

Preprocessing serta Normalisasi pada Dataset Kupu-Kupu untuk Ekstraksi Fitur Warna, Bentuk dan Tekstur

Dhian Satria Yudha Kartika ¹, Hendra Maulana ^{2,*}

¹ Universitas Pembangunan Nasional Veteran Jawa Timur; dhian.satria@upnjatim.ac.id

² Universitas Pembangunan Nasional Veteran Jawa Timur; hendra.maulana.if@upnjatim.ac.id

* Correspondence: dhian.satria@upnjatim.ac.id.

Abstrak: Perkembangan penelitian dalam bidang *image processing* atau *citra digital* sangat bermanfaat pada beberapa bidang, diantaranya pertanian, kesehatan ataupun infrastruktur dan tata kelola pemerintahan. Penelitian dalam *citra digital* bisa membantu mendapatkan keputusan terbaik agar hasil yang didapatkan sesuai dengan rencana dan analisa permasalahan. Penelitian dalam bidang *citra digital* yang berkembang salah satunya penelitian dalam hal serangga. Dalam penelitian ini jenis serangga yang digunakan adalah kupu-kupu sebanyak 890 data. Dataset tersebut dibagi menjadi 10 kelas, masing-masing kelas sebanyak 89 data. Dataset akan dilakukan proses ekstraksi fitur warna, fitur tekstur dan fitur bentuk. Sebelum proses ekstraksi fitur hal terpenting adalah melakukan *preprocessing* dan normalisasi ukuran piksel. *Preprocessing* dilakukan untuk menghilangkan *noise* pada gambar. *Noise* dalam penelitian citra digital merupakan hal yang harus dihilangkan agar tidak mengurangi hasil yang akan didapatkan. Selain menghilangkan *noise*, proses normalisasi ukuran setiap objek dataset juga disesuaikan. Ukuran masing-masing dataset disamakan sehingga hasil yang didapatkan juga mempunyai standarisasi nilai. Hasil yang didapatkan pada proses klasifikasi kupu-kupu mempunyai nilai akurasi sebesar 75% penggabungan dari ketiga ekstraksi fitur warna, fitur tekstur dan fitur bentuk pada ukuran piksel 256x160.

Kata Kunci: Citra digital; Preprocessing; Normalisasi Piksel, Ekstraksi Fitur, Kupu-Kupu

1. Pendahuluan

Pengambilan informasi berupa gambar banyak digunakan di berbagai bidang penelitian misalnya di bidang pertanian untuk bidang penyakit pada buah-buahan. Pada buah-buahan, data dapat diambil dan dicocokkan ke dalam dataset untuk menentukan buah mana yang terinfeksi penyakit atau tidak berdasarkan warnanya [1]. Penelitian dalam bidang kesehatan misalnya mendeteksi darah sebelum proses transfusi dari pendonor ke pasien. Aplikasi deteksi darah mampu mendeteksi gambar dengan mudah, cepat, aman, dan memiliki akurasi yang tinggi, sehingga terdeteksi darah yang akan diambil tersebut aman atau berpenyakit [2]. Beberapa peneliti juga menggunakan pengolahan citra untuk identifikasi spesies serangga [3], pengolahan citra juga digunakan untuk mendeteksi secara otomatis jenis dan klasifikasi batik [4] [5], dalam penelitian lain menggunakan pengolahan citra untuk mengidentifikasi jaringan jalan perkotaan menggunakan data dengan resolusi tinggi [6]. Penelitian dalam hal normalisasi dan standarisasi dataset ini merupakan rangkaian dari penelitian sebelumnya terkait klasifikasi warna pada ikan koi dan kupu-kupu [7] [8] [9].

Salah satu studi yang berkembang di bidang pengolahan citra saat ini adalah identifikasi spesies serangga. Lebih dari 170.000 spesies serangga ditemukan oleh para peneliti, berdasarkan warna,

bentuk, dan tekstur pada sayapnya. Kupu-kupu merupakan salah satu jenis serangga yang memiliki keunikan baik dari segi warna, corak, dan tekstur pada bagian fisiknya, keunikan ini semakin menjadi ciri khusus kupu-kupu untuk dibedakan berdasarkan spesiesnya. Beberapa metode umum yang digunakan untuk mendeteksi karakteristik tekstur menggunakan local binary pattern (LBP) kemudian hasilnya diklasifikasikan menggunakan jaringan saraf tiruan (JST) [3]. Meskipun hasil penelitian sebelumnya [10] [11] menunjukkan LBP sedikit lebih rendah dibandingkan metode lain, namun LBP dianggap mampu menggambarkan citra dengan baik [12] dan banyak digunakan dalam beberapa penelitian misalnya pengolahan citra, computer vision, dan pengambilan citra, penginderaan jauh (satelit), dan analisis citra di bidang medis. Keuntungan menggunakan operator LBP adalah toleransi perubahan iluminasi, komputasi ringan yang memungkinkan analisis gambar secara real-time [13].

