

Klasifikasi Stempel Dokumen Berwarna Menggunakan Fuzzy Integral dan Artificial Neural Networks

¹A. Budiman, ²B.H Iswanto

^{1,2}Lab Instrumentasi dan Komputasi

Jurusan Fisika, FMIPA, Universitas Negeri Jakarta

JL. Pemuda 10 Rawamangun, Jakarta Timur 1322, Indonesia

Tel : 021-47885283

¹ budimanok@gmail.com,

² bhionet@gmail.com

Received date : 5 Desember 2010

Accepted date : 26 Februari 2011

Abstrak

Berkembangnya teknologi pencetakan meningkatkan rawannya tindakan pemalsuan, maka dibutuhkan sistem identifikasi dan klasifikasi yang bersifat non-destruktif yang dapat diterapkan di dokumen. Stempel pada dokumen berwarna dapat dijadikan objek penciri keaslian suatu dokumen. Integral fuzzy Sugeno dan Choquet dipilih sebagai operator terbobot fusi data sinyal multi-kanal untuk ekstraksi stempel sebagai objek penciri. Metode ini dapat memberikan variasi maksimum pada citra keabuan terhadap cluster stempel warna yang ada pada citra dokumen. Hasil ekstraksi ciri diukur tingkat *eccentricity*, yang menunjukkan hasil ekstraksi ciri stempel masih berbentuk melingkar (*roundnes*). Tingkat pengenalan data uji menggunakan ANN bernilai 100% dari keseluruhan citra uji, yang berarti *artificial neural networks* dapat mengenali semua stempel uji. Pada tahap pengujian keaslian dokumen, menguji stempel hasil cetak printer menghasilkan tingkat pengenalan keaslian sebesar 80,1%.

Kata kunci: *integral fuzzy, artificial neural networks, pengolahan citra.*

Abstract

The development of printing technology, increase the susceptibility of counterfeiting measures, the identification and classification system is needed that is non-destructive that can be applied in the document. Stamp color at documents can be made as the object feature of the authenticity of a document. Sugeno and Choquet fuzzy integral is selected as the operator weighted data fusion of multi-channel signal for the extraction of stamps as objects feature. This method can give a maximum variation in the gray image of the cluster color stamp on the document image. The result of feature extraction measured levels of eccentricity, which shows the results of feature extraction is still a circular stamp (roundness). Recognition level test data using ANN worth 100% of the overall image of the test, which means the artificial neural networks identify all test seals. At the authenticity of documents testing that examine stamp that prints by printer, resulting in recognition rate of 80.1% authenticity.

Keywords: *fuzzy Integral, artificial neural network, image processing.*

1 Pendahuluan

Penentuan keaslian suatu dokumen memerlukan objek-objek penciri sebagai indikator, misalnya logo, tanda tangan, jenis kertas, warna dan jenis tulisan, dan penciri lain seperti penggunaan pengkode garis (*barcode*), nomor seri, tanggal, bahan warna tertentu dan penanda air (*watermarking*).

Objek-objek seperti logo, stempel, tanda tangan dan penanda khas lainnya, memiliki keunikan atau ciri-ciri yang khusus (pembeda). Stempel pada dokumen dapat digunakan

sebagai autentifikasi dan syarat sahnya suatu dokumen. Pengujian keaslian secara konvensional seringkali gagal karena keterbatasan kemampuan visual kita, apalagi jika pengujian harus dilakukan terhadap sejumlah besar dokumen. Karena itu suatu sistem yang cerdas dan secara otomatis dapat mengidentifikasi keaslian suatu dokumen sangat diperlukan.

2 Fuzzy Integral

Salah satu pendekatan fuzzy dalam fusi data adalah teori fuzzy Sugeno [7]. Teori fuzzy ini mentransformasikan harga input x_i ke dalam bilangan fuzzy $[0,1]$. Transformasi ini secara matematik dapat dituliskan.

