
ANALYSIS OF EUCLIDEAN DISTANCE ALGORITHM EFFECTIVENESS FOR COLOR DETECTION IN CIE LAB FORMAT

Kemal Ardiansyah^{*1}, Fathoni Mahardika², Deris Santika³

^{1,2,3}Informatics, Faculty of Information Technology, Sebelas April University Sumedang, Indonesia
Email: ¹230660121174@student.unsap.ac.id, ²fathoni@unsap.ac.id, ³deris@unsap.ac.id

(Article received: 23-07-2024; Revision: 25-07-2024; published: 07-12-2024)

Abstract

An image is a collection of two-dimensional colored points or pixels in the form of $f(x,y)$ that can be displayed on a monitor screen as a discrete value in digital form called a pixel. Color detection is a method used to classify one or more colors. There are many methods that can be used to identify colors in digital images and this research uses the Euclidean Distance method with reference points in CIE LAB format for each pixel. The image entered into Euclidean Distance is LAB color space. Euclidean Distance will calculate the difference in distance from a point to the reference point. The reference point is obtained by finding the mode value of all orange image data. Previously, all image data must first be converted into pixels. Determination of this reference point is strongly influenced by the total pixel size for each image, there is a pixel data distribution factor that affects the reference point. For the orange color with a dataset taken from the kaggle website with the project name is color dataset for color recognition as many as 25 images obtained an average accuracy of 96%.

Keywords: accuracy, color identification, euclidean distance, LAB color space, mode, orange color

ANALISIS EFEKTIVITAS ALGORITMA EUCLIDEAN DISTANCE UNTUK DETEKSI WARNA DALAM FORMAT CIE LAB

Abstrak

Citra merupakan kumpulan titik atau piksel berwarna dua dimensi berupa $f(x,y)$ yang bisa ditampilkan pada layar monitor sebagai nilai diskrit dalam bentuk digital yang disebut dengan piksel. Deteksi warna merupakan metode yang digunakan untuk mengklasifikasikan satu atau beberapa warna. Ada banyak metode yang bisa digunakan untuk mengidentifikasi warna pada citra digital dan penelitian ini mengidentifikasi warna menggunakan metode *Euclidean Distance* dengan titik acuan dalam format *CIE LAB* untuk setiap piksel. Citra yang dimasukkan ke dalam *Euclidean Distance* adalah *LAB color space*. *Euclidean Distance* akan menghitung selisih jarak dari suatu titik terhadap titik acuan. Titik acuan didapat dengan mencari nilai modus dari seluruh data citra warna *orange*. Sebelumnya seluruh data citra harus diubah dulu menjadi piksel-piksel. Penentuan titik acuan ini sangat dipengaruhi oleh ukuran total piksel untuk setiap citra, terdapat faktor sebaran data piksel yang mempengaruhi titik acuan. Untuk warna *orange* dengan *dataset* diambil dari *website kaggle* dengan nama proyek adalah *color dataset for color recognition* sebanyak 25 citra didapat akurasi rata - rata 96%.

Kata kunci: akurasi, euclidean distance, identifikasi warna, LAB color space, modus, warna orange

1. Pendahuluan

Citra merupakan kumpulan titik atau piksel berwarna dua dimensi [1] berupa $f(x,y)$ yang bisa ditampilkan pada layar monitor sebagai nilai diskrit dalam bentuk digital yang disebut dengan piksel [2]. Deteksi warna merupakan metode yang digunakan untuk mengklasifikasikan satu atau beberapa warna [3]. Pada penelitian sebelumnya identifikasi warna ada yang bertujuan untuk mengidentifikasi mutu buah berdasarkan warna seperti [4] ada pula penelitian lain yaitu sistem pengenalan warna