Komponen dasar yang biasa digunakan dalam penelitian di bidang pencitraan antara lain fitur warna, bentuk, dan tekstur [7] [9]. Metode yang digunakan untuk proses ekstraksi ciri warna menggunakan histogram warna. Histogram warna merupakan representasi dari sebaran warna pada citra. Distribusi warna pada gambar menunjukkan probabilitas warna sebelum ditentukan. Histogram warna dapat dibangun untuk semua jenis dimensi warna, beberapa diantaranya sering digunakan untuk ruang warna tiga dimensi misalnya RGB, HSV, L^*a^*b , YCbCr, dan lain-lain [14]. Dalam penelitian ini [5] menyebutkan dalam hal ekstraksi ciri dan warna pada batik yang menggunakan Fast Discrete Curvelet Transform (FDCT) dan ruang warna HSV untuk ekstraksi warna disebutkan proses ekstraksi ciri warna RGB menjadi HSV ditambah dengan Color Quantization (CQ). Tujuan dari Color Quantization digunakan untuk mengurangi waktu komputasi tanpa mengurangi kualitas gambar. Kombinasi yang bagus untuk mengukur performa dan waktu komputasi dengan menggunakan 72 fitur warna yang terdiri dari 8 komponen Hue, 3 komponen Saturation dan 3 komponen Value [5]. Dalam penelitian terkait, Yousef melakukan pengujian ekstraksi ciri warna hanya dengan menggunakan nilai Hue saja dan membandingkannya dengan menggunakan Hue, Saturation, Value secara bersamaan. Hasil nilai Saturasi dan Nilai memberikan peningkatan pada nilai dimensi dan menambah lebih banyak informasi tentang gambar [15].

Dalam penelitian ini akan dilakukan proses menghilangkan *background* pada objek dataset kupu-kupu. Noise dalam penelitian citra merupakan hal yang harus dihilangkan agar tidak mengurangi hasil yang akan didapatkan. Selain menghilangkan noise, proses normalisasi ukuran setiap objek dataset juga disesuaikan. Ukuran masing-masing dataset disamakan sehingga hasil yang didapatkan juga mempunyai standarisasi nilai.