Jika X suatu himpunan dan B bagian dari X . Dengan A fungsi dari $g: \rightarrow [0,1]$, fuzzy measure B terpenuhi dengan tiga kondisi:

1. *Boundary condition* $g(\phi) = 0, g(B) = 1$
2. *Monotonicity* if $g(A) \leq g(B)$ if $A \subset B$, dan $A, B \in B$
3. *Continuity* : $\lim g(A_i) = g(\lim A_i)$ if $A_i \in B$

Fuzzy measure dijelaskan oleh Sugeno [7] yang memenuhi kondisi yang disebut λ -rule ($\lambda > -1$):

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B) \tag{1}$$

Jika $X = \{x_1, x_2, \dots, x_n\}$, dan $g_i = g(\{x_i\}), i = 1, \dots, n$, maka fuzzy measure $g_\lambda(A)$:

$$\begin{aligned} g_\lambda(A) &= \sum_{i=1}^n g_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n g_i g_j + \dots + \lambda^{n-1} g_1 g_2 \dots g_n \\ &= \frac{1}{\lambda} \left[\prod_{x_i \in A} (1 + \lambda g_i) - 1 \right], \lambda \neq 0 \end{aligned} \tag{2}$$

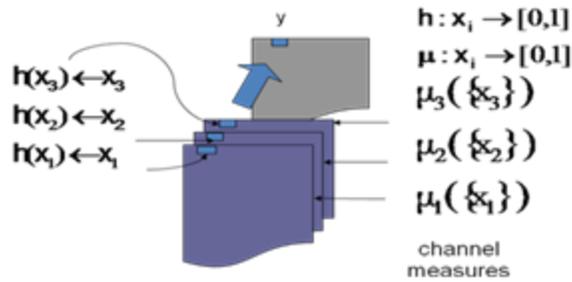
Fuzzy integral adalah suatu operator terbobot, yang disebut ukuran fuzzy (fuzzy measure) yang memboboti data input. Metode ini dapat memberikan variasi maksimum pada citra keabuan terhadap cluster stempel warna yang ada pada citra dokumen. Dengan teknik ini stempel pada dokumen dapat diekstrak sebagai objek pencari. Fuzzy integral atau biasa disebut Sugeno integral dapat dipandang sebagai operator penjumlahan :

$$I_s(k) = \bigvee_{i=1}^n [h_{(i)}(x_i) \wedge \mu(A_{(i)})] \tag{3}$$

Choquet integral:

$$C_\mu[x_1 \dots x_n] = \sum_{i=1}^n h_{(i)}(x_i) \cdot [\mu(A_{(i)}) - \mu(A_{(1-i)})] \tag{4}$$

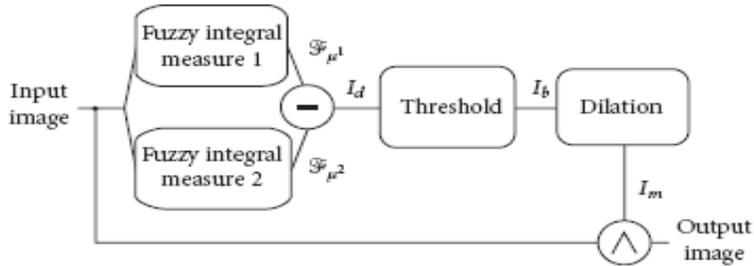
Dalam konteks pengolahan citra, penerapan integral fuzzy Sugeno dapat digambarkan seperti pada Gambar 1.



Gambar 1 Representasi Fuzzy Integral dalam citra multikanal

3 Langkah Kerja untuk Deteksi Kluster Warna

Strategi yang digunakan meninjau komputasi dari integral fuzzy pada kanal-kanal warna daripada input citra dengan dua ukuran fuzzy (*fuzzy measure*) yang berbeda. Ukuran fuzzy ini dipilih agar cluster warna dideteksi untuk memberi dampak maksimal dengan mengubah koefisien-koefisiennya. Kemudian dua citra keabuan diperoleh (F_1 dan F_2). Berikutnya citra-citra yang berbeda dari hasil ini dikomputasi untuk mendapatkan citra selisih (*different image*) I_d dan dipotong (*thresholded*) untuk mendapatkan suatu *mask biner*. Mask ini digunakan untuk proses dilasi (*dilation*) terhadap citra input untuk *segmentasi stempel*.



Gambar 2 Proses segmentasi objek penciri dari citra berwarna

Integral fuzzy akan dipilih dalam penelitian ini sebagai operator terbobot fusi data sinyal multi-kanal. Ada beberapa tipe integral fuzzy, tapi disini akan digunakan integral fuzzy Sugeno dan Choquet. Model warna RGB pada citra digital dari dokumen selanjutnya dipandang sebagai sumber informasi dari tiga kanal sinyal warna. Fuzzy integral adalah suatu operator terbobot, yang disebut ukuran fuzzy (*fuzzy measure*) yang memboboti data input. Metode ini dapat memberikan variasi maksimum pada citra keabuan terhadap cluster stempel warna yang ada pada citra dokumen. Dengan teknik ini stempel pada dokumen dapat diekstrak sebagai objek penciri.