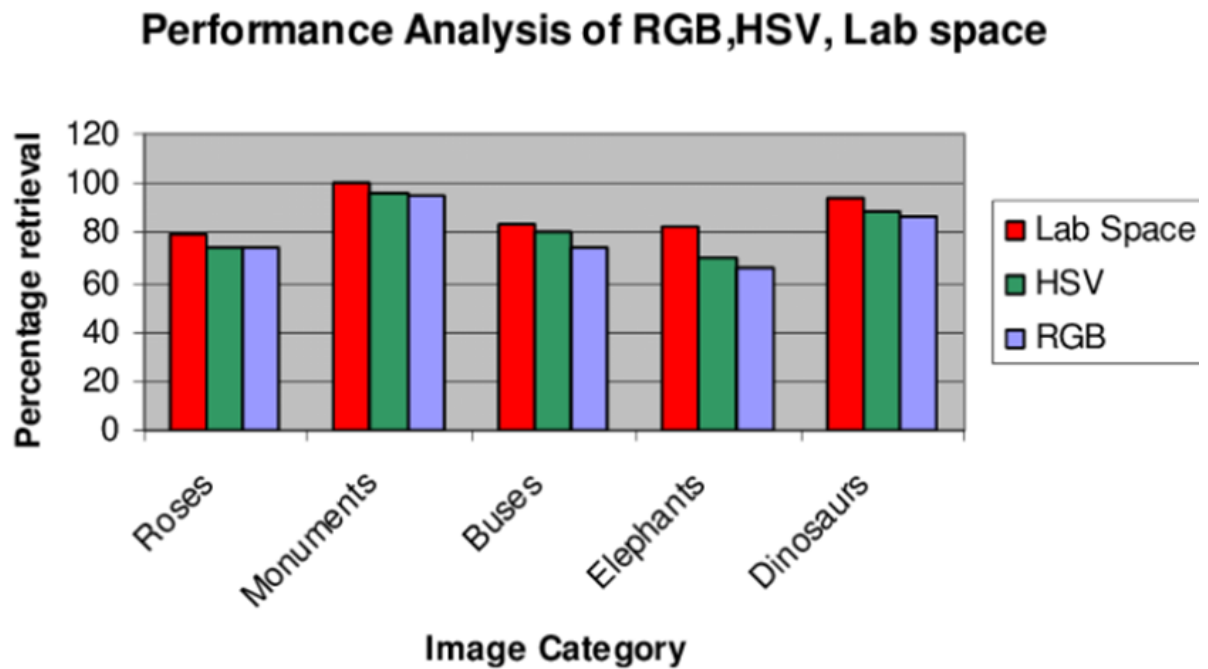
secara *real-time* untuk membantu orang-orang yang mengalami gangguan penglihatan. Sistem ini berbasis teknologi bantuan berbasis komputer yang dapat membantu orang-orang yang mengalami gangguan penglihatan dalam kebutuhan sehari-hari dan berintegrasi dengan baik dalam masyarakat. Penelitian ini menghasilkan sistem *embedded real-time* yang dapat membantu orang-orang yang mengalami gangguan penglihatan mengenali warna, berinteraksi, dan mengambil keputusan berdasarkan persepsi mereka terhadap warna [5], deteksi warna pada gambar RGB menggunakan teknik *Machine Learning* [6]. Semua penelitian tersebut menggunakan identifikasi warna sebagai dasar dari penelitiannya, maka dari itu kesalahan pada proses identifikasi warna akan membuat hasil penelitian menjadi berbeda. Terdapat berbagai metode untuk melakukan identifikasi warna, seperti pada penelitian [6] dimana pada penelitian tersebut mengusulkan *classifier K-Nearest Neighbour (KNN)* untuk deteksi warna RGB yang efisien.

Berdasarkan penelitian [7] dan [8] *Euclidean Distance measurement (EDM)* adalah salah satu algoritma yang simpel untuk diterapkan dalam menghitung jarak. *EDM* memiliki kelebihan dalam perhitungan jarak lurus antara satu titik dengan titik lain [8]. *EDM* juga baik untuk perhitungan jarak yang tidak memiliki banyak variabel sesuai dengan hasil penelitian [9]. Jika dibandingkan dengan perhitungan *Manhattan Distance* dalam algoritma *K-Means*, kinerja dari *Euclidean Distance* lebih baik [10]. Ini dibuktikan dengan gambar tabel di bawah ini, perhitungan dengan metode *Silhouette Coefficient* nilai *Euclidean* lebih tinggi dari *Manhattan* yang artinya *Euclidean* lebih baik. Sedangkan nilai standar deviasi *Euclidean* lebih rendah dari *Manhattan* artinya nilai standar deviasi *Euclidean* lebih mendekati rata – rata (semakin baik).

