

**IMPLEMENTASI *PRINCIPAL COMPONENT ANALYSIS* (PCA)
UNTUK TEMU KEMBALI CITRA MOTIF KAIN TENUN NTT
BERDASARKAN WARNA DAN TEKSTUR**

Marwan E. Lubis¹, Silvester Tena² dan Sarline O. Manu³

¹*Program Studi Teknik Elektro, Universitas Nusa Cendana, Jl. Adisucipto Penfui Kupang
Email: marwanlubis51@gmail.com*

²*Studi Teknik Elektro, Universitas Nusa Cendana, Jl. Adisucipto Penfui Kupang
Email: siltena@staf.undana.ac.id*

³*Studi Teknik Elektro, Universitas Nusa Cendana, Jl. Adisucipto Penfui Kupang
Email: sarlince_manu@staf.undana.ac.id*

ABSTRAK

Nusa Tenggara Timur is one of the island province that has a diversity of cultures in each region. One of them is a woven fabric with distinctive motifs that vary both in terms of color, shape and texture. To classify the motifs of woven fabric manually will experience difficulty and longer time. This research aims to to apply the CBIR system method for Retrieval of NTT woven fabric motifs. Content Based Image Retrieval (CBIR) is a image search method based on the similarity of color, shape and texture. Feature extraction methods used are color features using color statistics. Texture features using texture histograms. Principal Component Analysis (PCA) used in the feature classification process. To compare the similarity of the test image and the image in the database using the Euclidean Distance method. The number of sample images is 12, consisting of 36 images in the data base and 24 test images. Testing process is also done by adding noise salt & pepper and noise gaussian. Based on the test results, the level of success and accuracy of image retrieval up to 94,3% of the 12 samples tested. On the addition of noise salt & pepper the method still survives at the 0.05 density level of 61% and the gaussian noise at the mean & variant 0.01 level of 66.5%. In order to prevent a mis retrieval image outside the database, the application uses a thresholding technique.

Keywords: CBIR, Motif Kain NTT, PCA, Color & Texture

Author : Marwan E. Lubis, Silvester Tena dan Sarline O. Manu

1. PENDAHULUAN

Nusa Tenggara Timur merupakan salah satu provinsi dengan jumlah pulau terbanyak di Indonesia, diantaranya Pulau Flores, Sumba, Timor, Alor, Lembata, Rote, Sabu dan pulau-pulau kecil lainnya. Setiap daerah memiliki keragaman budaya lokal yakni bahasa, alat musik, tarian dan rumah adatnya. Salah satu yang paling populer adalah kain tenun.

Kain tenun yang dimiliki berbagai daerah di NTT menjadi salah satu ikon nasional di bidang *fashion*. Kain tenun tersebut memiliki ciri khasnya masing-masing berdasarkan geografis, demografis dan kultur budaya daerah sehingga memiliki perbedaan warna, bentuk dan teksturnya. Pengelompokan motif kain tenun secara manual memerlukan waktu yang cukup lama. Pemanfaatan teknologi komputer sangat dibutuhkan dalam pengelompokan motif kain tenun. Salah satu proses penting dalam pengelolaan data yaitu temu kembali informasi yang dibutuhkan pengguna. Proses ini disebut dengan istilah temu balik informasi atau *Information Retrieval* (IR). Salah satu *Information retrieval* dalam temu kembali citra atau *Content Based Image Retrieval* (CBIR). *Content Based Image Retrieval* (CBIR) merupakan metode pencarian citra dengan melakukan perbandingan antara citra *query* dengan citra yang ada di *database* berdasarkan informasi yang ada pada citra tersebut.

Dalam studi kasus penelitian yang mengangkat kain tenun telah dilakukan oleh Agista Regiaswuri (2017), yaitu tentang penemuan kembali citra tenun dengan memanfaatkan transformasi wafalet diskrit untuk menganalisa fitur tekstur pada citra kain. Penelitian selanjutnya dilakukan oleh (Lamabelawa, 2016) yaitu tentang *Content Based Image Retrieval System* (CBIRS) pada kain tenun dengan memanfaatkan analisis tekstur. Berdasarkan studi kasus tersebut proses temu kembali atau CBIR hanya berdasarkan fitur tekstur yang digunakan, sehingga dalam tulisan ini ditambahkan ciri warna dalam proses temu kembali, sehingga hasil yang diharapkan akan memiliki keakuratan yang lebih tinggi.