Ekstraksi objek penciri dari stempel berwarna pada dokumen dilakukan dengan memindai (*scanning*) secara optik ke dalam bentuk citra digital berwarna. Citra dalam model warna RGB diperoleh sebagai citra input sebagai berikut:

$$I(x,y) = [I_r(x,y), I_G(x,y), I_B(x,y)] \quad (5)$$

Selanjutnya citra beda $I_d(x, y)$ diperoleh dengan menerapkan

$$I(x, y) = [I_r(x, y), I_G(x, y), I_B(x, y)] \quad (6)$$

F_{μ^i} adalah citra-citra yang dihasilkan dari komputasi integral fuzzy dengan ukuran fuzzy μ^i pada tiap piksel warna $x_1 = I_R(x, y)$, $x_2 = I_G(x, y)$, dan $x_3 = I_B(x, y)$. Integral fuzzy Sugeno untuk suatu piksel selanjutnya dapat dihitung:

$$S(x_1, x_2, \dots, x_n) = \bigvee_{i=1}^n [h(x_i) \wedge \mu(A_i)] \quad (7)$$

Dalam citra biner setiap piksel hanya mempunyai dua kemungkinan nilai yaitu 0 dan 1. Sebuah citra biner dapat dianggap sebagai tipe khusus dari citra intensitas yang hanya berisi hitam dan putih. Citra biner $I_b(x, y)$ didapat dengan menerapkan pembatas (threshold) sbb.

$$I_b(x, y) = \begin{cases} 1, & I_d(x, y) \geq \theta, \\ 0, & I_d(x, y) < \theta, \end{cases} \quad (8)$$

Pada algoritma citra biner berkerja dengan sekumpulan grup piksel yang disebut *neighborhoods*, yang didasarkan pada lokasi relative terhadap piksel tersebut. Sembarang prosedur biner didasarkan pada analisis histogram dapat diterapkan untuk tujuan ini. Untuk mendapatkan bagian yang mungkin salah deteksi pada citra, suatu citra mask $I_m(x, y)$ hasil-hasil dari aplikasi dari suatu dilasi morfologi diiterasi sekali pada citra ini:

$$I_m(x, y) = I_b(x, y) \oplus S \quad (9)$$

Dimana S adalah suatu elemen struktur yang biasanya diambil sebagai suatu 4-tetangga terdekat. Proses masking dilakukan untuk menutupi bagian sudut-sudut citra sehingga mengurangi kemungkinan variasi data yang muncul pada bagian-bagian tersebut. Terakhir citra keluaran $I_o(x, y)$ dihitung dengan memfilter citra input dengan mask yang diperoleh dengan operator logika AND:

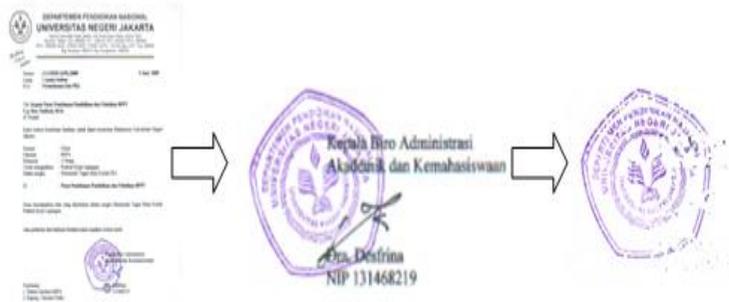
$$I_o(x, y) = I(x, y) \wedge I_m(x, y) \quad (10)$$

Moment citra sangat berguna untuk menggambarkan objek setelah segmentasi. *Moment* dapat menggambarkan suatu objek dalam hal area, posisi, orientasi dan parameter terdefinisi lainnya. Dalam pembuatan vektor ciri untuk pengenalan objek. Besarnya nilai *exentricity* suatu objek dapat diukur sebagai :

$$\varepsilon = \frac{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}{\mu_{00}} \quad (11)$$