Tabel 1. Kinerja *K-Means* Menggunakan *Euclidean Distance* dan *Manhattan Distance*

Metode	<i>Silhouette Coefficient</i>	Standart Deviasi
<i>K-Means Euclidean Distance</i>	0,71	0,46466002
<i>K-Means Manhattan Distance</i>	0,64	0,4961977

Dalam jurnal "*Content-Based Image Retrieval using Color and Shape Descriptors*" dilakukan pengujian satu algoritma dengan format warna *HSV*, *RGB*, dan *LAB*. Pada penelitian tersebut algoritmanya menggunakan *Content-Based Image Retrieval (CBIR)* yang cara klasifikasi objeknya menggunakan warna dan bentuk dari objek. Dari gambar di bawah merupakan hasil dari penelitiannya, terlihat jelas bahwa kinerja *CBIR* dalam format warna *LAB* sangat baik dibandingkan dengan *HSV* dan *RGB* [11].

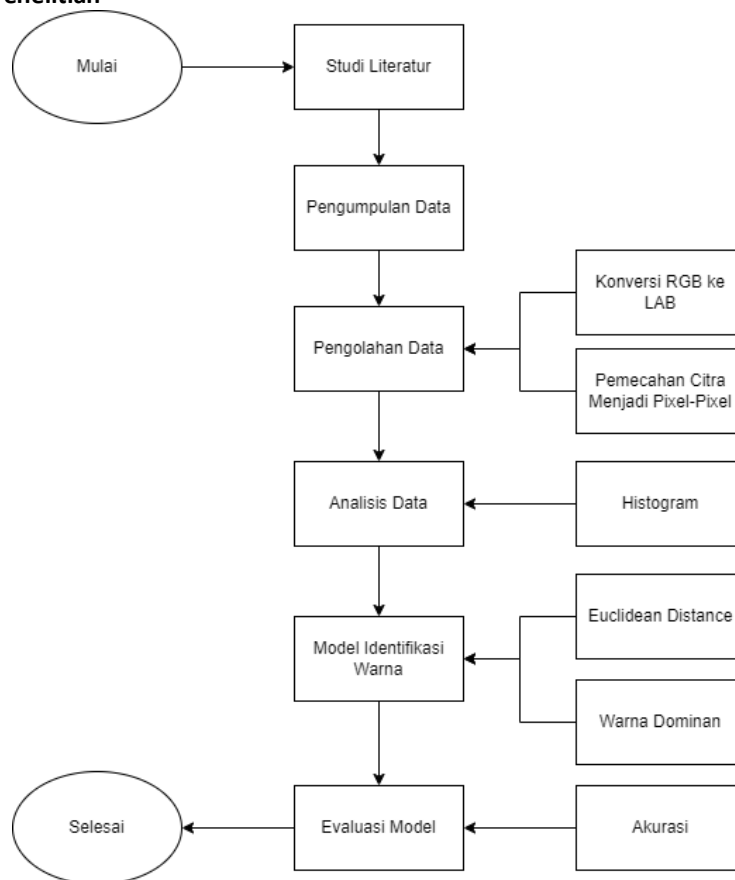


Gambar 1. Analisis Kinerja Untuk *Color Space* RGB, HSV, dan LAB

Dari penjelasan tersebut maka peneliti akan melakukan penelitian yang berjudul identifikasi warna menggunakan *LAB color space* dan *Euclidean Distance*. Format warna LAB akan digunakan sebagai format warna dari gambar, karena berdasarkan penelitian sebelumnya format warna *LAB* memiliki tingkat kinerja yang lebih baik dari warna *HSV*, dan *RGB*. Sedangkan untuk pengklasifikasian warna dari gambar akan digunakan metode *Euclidean Distance*. Dimana setiap piksel dari gambar akan diklasifikasikan berdasarkan hasil perhitungan *Euclidean Distance*. Dari penelitian sebelumnya metode *Euclidean* lebih baik dari *Manhattan* dalam penerapannya di algoritma *K-Means*.