2. Metode dan Dataset

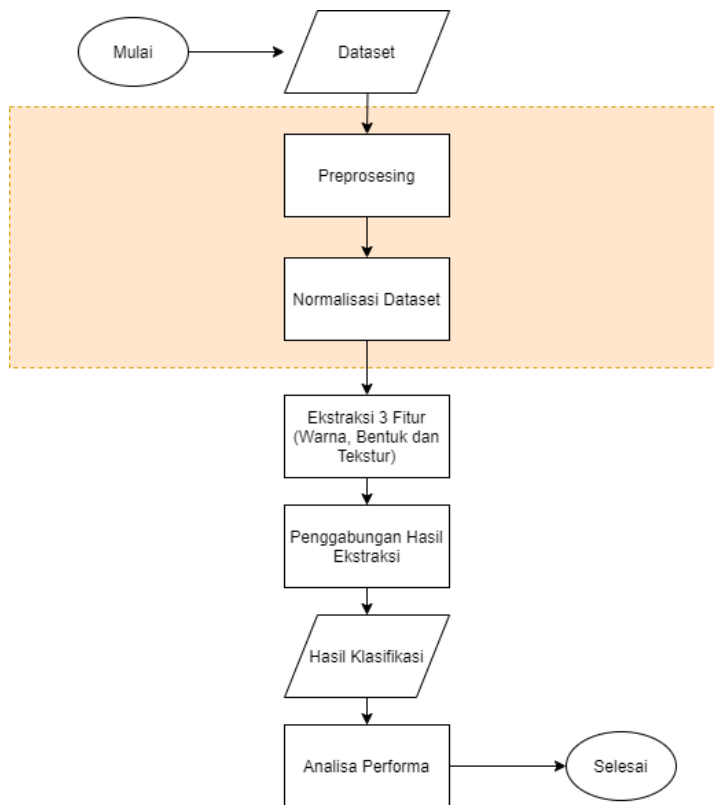
2.1. Metode

Pada penelitian ini menggunakan beberapa metode seperti Color Quantization pada HSV Color Space yang digunakan untuk ekstraksi ciri warna, Local Binary Pattern digunakan untuk ekstraksi ciri tekstur dan Region props digunakan untuk ekstraksi ciri bentuk. Proses menggunakan metode tersebut akan dilakukan preprocessing dan normalisasi data. Proses tersebut digunakan untuk menyiapkan data agar siap untuk setiap proses ekstraksi fitur. Metodologi penelitian dijelaskan pada gambar 1.











Penelitian ini fokus pada preprosesing dan normalisasi data. Membersihkan dataset dari noise pada objek kupu-kupu. Noise yang dimaksud seperti daun, bunga, kayu dan lain lain. Sehingga yang didapatkan nantinya fokus pada objek kupu-kupu. Setelah dilakukan preprosesing, gambar akan dilakukan cropping (pemotongan) sesuai dengan ukuran objek. Setelah proses cropping dilakukan, ukuran masing-masing dataset dilakukan standarisasi pixel.

2.2. Dataset

Pada tahap ini proses pengumpulan dan analisis data untuk dijadikan dataset. Data yang digunakan dalam penelitian ini adalah gambar kupu-kupu yang diambil dari penelitian [17] yang berjudul Model Pembelajaran Pengenalan Objek dari Deskripsi Bahasa Alami. Gambar kupu-kupu



Gambar 1. Flowchart yang Dilakukan dalam Penelitian

				
<i>Danaus plexippus</i>	<i>Heliconius charitonius</i>	<i>Nymphalis antiopa</i>	<i>Papilio cresphontes</i>	<i>Heliconius Erato</i>
				
<i>Pieris rapae</i>	<i>Vanessa atalanta</i>	<i>Junonia coenia</i>	<i>Lycaena phlaeas</i>	<i>Vanessa cardui</i>

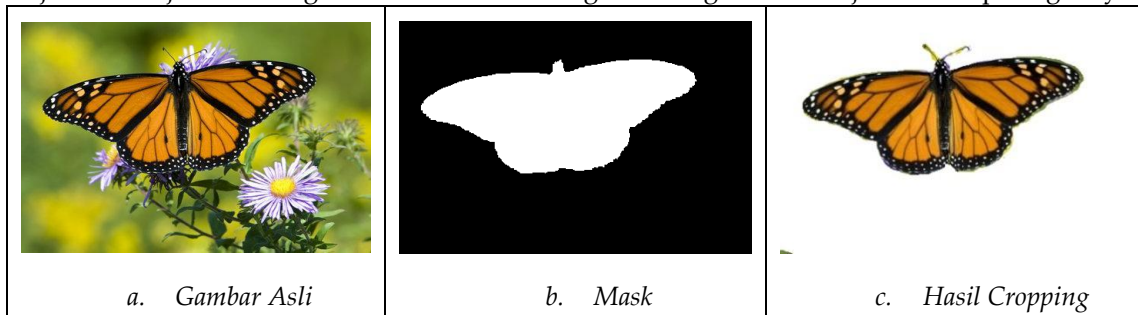
Gambar 2. Dataset Foto Kupu Kupu

yang digunakan sebanyak 890 gambar dalam format JPEG dan PNG, masing-masing jenis kupu-kupu berjumlah 89 gambar. Proses pengambilan gambar dataset bervariasi, dari sisi atas, depan, belakang, kanan atau kiri. Dataset yang digunakan dalam penelitian dijelaskan pada gambar 2.