Pada penelitian ini diterapkan metode *Principal Component Analysis* (PCA) sebagai ekstraksi fitur pada citra kain tenun. *Principal Component Analysis* (PCA) banyak digunakan dalam proses klasifikasi dengan cara

mengidentifikasi pola data dan kemudian mengekspresikan data tersebut ke bentuk yang lain untuk menunjukkan perbedaan dan persamaan antar pola.

2. DASAR TEORI

Content Based Image Retrieval (CBIR)

CBIR merupakan suatu aplikasi *Computer Vision* untuk masalah *image retrieval* (pengambilan citra), yaitu sebuah teknik yang menggunakan *content visual* dari citra tetap dalam mencari kesamaannya dalam *database* yang sangat besar menurut input atau sesuai dengan keinginan pengguna (*user*). Metode CBIR yang sering digunakan pada pencarian berdasarkan kemiripan warna, bentuk, dan tekstur. CBIR juga dapat diartikan sebagai teknik untuk mencari gambar yang berhubungan dan mempunyai karakteristik dari suatu kumpulan gambar.

Statistika Warna

Fitur ekstraksi ciri warna dapat dihitung dengan menggunakan metode statistik seperti rerata, deviasi standar, *skewness*, dan *kurtosis*. Sebagai contoh, penelitian yang dilakukan Kadir, dkk., 2011 dalam identifikasi tanaman hias dengan perhitungan dikenakan pada setiap komponen R, G dan B.

Nilai rerata (μ) menunjukkan ukuran distribusi dan dihitung menggunakan persamaan :

$$\mu = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^N p_i \quad (1)$$

Dengan M = nilai panjang piksel, N = nilai lebar piksel, p_i = rata-rata nilai piksel

Varian (σ) menyatakan luas sebaran distribusi. Akar kuadrat varians dinamakan sebagai deviasi standar. Rumus yang digunakan untuk menghitung deviasi standar :

$$\sigma = \sqrt{\frac{1}{M} \sum_{i=1}^M \sum_{j=1}^N (p_i - \mu)^2} \quad (2)$$

Skewness (θ) atau condongan adalah ukuran ketidak simetrisan. Apabila nilai Distribusi condong ke kiri maka nilai *skewness* berubah nilai negatif. Sebaliknya distribusi dikatakan condong ke kanan maka nilai *skewness* berupa bilangan positif. Persamaan untuk menghitung *Skewness* sebagai berikut :

$$\theta = \frac{\sum_{i=1}^M \sum_{j=1}^N (p_i - \mu)^3}{M^3} \quad (3)$$

Kurtosis (γ) adalah ukuran yang menunjukkan sebaran data bersifat meruncing atau menumpul. Kurtosis dihitung dengan persamaan seperti berikut :

$$\gamma = \frac{\sum_{i=1}^M \sum_{j=1}^N (p_i - \mu)^4}{M^4} \quad (4)$$

Ciri Tekstur

Untuk menghitung nilai tekstur suatu citra dapat digunakan metode histogram tekstur. Ada dua jenis citra yang memiliki tekstur yang berbeda, yaitu citra bertekstur halus memiliki daerah perubahan intensitas yang sempit dan citra bertekstur kasar memiliki kontras tinggi dengan intensitas yang lebar. Untuk mendapatkan nilai perbedaan histogram dapat dilakukan dengan cara menghitung nilai statis, *skewness*, energi, entropi dan *smoothness*. Fitur pertama yang dihitung secara statis adalah rerata intensitas, komponen fitur ini dihitung berdasarkan persamaan :

$$m = \sum_{i=0}^{L-1} i \cdot p \quad (5)$$

Pada persamaan diatas (m) adalah rerata intensitas dan i adalah aras keabuan pada citra f dan $p(i)$ menyatakan probabilitas kemunculan i dan L adalah nilai aras keabuan tertinggi, sehingga menghasilkan rerata kecerahan objek. Fitur kedua yang dihitung adalah deviasi standar dengan persamaan :

$$\sigma = \sqrt{\sum_{i=1}^{L-1} (i - m)^2 \cdot p(i)} \quad (6)$$

Pada persamaan diatas, σ^2 adalah varians atau momen orde dua ternormalisasi karena $p(i)$ merupakan fungsi peluang yang memberikan ukuran kekontrasan.