2. Metode

2.1. Tahapan Penelitian



Gambar 2. Tahapan Penelitian

Berdasarkan gambar 2, penelitian ini dimulai dari studi literatur, pengumpulan data, pengolahan data dengan melakukan konversi RGB ke LAB dan citra dipecah menjadi pixel - pixel, analisis data menggunakan histogram, pembuatan model identifikasi menggunakan metode Euclidean Distance serta warna dominan dan evaluasi model melalui perhitungan akurasi.

2.2. Dataset

Data yang digunakan dalam penelitian ini berasal dari website kaggle dengan nama proyek adalah color dataset for color recognition. Datasetnya berisi kumpulan gambar – gambar dari warna *black, blue, brown, green, grey, orange, red, violet, white, yellow*. Setiap warna tersebut terdiri dari 25 gambar dalam RGB color space. Dengan demikian terdapat 250 gambar. Sebagai pengambilan sampel maka warna yang digunakan hanya warna *orange* sebanyak 25 citra.

Citra akan dikonversi dari RGB menjadi LAB kemudian citra dipecah menjadi pixel - pixel. Pemecahan citra menjadi pixel dilakukan dengan melakukan iterasi. Hal ini dapat dilakukan karena citra (dalam penelitian ini menggunakan citra digital) merupakan suatu matriks height x width. Posisi dari pixel ditentukan dari koordinat height dan width, $f(i,j)$. Setiap pixel terdiri dari 3 channel yang merepresentasikan komponen Luminance (L), Green-Red (a), serta Blue-Yellow (b).

2.3. Analisis Data

Citra yang sudah dalam pixel – pixel selanjutnya dianalisis menggunakan histogram. Dengan histogram maka akan diketahui distribusi frekuensi dari setiap komponen Luminance (L), Green-Red (a), serta Blue-Yellow (b) suatu citra digital. Hal ini penting karena nilai dengan distribusi frekuensi tertinggi akan untuk setiap komponen akan digunakan sebagai titik acuan dalam menentukan warna dari pixel. Titik acuan ini akan dimasukkan kedalam model classifier untuk dibandingkan dengan setiap pixel dari citra.

2.4. Euclidean Distance

Euclidean Distance merupakan salah satu metode yang digunakan untuk menghitung selisih jarak terdekat dari suatu point atau titik. Titik ini salah satunya adalah pixel dalam citra, setiap pixel dalam citra terdiri dari komponen penyusunnya seperti jika citranya RGB maka satu pixel akan terdiri dari nilai warna merah, hijau, dan biru (membentuk 3 channel), sedangkan untuk penelitian ini menggunakan CIELab maka satu pixelnya terdiri dari nilai luminance, green-red, serta blue-yellow. Secara sederhana Euclidean Distance akan menghitung nilai kedekatan antara dua titik dengan dimana dalam penelitian ini dua titik merupakan nilai dari komponen luminance pada pixel ke (i,j) dikurangi dengan nilai luminance dari titik acuan yang didapatkan lewat histogram. Begitupun untuk green-red, serta blue-yellow. Terakhir semua selisih jaraknya ditambahkan. Dengan demikian persamaan untuk menghitung selisih jarak terdekat adalah sebagai berikut

$$\begin{aligned} & \text{pixel}(i,j) \\ & = \sqrt{(\text{luminance}_{(i,j)} - \text{luminance}_{\text{titik acuan}})^2 + (\text{green red}_{(i,j)} - \text{green red}_{\text{titik acuan}})^2 + (\text{blue yellow}_{(i,j)} - \text{blue yellow}_{\text{titik acuan}})^2} \end{aligned}$$

Semakin kecil nilai dari Euclidean Distance artinya semakin mirip hubungan antara dua buah pixel tersebut. Dalam penelitian ini Euclidean Distance digunakan sebagai classifier atau predictor atau yang akan menentukan apakah suatu pixel merupakan warna orange atau bukan.