3. Preprocessing dan Normalisasi

3.1. Preprocessing

Segmentasi menjadi pembahasan yang sangat penting dalam bidang analisa gambar dan computer vision. Banyak metode yang digunakan untuk proses segmentasi. Beberapa penelitian menjelaskan tujuan dari segmentasi untuk membagi sebuah gambar menjadi beberapa bagian yang



Gambar 3. Preprocessing pada Gambar Kupu Kupu

sesuai dengan kebutuhan. Ada 2 syarat dasar melakukan metode segmentasi citra diantaranya : proses segmentasi menghasilkan nilai paling akurat dan kecepatan pemrosesan citra yang paling cepat, kedua hal tersebut akan mendukung proses selanjutnya. Dalam praktiknya kedua syarat itu bertentangan, karena akurasi yang tinggi mungkin membutuhkan komputasi yang lama dan komputasi yang cepat menghasilkan akurasi yang rendah. Sehingga perlu ditambahkan algoritma lagi pada proses segmentasi untuk menyeimbangkannya.

Pada penelitian citra kupu-kupu ini akan melakukan preprocessing dan tidak melakukan segmentasi. Perbedaan antar keduanya yaitu proses segmentasi menggunakan beberapa metode untuk mengaplikasikannya, sedangkan preprocessing adalah teknik untuk memastikan data siap digunakan tanpa menggunakan metode. Peneliti Josiah Wang [17] memanfaatkan kupu-kupu sebagai dataset dan melakukan proses segmentasi. Dataset sebelum dilakukan proses segmentasi masih mempunyai banyak noise. Noise yang dimaksud adalah gambar masih ada background berupa bunga, ranting, atau benda lainnya. Dalam penelitiannya [17] proses segmentasi memberikan hasil berupa masking atau hasil potongan noise berupa background warna hitam. Sehingga perlu dilakukan beberapa tahap untuk menggabungkan antara gambar asli dengan mask untuk mendapatkan gambar kupu-kupu utuh (badan dan sayap) tanpa noise. Karena preprocessing ini sangat menentukan kesuksesan metode berikutnya, oleh karena itu harus dipastikan menghasilkan gambar dengan kualitas dan akurasi yang baik.

Tahap preprocessing bisa dilihat pada gambar 3 berikut ini, yaitu menggabungkan antara gambar (a) gambar asli kupu-kupu (yang masih ada noise, seperti bunga dan background yang lain) dengan gambar (b) mask pada penelitian sebelumnya sehingga didapatkan gambar (c) hasil cropping.

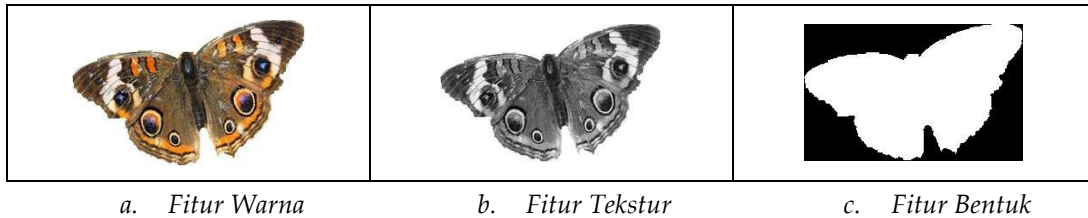
3.2. Normalisasi

Proses normalisasi adalah sebuah proses yang digunakan untuk mempersiapkan data agar bisa digunakan untuk memenuhi kebutuhan pemakai dan mendukung proses berikutnya mendapatkan hasil yang lebih baik. Pentingnya proses normalisasi agar inputan mempunyai nilai yang seragam, nilai yang dimaksud adalah ukuran gambar dengan piksel yang sama.

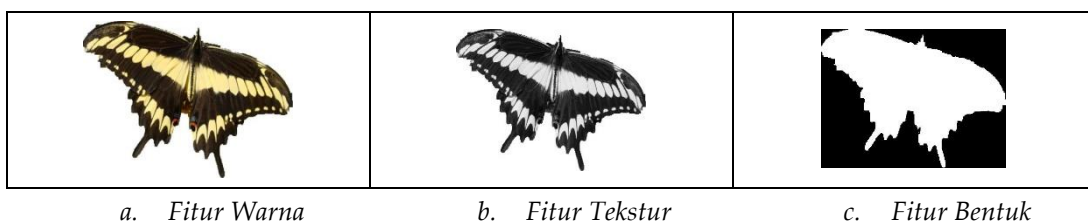
Inputan data yang tidak sama akan mempengaruhi proses yang dihasilkan dan juga mempengaruhi hasil akhir pada aplikasi. Beberapa skenario proses normalisasi dataset diantaranya, menghilangkan *background* gambar (warna putih), mengubah gambar menjadi *blackwhite*, rekonstruksi citra dengan menambahkan salah satu fungsi, melakukan *cropping* dan *resize* pada gambar sesuai dengan ukuran piksel yang diinginkan. Ada beberapa skenario yang dilakukan pada

saat resize gambar yaitu ukuran gambar menjadi 256 x 160 piksel, mengubah ukuran menjadi 420 x 315 piksel dan mengubah gambar dengan ukuran 768 x 576 piksel.

