ada.



Gambar 11. Sampel uji coba pengenalan abjad B

Berdasarkan gambar 11, pengguna dapat menyesuaikan warna gestur yang ingin ditangkap menggunakan trackbars yang ada pada jendela "Trackbars". Hasil deteksi warna tersebut dijadikan mask untuk mengisolasi area yang menjadi fokus deteksi gestur. Selanjutnya, untuk mendapatkan nilai akurasi dapat dihitung menggunakan persamaan:

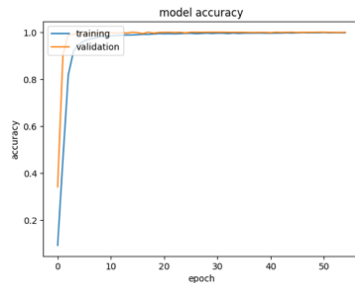
$$akurasi = \frac{\text{jumlah prediksi benar}}{\text{total jumlah data}} \quad (2)$$

Adapun jumlah prediksi benar adalah jumlah data yang benar-benar terklasifikasi dengan benar oleh model. Sedangkan total jumlah data adalah jumlah keseluruhan data yang dievaluasi oleh model. Akurasi tertinggi pada penelitian ini adalah pada model dengan inlay epoch 55 dengan nilai akurasi sebesar 99.78%.

Tabel 1. Nilai akurasi

Skenario	Epoch	Batch Size	Learning Rate	Akurasi
Pertama	15	32	0.01	99.05%
Kedua	25	32	0.01	99.60%
Ketiga	35	32	0.01	99.67%
Keempat	45	32	0.01	99.69%
Kelima	55	32	0.01	99.78%
Keenam	65	32	0.01	99.70%

Nilai akurasi uji coba pada epoch 15 sampai epoch 55 selalu mengalami peningkatan. Sedangkan, pada epoch 65 nilai akurasi mengalami penurunan sebesar 0.08%. Hal tersebut menandakan bahwa semakin besar epoch tidak membuat akurasi semakin meningkat.

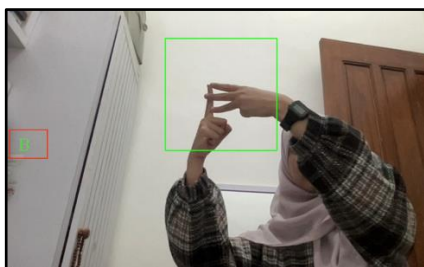


Gambar 12. Grafik akurasi pada epoch 55

Dari grafik pada gambar 12 terlihat bahwa nilai akurasi pada epoch 55 sebesar 99.70% atau 0.9970. Secara keseluruhan cara kerja program pengenalan ini adalah melakukan perbandingan antara citra uji dengan citra-citra referensi yang tersimpan. Ini dilakukan dengan menggunakan metode pencocokan fitur yang digunakan untuk mencocokkan deskriptor fitur antara citra uji dan citra referensi.

Proses ini menghasilkan kumpulan poin-poin yang sesuai antara dua citra. Selanjutnya, dilakukan pengecekan terhadap jumlah poin yang sesuai. Jika jumlah poin yang sesuai melebihi ambang tertentu, maka citra uji dianggap cocok dengan citra referensi yang sesuai. Jika tidak terdapat korelasi yang cukup dengan citra referensi yang ada, maka dilanjutkan dengan hasil prediksi dari model terhadap citra uji.

Hasil prediksi ini akan mengembalikan label yang sesuai dengan kelas yang diidentifikasi oleh model, seperti huruf A sampai Z. Secara keseluruhan, fungsi predictor ini bertujuan untuk melakukan identifikasi citra uji terhadap citra referensi yang telah disimpan, dengan langkah-langkah pencocokan fitur dan prediksi kelas menggunakan model yang telah dilatih sebelumnya.