2.5. Akurasi

Akurasi berfungsi untuk mengevaluasi kinerja dari algoritma yang dibuat atau dikembangkan. Semakin tinggi nilai akurasi artinya semakin baik kinerja dari algoritma atau model yang dibuat atau dikembangkan. Untuk memperoleh akurasi pertama – tama perlu menghitung terlebih dahulu setiap hasil yang diprediksi oleh algoritma atau model dengan hasil yang sebenarnya, apakah terprediksi dengan benar atau salah memprediksi. Untuk mendapatkan akurasi maka setiap percobaan pengujian yang diprediksi dengan benar oleh algoritma akan dibagi oleh total dari percobaan pengujian lalu dikalikan dengan angka 100 sehingga akurasi yang didapat dalam persen. Dengan demikian persamaan untuk menghitung akurasi adalah sebagai berikut:

$$\text{akurasi (\%)} = \frac{\text{percobaan pengujian benar}}{\text{total percobaan pengujian}}$$

3. Hasil dan Pembahasan

3.1 Histogram

Data yang divisualisasikan melalui histogram merupakan pixel – pixel dari setiap citra yang dikumpulkan ke dalam setiap komponen. Setiap komponen ini disimpan ke dalam struktur data larik. Larik ini merupakan hasil dari tahap pemecahan citra menjadi pixel – pixel dengan tujuan analisis. Selanjutnya setiap komponen dimasukkan kedalam fungsi hist() dari library matplotlib. Dengan fungsi hist() akan diperlihatkan grafik dari histogram untuk setiap komponen color space LAB. Parameter dari fungsi hist() dalam penelitian ini hanya diisi x yang merupakan data, diisi dengan larik yang merupakan kumpulan pixel – pixel untuk setiap komponen color space LAB, parameter bins yang akan mengelompokkan data, diisi dengan 'auto' karena setiap komponen dari color space LAB memiliki range yang berbeda – beda tidak seperti RGB yang setiap komponennya memiliki range yang sama yaitu 0 sampai 256, dan parameter color, untuk komponen luminance diisi dengan 'red' sedangkan untuk komponen green-red dan blue-yellow masing – masing diisi 'green' dan 'blue'. Parameter color akan memberikan warna ketika data divisualisasikan dalam histogram. Dalam grafik histogram sumbu x diberi keterangan nilai intensitas (intensity value) yang diatur dalam xlabel dan sumbu y diberi keterangan frekuensi kemunculan dari data (count) yang diatur dalam ylabel.

```
import matplotlib.pyplot as plt

_ = plt.hist(luminance, bins = 'auto', color = 'red')

_ = plt.hist(green_red_a, bins = 'auto', color = 'green')

_ = plt.hist(blue_yellow_b, bins = 'auto', color = 'blue')

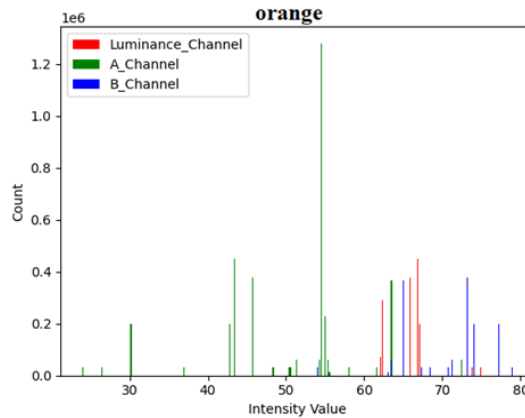
_ = plt.xlabel('Intensity Value')

_ = plt.ylabel('Count')

_ = plt.legend(['Luminance_Channel', 'A_Channel', 'B_Channel'])

plt.show()
```

Grafik histogram untuk warna *orange* dapat dilihat pada gambar 3. Pada gambar 3 terlihat bahwa nilai dominan untuk setiap komponen LAB dimana komponen luminance ada di *range* 60-an, sedangkan untuk A ada di 50-an dan B ada di 70-an.