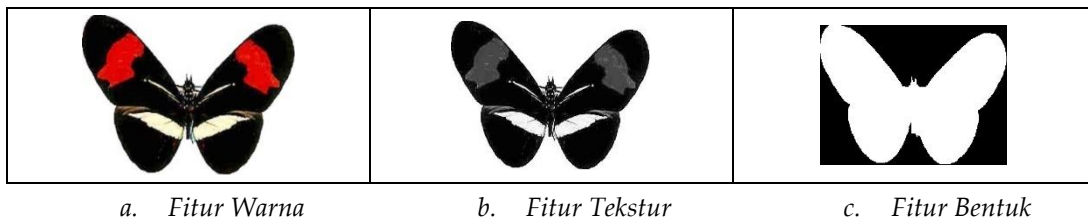
Proses normalisasi diatas akan mengubah ukuran gambar yang awalnya beragam ukuran menjadi seragam, dengan 3 ukuran berdasarkan skenario diatas. Hasil normalisasi beberapa dataset bisa dilihat pada gambar 4, gambar 5 dan gambar 6.



Gambar Error! No text of specified style in document. Hasil Normalisasi ukuran 256x160 piksel



Gambar 5. Hasil Normalisasi ukuran 420x315 piksel



Gambar 6. Hasil Normalisasi ukuran 768x576 piksel

4. Hasil Penelitian dan Analisis

Pada tahapan preprosesing dan normalisasi digunakan untuk proses klasifikasi. Setelah ukuran dari masing-masing gambar dataset sudah seragam, maka dilakukan proses ekstraksi fitur warna, bentuk dan tekstur. Tahapan setelah ekstraksi selanjutnya dilakukan proses klasifikasi. Hasil klasifikasi yang sudah dilakukan, selanjutnya dilakukan proses pengujian. Dari proses pengujian didapatkan nilai akurasi masing-masing ukuran piksel, piksel 256x160, piksel 420x315 dan piksel 768x576. Nilai akurasi bisa dilihat pada tabel 1.

Tabel 1. Nilai Akurasi

Ukuran piksel	Fitur Warna (FW)	Fitur Tekstur (FT)	Fitur Bentuk (FB)	FW & FT	FW & FB	FT & FB	FW & FT & FB
256x160	72 %	68 %	49 %	70 %	71%	66 %	75 %
420x315	75 %	60 %	39 %	72 %	74 %	64 %	70 %
768x576	78 %	60 %	39 %	71 %	74 %	64 %	71 %

Berdasarkan tabel 1 nilai akurasi. Hasil ekstraksi fitur warna terbaik pada ukuran piksel 768x576. Sedangkan hasil ekstraksi fitur tekstur pada ukuran piksel 256x160. Dan untuk hasil ekstraksi fitur bentuk rata-rata dibawah 50%.

5. Kesimpulan

Nilai akurasi pada proses penggabungan dari ketiga ekstraksi fitur warna, bentuk dan tekstur didapatkan nilai sebesar 75% dengan piksel 256x160. Nilai akurasi sebesar 75% dikatakan baik dibandingkan dengan hasil pengujian yang lain pada ukuran piksel yang berbeda. Nilai akurasi yang baik ini, tidak lepas dari tahapan preprosesing dan normalisasi yang dilakukan sebelum ekstraksi fitur.