Fitur *skewness* adalah ukuran ketidaksimetrisan terhadap rerata intensitas dengan persamaan :

$$\sum_{i=1}^{L-1} (i - m)^3 p(i) \quad (7)$$

Deskriptor atau energi adalah ukuran distribusi intensitas piksel terhadap jangkauan aras keabuan menggunakan persamaan berikut:

$$\sum_{i=0}^{L-1} [p(i)]^2 \quad (8)$$

Entropi adalah kompleksitas citra, dimana semakin tinggi nilai entropi semakin kompleks citra tersebut. Entropi dapat mempersentasikan jumlah informasi yang terkandung pada sebaran data.

$$\sum_{i=0}^{L-1} p(i) \log_2(p(i)) \quad (9)$$

Properti kehalusan adalah tingkat nilai kehalusan/kekasaran intensitas pada citra, menggunakan persamaan:

$$R = 1 - \frac{1}{1+\sigma^2} \quad (10)$$

Dari rumus di atas semakin nilai R rendah menunjukkan bahwa citra memiliki intensitas kasar.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) adalah sebuah teknik *statistika* yang biasa dipakai dalam pengenalan, klasifikasi dan mereduksi data citra. Penggunaan PCA bertujuan untuk mengklasifikasi data yang lebih sederhana dengan mengurangi kumpulan data kompleks menjadi dimensi lebih kecil dengan mempertahankan variasi data yang ada.

Alur kerja PCA adalah menyediakan nilai variabel pada citra dengan cara mereduksi dimensinya, mentransformasi variabel-variabel asal yang saling berkorelasi menjadi variabel-variabel baru tanpa menghilangkan informasi penting yang ada di dalamnya atau disebut *principal component*. *Principal Component Analysis* (PCA) banyak digunakan untuk memproyeksikan atau mengubah data berukuran besar menjadi berukuran lebih kecil. Tahapan *Principal Component Analysis* dapat dijelaskan sebagai berikut :

1. Mengambil data pelatihan dari citra 2D.
2. Menghitung rata-rata data pelatihan
3. Melakukan normalisasi data pelatihan dengan melakukan pengurangan vektor citra dengan nilai vektor rata-rata.
4. Menghitung matriks kovarian
5. Hitung Eigen Value (λ_i) dan Eigen Vektor (V_i) dengan :

Euclidean Distance (ED)

Euclidean distance adalah perhitungan jarak dari dua buah titik dalam *euclidean space*. *Euclidean space* diperkenalkan oleh seorang matematikawan dari Yunani sekitar 300 Tahun Sebelum Masehi. Metode ED merupakan metode pengukuran jarak garis lurus antara dua titik, misalnya titik $X(X_1, X_2, \dots, X_n)$ terhadap titik $Y(Y_1, Y_2, \dots, Y_n)$

$$d_i(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (11)$$

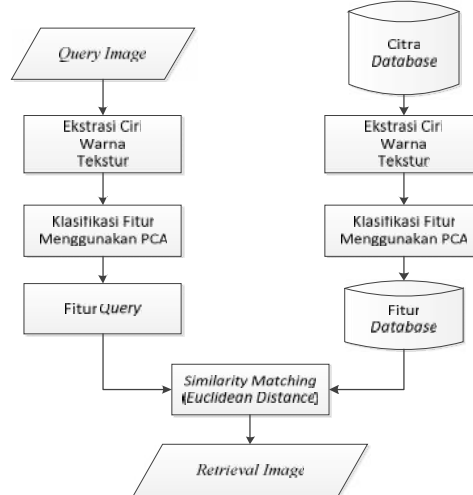
Dengan d = jarak *Euclidean* x = nilai vektor fitur data uji y = nilai vektor fitur data referensi atau data latih

3. .METODOLOGI PENELITIAN

Pada penelitian ini diambil citra kain tenun sebanyak 12 sampel. Pengambilan citra pada setiap sampel dilakukan sebanyak 5 kali. Dari ke-5 citra tersebut akan dibagi menjadi dua yakni tiga citra dilakukan sebagai citra latih dan dua citra sebagai citra uji. Jumlah total citra yakni 60 citra kain tenun yang terdiri dari 36 buah citra latih dan 24

buah citra uji. Untuk mendapatkan waktu komputasi yang lebih efisien setiap data citra diubah menjadi ukuran 200x150 piksel. Parameter yang digunakan dalam penelitian ini adalah Ciri Warna dan Tekstur dan Metode *Principal Component Analysis* Untuk mengukur tingkat keakuratan pengenalan citra kain tenun digunakan metode *Euclidean Distance*.