Gambar 3. Histogram Seluruh Citra Warna *Orange*

Jika hanya melihat histogram saja, seperti terlihat pada gambar 3, untuk mencari titik acuan masih kurang jelas. Dengan demikian untuk mencari titik acuan, dalam penelitian ini titik acuan diartikan sebagai nilai dari setiap komponen yang paling sering muncul, maka menggunakan nilai modus. Nilai modus, yang juga merupakan nilai frekuensi tertinggi dari histogram, dapat dicari menggunakan library scipy fungsi mode() di bagian stats. Pada fungsi mode() diisi dengan setiap komponen dari LAB dan untuk mendapatkan nilainya tambahkan atribut mode. Sehingga didapat hasil untuk modus menggunakan library scipy dengan fungsi mode(), yang menjadi titik acuan, dari setiap warna *orange* pada penelitian ini dengan dataset color dataset for color recognition dari situs kaggle untuk komponen luminance adalah 62.479811150909384, komponen A (green-red) 54.55919834106504 dan untuk komponen B (blue-yellow) yaitu 71.41677061953379.

Hasil pengujian untuk warna *orange* sebanyak 25 citra terlihat seperti pada gambar 5. Rata - rata akurasi untuk ke-25 citra tersebut adalah 96% yang didapat dengan menjumlahkan semua akurasi dibagi dengan banyaknya data (25).

Tabel 2. Hasil Prediksi

no	akurasi(%)	jumlah pixel prediksi benar	jumlah salah prediksi	total pixel
			yellow	
1	100	30000	0	30.000
2	100	30000	0	30.000
3	100	30000	0	30.000
4	100	30000	0	30.000
5	100	30000	0	30.000
6	100	30000	0	30.000
7	100	30000	0	30.000
8	100	30000	0	30.000
9	100	30000	0	30.000
10	100	30000	0	30.000
11	100	30000	0	30.000
12	100	30000	0	30.000
13	100	30000	0	30.000
14	0	0	30000	30.000
15	100	640000	0	640.000
16	100	640000	0	640.000
17	100	375765	0	375.765
18	100	199808	0	199.808
19	100	74520	0	74.520
20	100	198912	0	198.912
21	100	365118	0	365.118
22	100	420660	0	420.660
23	100	197578	0	197.578
24	100	30000	0	30.000
25	100	30000	0	30.000

4. DISKUSI

Pada penelitian ini yang bertujuan mengidentifikasi citra warna *orange* mencapai hasil yang memuaskan yaitu akurasi mencapai 96%. Dengan demikian untuk penerapan LAB *color space* dan metode Euclidean Distance sebagai *classifier* pada citra yang telah dipecah menjadi pixel - pixel dapat dikatakan berhasil. Bagian yang paling penting dari penelitian ini yaitu penentuan titik acuan atau *color palette*. Untuk kasus warna *orange* titik acuannya yang menggunakan nilai modus dapat dikatakan tepat, komponen luminance adalah 62.479811150909384, komponen A (green-red) 54.55919834106504 dan untuk komponen B (blue-yellow) yaitu 71.41677061953379. Tetapi jika melihat gambar 4.2 terlihat terdapat sembilan citra yang memiliki jumlah pixel lebih besar dari pada yang lainnya. Jumlah pixel setiap citra sangat berpengaruh terhadap penentuan titik acuan, jika titik acuan tidak tepat bisa menyebabkan akurasi model buruk.

5. KESIMPULAN

Pada penelitian ini, prediksi warna dengan menggunakan euclidean Distance serta LAB *color space* sangat dipengaruhi oleh titik acuan, jika titik acuannya tepat maka akan menghasilkan prediksi yang tepat. Untuk kasus warna *orange* penentuan titik acuan menggunakan statistik yaitu modus memperoleh hasil yang memuaskan. Perlu penelitian lanjutan untuk menentukan titik acuan menggunakan metode yang lainnya.