Referensi

1. Bhange, M., & Hingoliwala, H. A. (2015). Smart Farming: Pomegranate Disease Detection Using Image Processing. *Procedia Computer Science*, 58, 280–288. <https://doi.org/10.1016/j.procs.2015.08.022>
2. Ferraz, A., Carvalho, V., & Machado, J. (2016). Determination of Human Blood Type Using Image Processing Techniques. *Measurement*, 97. <https://doi.org/http://dx.doi.org/10.1016/j.measurement.2016.11.015>
3. Kaya, Y., Kayci, L., & Uyar, M. (2015). Automatic identification of butterfly species based on local binary patterns and artificial neural networks. *Applied Soft Computing Journal*, 28, 132–137. <https://doi.org/10.1016/j.asoc.2014.11.046>
4. Kurniawardhani, A., Suciati, N., & Arieshanti, I. (2014). Klasifikasi Citra Batik Menggunakan Metode Ekstraksi Ciri yang Invariant Terhadap Rotasi. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 12(2), 48. <https://doi.org/10.12962/j24068535.v12i2.a322>
5. Suciati, N., Kridanto, A., Naufal, M. F., Machmud, M., & Wicaksono, Y. (2015). Fast Discrete Curvelet Transform And HSV Color Features For Batik Image Classification, 99–104.
6. Herumurti, D., Uchimura, K., & Koutaki, G. (2013). Urban Road Network Extraction Based on Zebra Crossing Detection From a Very High-Resolution RGB Aerial Image and DSM Data. <https://doi.org/10.1109/SITIS.2013.24>
7. Kartika, D. S. Y., & Herumurti, D. (2016, October). Koi fish classification is based on HSV color space. In *2016 International Conference on Information & Communication Technology and Systems (ICTS)* (pp. 96–100). IEEE.
8. Kartika, D. S. Y., Herumurti, D., & Yuniarti, A. (2018). Local binary pattern method and feature shape extraction for detecting butterfly image. *International Journal*, 15(50), 127–133.
9. Satria, D., Kartika, Y., & Herumurti, D. (2016). Koi Fish Classification based on HSV Color Space. *International Conference on Information, Communication Technology, and System (ICTS)*, 5, 96–100. <https://doi.org/10.1109/ICTS.2016.7910280>
10. Kaya, Y., Kayci, L., & Tekin, R. (2013). A Computer Vision System for the Automatic Identification of Butterfly Species via Gabor- Filter-Based Texture Features and Extreme Learning Machine : GF + ELM. *TEM Journal*, 2(1).
11. Kayci, L., & Kaya, Y. (2014). A vision system for automatic identification of butterfly species using a grey-level co-occurrence matrix and multinomial logistic regression. *Zoology in the Middle East*, 60(1), 57–64. <https://doi.org/10.1080/09397140.2014.892340>
12. Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971–987. <https://doi.org/10.1109/TPAMI.2002.1017623>
13. Burçin, K., & Vasif, N. V. (2011). Down syndrome recognition using local binary patterns and statistical evaluation of the system. *Expert Systems with Applications*, 38(7), 8690–8695. <https://doi.org/10.1016/j.eswa.2011.01.076>

14. Singh, C., & Preet Kaur, K. (2016). A fast and efficient image retrieval system based on color and texture features. *Journal of Visual Communication and Image Representation*, 41, 225–238. <https://doi.org/10.1016/j.jvcir.2016.10.002>
15. Youssef, S. M. (2012). ICTEDCT-CBIR: Integrating curvelet transform with enhanced dominant color extraction and texture analysis for efficient content-based image retrieval. *Computers and Electrical Engineering*, 38(5), 1358–1376. <https://doi.org/10.1016/j.compeleceng.2012.05.010>
16. VijayaLakshmi, B., & Mohan, V. (2016). Kernel-based PSO and FRVM: An automatic plant leaf type detection using texture, shape, and color features. *Computers and Electronics in Agriculture*, 125, 99–112. <https://doi.org/10.1016/j.compag.2016.04.033>
17. Wang, J., Markert, K., & Everingham, M. (2009). Learning models for object recognition from natural language descriptions. *Learning*, 2.1-2.11. Retrieved from <http://eprints.pascal-network.org/archive/00006257/>
18. Junhua, C., & Jing, L. (2012). Research on Color Image Classification Based on HSV Color Space. 2012 Second International Conference on Instrumentation, Measurement, Computer, Communication, and Control, 255(3), 944–947. <https://doi.org/10.1109/IMCCC.2012.226>.
19. Kartika, D. S. Y., Herumurti, D., & Yuniarti, A. (2018). Butterfly image classification using color quantization method on hsv color space and local binary pattern. *IPTEK Journal of Proceedings Series*, (1), 78-82.
20. Khotimah, W. N., Arifin, A. Z., Yuniarti, A., Wijaya, A. Y., Navastara, D. A., & Kalbuadi, M. A. (2015, October). Tuna fish classification using decision tree algorithm and image processing method. In 2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA) (pp. 126-131). IEEE.



© 2019 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