Alur kerja system temu kembali citra kain tenun dengan metode PCA dapat dilihat pada Gambar1.



Gambar 1 : Alur Kerja CBIR

Berdasarkan Gambar 1 tahap awal proses aplikasi dengan melakukan baca citra kain tenun. Citra yang dibaca yaitu berupa citra digital yang dimasukkan oleh *user* ke aplikasi (citra *query*) dan citra yang telah disimpan ke dalam media penyimpanan *database*. Setelah melakukan proses baca citra dilanjutkan dengan proses ekstraksi ciri pada gambar, ada dua fitur yang diekstraksi yaitu ciri warna dan ciri tekstur. Hasil proses ekstraksi menghasilkan vektor ciri pada citra uji. Proses yang sama juga dilakukan pada data-data gambar yang tersimpan dalam *database* seperti gambar *query* sehingga menghasilkan vektor ciri pada *database* yang tersimpan.

Klasifikasi fitur dilakukan pada vektor ciri warna dan tekstur untuk citra *query* dan citra di *database* dengan melakukan proses *Principal Component Analysis* (PCA). Data yang sudah mengalami klasifikasi fitur dilakukan proses perbandingan antara citra *query* dengan data yang tersimpan dalam basis data menggunakan perhitungan *euclidean distance* (perhitungan jarak kedua vektor). Data hasil perbandingan dilakukan *indexing* dan *retrieval* mulai data dari nilai jarak terkecil sampai nilai paling besar.

4. HASIL DAN PEMBAHASAN

Proses pengujian temu kembali citra dilakukan dengan perhitungan *euclidean distance* yaitu dengan membandingkan nilai PCA antara citra uji dengan citra latih dalam *database*. Setiap data citra uji yang diinputkan akan dibandingkan dengan 36 citra kain tenun di *database* dan aplikasi akan menampilkan proses temu kembali citra berdasarkan nilai *euclidean distance* yang didapatkan. Pengujian dilakukan dengan menggunakan sampel kain tenun Sumba.



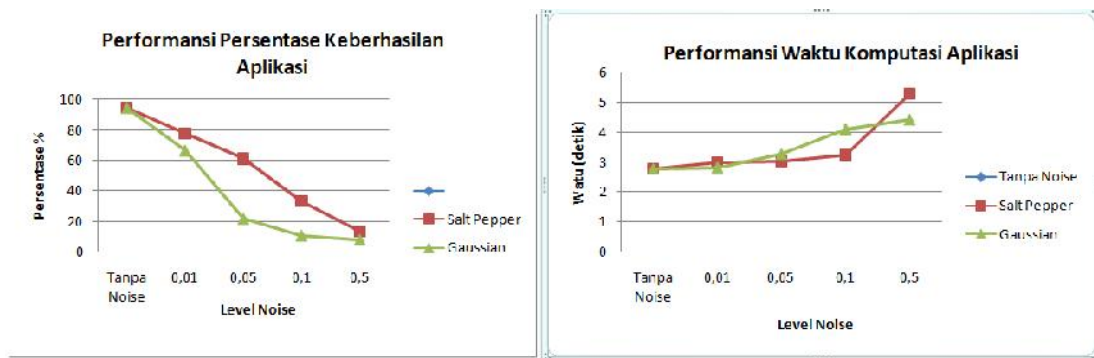
Gambar 2 : Pengujian temu kembali citra tanpa menggunakan noise

Berdasarkan hasil pengujian, jika semakin kecil nilai *euclidean distance* yang didapatkan, maka akan ditemukan kemiripan dengan citra uji. Nilai *euclidean distance* tersebut diurutkan dan ditampilkan pada tabel informasi aplikasi beserta nama kain tenun dari nilai yang terkecil sampai terbesar. Hasil Pengujian dapat dilihat pada Tabel 1.