Kemudian sebaran data, jumlah pixel untuk setiap citra, sangat mempengaruhi titik acuan sehingga perlu diperhatikan agar menjaga sebaran pixel datanya tidak terlalu jauh. Untuk itu diperlukan proses *preprocessing* yang bertujuan menyeragamkan ukuran citra, karena jumlah pixel ditentukan oleh ukuran citra dimana total pixel dari suatu citra merupakan perkalian dari height dan width dari citra.

Penelitian identifikasi warna dengan LAB *color space* dan metode Euclidean Distance ini hanya menggunakan satu jenis warna yaitu warna *orange* saja sehingga perlu dilakukan penelitian lanjutan untuk mengukur warna yang lainnya.

DAFTAR PUSTAKA

- [1] Jumadi, Juju, Yupianti Yupianti, and Devi Sartika. 2021. "Pengolahan Citra Digital Untuk Identifikasi Objek Menggunakan Metode Hierarchical Agglomerative Clustering." JST (Jurnal Sains Dan Teknologi). <https://doi.org/10.23887/jst-undiksha.v10i2.33636>
- [2] Goenawan, A. D., Rachman, M. B. A., & Pulungan, M. P. (2022, January 29). Identifikasi Warna Pada Objek Citra Digital Secara Real Time Menggunakan Pengolahan Model Warna HSV. Jurnal Teknik Informatika Dan Elektro, 4(1), 68-74. <https://doi.org/https://doi.org/10.55542/jurtie.v4i1.430>
- [3] Dessy Ana Laila Sari, Adi Mulyadi, Adi Pratama, and Rezki Nalandari, "DETEKSI OBJEK BERWARNA REAL TIME BERDASARKAN VISUALISASI WEBCAM ", ZTR, vol. 2, no. 1, pp. 21-24, Mar. 2020.
- [4] Dhiman, B., Kumar, Y. & Kumar, M. Fruit quality evaluation using machine learning techniques: review, motivation and future perspectives. Multimed Tools Appl 81, 16255–16277 (2022). <https://doi.org/10.1007/s11042-022-12652-2>.

- [5] Samara, M., Deriche, M., Al-Sadah, J. et al. Design and Implementation of a Real-Time Color Recognition System for the Visually Impaired. *Arab J Sci Eng* 48, 6783–6796 (2023). <https://doi.org/10.1007/s13369-022-07506-w>.
- [6] Awotunde, J.B., Misra, S., Obagwu, D., Florez, H. (2022). Multiple Colour Detection of RGB Images Using Machine Learning Algorithm. In: Florez, H., Gomez, H. (eds) *Applied Informatics. ICAI 2022. Communications in Computer and Information Science*, vol 1643. Springer, Cham. https://doi.org/10.1007/978-3-031-19647-8_5.
- [7] Dokmanic, Ivan, Reza Parhizkar, Juri Ranieri, and Martin Vetterli. 2015. "Euclidean Distance Matrices: Essential Theory, Algorithms, and Applications." *IEEE Signal Processing Magazine* 32(6):12–30.
- [8] D'Agostino, Marcello and Valentino Dardanoni. 2009. "What's so Special about Euclidean Distance? A Characterization with Applications to Mobility and Spatial Voting." *Social Choice and Welfare* 33(2):211–33.
- [9] Selvarasu, N., Alamelu Nachiappan, and N. M. Nandhitha. 2010. "Euclidean Distance Based Color Image Segmentation of Abnormality Detection from Pseudo Color Thermographs." *International Journal of Computer Theory and Engineering* 2(4):514–16.
- [10] Wahyu Pribadi, W., Yunus, A., & Wiguna, A.S. (2022). PERBANDINGAN METODE K-MEANS EUCLIDEAN DISTANCE DAN MANHATTAN DISTANCE PADA PENENTUAN ZONASI COVID-19 DI KABUPATEN MALANG. *JATI (Jurnal Mahasiswa Teknik Informatika)*.
- [11] J. Pujari, S. N. Pushpalatha and D. Padmashree, "Content-Based Image Retrieval using color and shape descriptors," 2010 International Conference on Signal and Image Processing, Chennai, India, 2010, pp. 239-242, doi: 10.1109/ICSIP.2010.5697476