Gambar 13. Sampel uji coba pengenalan abjad B

Seperti yang terlihat pada gambar 13, citra yang akan dikenali berada pada box yang berwarna hijau dan hasil output abjad akan muncul di ujung kiri. Berdasarkan hasil output abjad pada gambar 10 dapat diambil kesimpulan program yang dapat mengenali abjad B pada Bahasa isyarat jenis BISINDO.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian pengenalan Bahasa Isyarat Indonesia (BISINDO), beberapa kesimpulan dapat ditarik. Pertama, terdapat tantangan dalam pengenalan gerakan karena beberapa di antaranya memerlukan beberapa percobaan untuk mendeteksi dengan benar, disebabkan kemiripan dengan gerakan bahasa isyarat lain dan ketidakmerataan pencahayaan. Kedua, akurasi tertinggi tercapai pada model dengan nilai epoch 55, mencapai 99.78%. Namun, hasil pengujian menunjukkan bahwa peningkatan jumlah epoch tidak selalu berarti peningkatan akurasi, berbeda dengan penelitian sebelumnya. Dalam konteks pengembangan selanjutnya, disarankan untuk memperbanyak dataset dengan variasi lebih lanjut, menggunakan sensor kamera yang lebih baik, mengoptimalkan hyperparameter, mengembangkan program menjadi aplikasi mobile atau web, dan meningkatkan kemampuan program untuk beroperasi pada latar belakang yang beragam, tidak hanya berlatar putih.

DAFTAR PUSTAKA

- [1] M. Panwar, "Hand Gesture Recognition based on Shape Parameters," 2012.
- [2] A. Novianty and F. Azmi, "Sign Language Recognition using Principal Component Analysis and Support Vector Machine," *IJAIT (International Journal of Applied Information Technology)*, vol. 4, no. 01, p. 49, Mar. 2021, doi: 10.25124/ijait.v4i01.3015.
- [3] L. E. PRATAMA, "IMPLEMENTASI ALGORITMA SIFT (SCALE INVARIANT FEATURE TRANSFORM) UNTUK MELAKUKAN KLASIFIKASI BAHAN BAKAR KENDARAAN RODA EMPAT PADA SPBU," 2014. Accessed: May 17, 2023. [Online]. Available: <https://elib.unikom.ac.id/gdl.php?mod=browse&op=read&id=jbptunikompp-gdl-linggaekap-35759>
- [4] S. Saha, "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way." Accessed: Jan. 17, 2023. [Online]. Available: <https://medium.com/towards-data-science/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
- [5] I. Mahfudi, M. Sarosa, R. Andrie Asmara, and M. Azrino Gustalika, "Indonesian Sign Language Number Recognition using SIFT Algorithm," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Apr. 2018. doi: 10.1088/1757-899X/336/1/012010.
- [6] M. E. Walizad and M. Hurroo, "Sign Language Recognition System using Convolutional Neural Network and Computer Vision; Sign Language Recognition System using Convolutional Neural

- Network and Computer Vision,” 2020. [Online]. Available: www.ijert.org
- [7] M. I. Fachrurrozi, Y. Vita Via, W. S. J. Saputra, E. Y. Puspaningrum, F. Tri Anggraeny, and S. R. Nudin, “Object Localization and Detecting Alphabet in Sign Language BISINDO Using Convolution Neural Network,” 2023. [Online]. Available: www.techniumscience.com
- [8] M. I. FACHRURROZI, “ISLBISINDO1.” Accessed: Jan. 17, 2024. [Online]. Available: <https://www.kaggle.com/datasets/idhamozi/indonesian-sign-language-bisindo>
- [9] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. 2006.
- [10] R. Szeliski, *Computer Vision: Algorithms and Applications*. 2011.
- [11] K. Ayu Safitri, R. Wulanningrum, K. Kunci - Grafologi, and T. Tangan, “Aplikasi Pengenalan Pola Tulisan Tangan Menggunakan Metode Support Vector Machine,” 2020.
- [12] A. K. Hidayah, U. Juahardi, R. Toyib, and N. Ade Wijaya, “Designing an Android-Based Bisindo Dictionary Application Using the Boyer Moore Method,” *JURNAL KOMITEK*, vol. 2, no. 2, pp. 553–560, 2022, doi: 10.53697/jkomitek.v2i2.
- [13] H. Bestari, “Mengenal Bahasa Isyarat.” Accessed: Jan. 17, 2023. [Online]. Available: <https://www.ypedulikasihabk.org/2018/11/09/mengenal-bahasa-isyarat/>
- [14] D. G. Lowe, “Accepted for publication in the,” 2004.
- [15] R. Dunford, Q. Su, and E. Tamang, “The Pareto Principle,” 2014. [Online]. Available: <http://hdl.handle.net/10026.1/14054>