Tabel 1. Pengujian Temu Kembali Citra

Data Sampel	Hasil Temu Kembali	
	Nama Kain Tenun	<i>Euclidean Distance</i>
 Kain Tenun Sumba	1	Kain Tenun Sumba 03
	2	Kain Tenun Sumba 01
	3	Kain Tenun Sumba 02
	4	Kain Tenun Malaka 01
	5	Kain Tenun Malaka 02
	6	Kain Tenun Solor 02
	7	Kain Tenun Solor 03
	8	Kain Tenun Solor 01
	9	Kain Tenun Malaka 03
	10	Kain Tenun Atambua 01
	11	Kain Tenun Atambua 02
	12	Kain Tenun Atambua 03

Pengujian tanpa noise menggunakan sampel kain tenun Sumba terdapat pada Tabel. 1 menunjukkan bahwa system temu kembali citra dapat bekerja dengan baik dan hasil temu kembali citra berupa tiga citra kain tenun Sumba yang terdapat dalam *database*, dengan jarak *euclidean distance* secara berurut-turut adalah 0,1169, 0,1247, 0,2518. Sedangkan citra yang memiliki kemiripan atau *relevan* dengan citra kain tenun Sumba adalah citra kain tenun Malaka, Solor dan Atambua. Untuk mengetahui performansi metode yaitu pengujian yang dilakukan untuk menentukan persentase secara keseluruhan sistem, dengan cara membandingkan satu citra *query* dengan semua citra yang ada dalam *database*. Pengujian performansi dilakukan menggunakan citra tanpa noise dan menggunakan noise. Pengujian menggunakan 12 sampel citra kain tenun, yang di uji dan diambil tiga jarak nilai *euclidean distance* terkecil dalam perankingan. Hasil kaseluruhan pengujian performansi aplikasi dapat dilihat di Grafik pada Gambar 3.



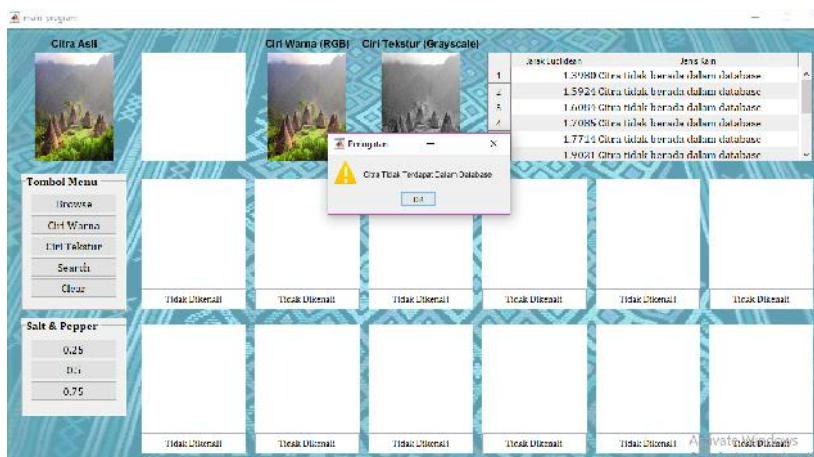
Gambar 3. Grafik Performansi Aplikasi dengan metode PCA

Berdasarkan grafik diatas menunjukkan bahwa hasil pengujian performansi metode PCA yang digunakan dalam ekstraksi fitur warna dan tekstur dalam proses temu kembali citra terhadap 36 citra pada *database* menghasilkan tingkat persentase keakuratan yakni 94,3%. Proses temu kembali citra dengan *noise salt & pepper* memiliki tingkat keberhasilan 77,66% pada nilai *density* 0.01 dan cenderung menurun pada nilai *density* 0.1 hingga 33,16%. *Noise gaussian* memiliki tingkat keberhasilan pengenalan mencapai 66,5 % pada nilai *mean & varian* 0.01 dan cenderung menurun pada nilai *mean & varian* 0.5 hingga 8,33%.

Dalam proses pengujian temu kembali citra rata-rata waktu komputasi yang diberikan sebesar 2,77 detik tanpa menggunakan *noise*. Pada penggunaan *noise salt & pepper* waktu komputasi yang diberikan sebesar 2,82 detik pada *noise* 0,01 dan makin meningkat pada *level noise* 0,5 dengan waktu komputasi sebesar 5,28 detik. Pada penggunaan *noise gaussian* waktu komputasi yang diberikan adalah 2,82 detik pada *level noise* 0,01 dan makin meningkat pada *level noise* 0,5 dengan waktu komputasi sebesar 4,43 detik.

Untuk membatasi temu kembali citra dari luar *database* maka digunakan nilai ambang. Nilai ambang diperlukan karena system akan menampilkan semua citra uji berdasarkan jarak terkecil. Nilai ambang dimaksudkan agar system hanya menampilkan citra uji yang memiliki rentang jarak *euclidean* antara 0 dan 1.

Hasil Pengujian menggunakan nilai ambang dapat dilihat pada Gambar 5. Dengan adanya nilai ambang, maka citra uji yang di luar *database* tidak dapat dikenali.



Gambar 4.5 Pengujian Citra Uji Luar Menggunakan Nilai Ambang

5. KESIMPULAN DAN SARAN

Kesimpulan

1. Penggunaan metode ekstraksi fitur warna dan tekstur menggunakan *Principal Component Analysis* dan *Euclidean Distance* dapat mempermudah proses temu kembali citra untuk klasifikasi kain Tenun Nusa Tenggara Timur. Persentase keberhasilan dan akurasi temu kembali citra tertinggi yakni 94,3% dari 12 citra uji yang digunakan.

2. Penambahan *noise salt and pepper* dan *gaussian* akan menurunkan kinerja metode yang digunakan, dalam proses temu kembali citra yang *relevan*. Penggunaan metode PCA, tingkat keberhasilan tertinggi *noise salt & pepper* pada nilai *density* 0.01 sebesar 77,6% dan pada *noise gaussian* dengan nilai *mean & varian* 0.01 sebesar 66,5%. Penambahan nilai *density, mean dan varian* pada citra uji akan menurunkan tingkat persentase keberhasilan aplikasi dalam proses temu kembali citra.
3. Waktu komputasi dalam proses temu kembali citra sangat dipengaruhi oleh proses kerja yang ditambahkan dalam aplikasi. Penambahan proses kerja *noise salt & pepper* dan *gasussian* pada citra akan memberikan waktu komputasi lebih besar dibandingkan tanpa *noise*. Rata-rata waktu komputasi yang diberikan aplikasi tanpa menggunakan *noise* adalah 2,98 detik dan kenaikan proses waktu komputasi terjadi pada pemberian *noise salt & pepper* 0.5 sebesar 5,28 detik dan *noisegaussian* 0,5 sebesar 4,43 detik.
4. Penggunaan nilai ambang bertujuan agar sistem hanya menampilkan citra yang *relevan* dalam *database*.

DAFTAR PUSTAKA

- Basuki, Ahmad. (2005). *Pengolahan Citra Digital Menggunakan Visual Basic*. Graha Ilmu: Yogyakarta.
- Hastuti, I; Hariadi, M.; Purnama, I.K.E. (2011). *Content Based Image Retrieval Berdasarkan fitur bentuk Menggunakan Metode Gradient Vektor Flow Snake*. Seminar Nasional Teknik Informatika. UPN Yogyakarta.
- Kadir, A. Nugroho, L.E.; Susanto, A. and Santosa. (2011). ‘‘Foliage Plant Retrieval Using Polar Fourier Transform, Color Moments and Vein Features. On Signal and Processing’’. *An International Journal*, Vol. 2, No. 3. Universitas Gadjah Mada Yogyakarta.
- Lamabelawa, M.J.L. (2016) . ‘‘Perbandingan Ekstraksi Fitur Tenun Ikat NTT Berbasis Analisis Tekstur’’. *Jurnal HOAQ-Teknologi Informasi* Vol.7, No.1. Teknik Informatika Stikom Uyelindo Kupang.
- Mardiyana, A. E. (2014). *Aplikasi Content Based Image Retrieval Motif Batik Menggunakan Metode Wavelet*. Skripsi S-1. Universitas Islam Negeri Maulana Malik Ibrahim Malang.
- Regiaswuri, Agista. (2017). *Penemuan Kembali Citra Tenun Dengan Kemiripan Motif menggunakan DWT*. Skripsi S-1. Institut Teknologi Sepuluh Nopember Surabaya.
- Susanto, A. K. (2013). *Teori dan Aplikasi Pengolahan Citra*. CV.Andi Offset. Yogyakarta.
- Utama, Juna.(2011). ‘‘Akusisi Citra Digital Menggunakan Pemrograman Matlab’’. *Majalah Ilmiah Unikom*, Vol.9, No.1. Universitas Komputer Indonesia Bandung.