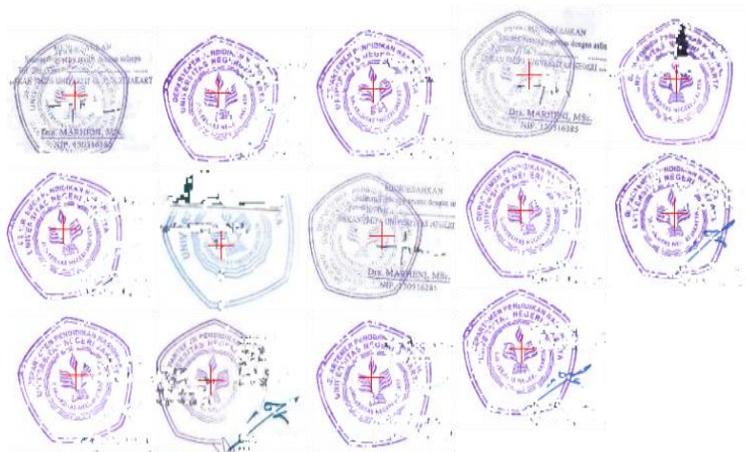
4 Pengujian Fuzzy Integral untuk Ekstraksi Ciri

Ekstraksi ciri menggunakan *algoritma fuzzy integral* yang telah diimplementasikan dalam bentuk software, diuji dengan pengujian segmentasi terhadap stempel dengan berbagai warna.



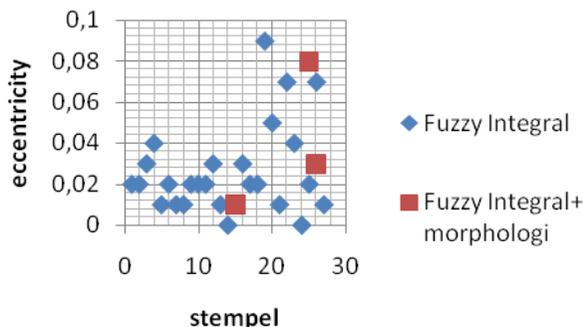
Gambar 3 Hasil ekstraksi ciri pada dokumen berwarna

Stempel yang tersegmentasi menyajikan sebuah bentuk melingkar (*circular*) seperti yang digambarkan dalam Gambar 4.



Gambar 4 Perhitungan momen citra pada Stempel yang telah diekstraksi dari dokumen

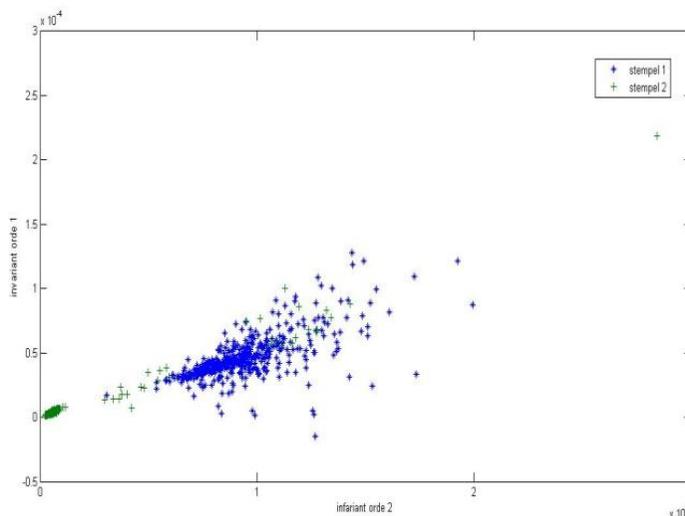
Koefisien eksentrisitas (11) dihitung pada gambar perbedaan $I_d(x, y)$, yang diperoleh dari Persamaan (6). Nilai riil ϵ , yang menjadi ciri bentuk informasi, memiliki rentang 0.0 untuk bentuk objek melingkar dan 1,0 untuk suatu objek yang berbentuk garis [11].



Gambar 5 Grafik stempel setelah proses morfologi

Momen (*moment*) dapat menggambarkan suatu objek dalam hal ini area, posisi, orientasi dan parameter terdefinisi lainnya [11]. Dari *moment* ternormalisasi sekumpulan moment invarian dapat didefinisikan.

Momen ini digunakan dalam pembuatan vektor pencari untuk pengenalan suatu objek. Objek yang sama akan memiliki moment yang sama walaupun ukurannya berbeda. Pada Gambar 6 dapat dilihat pada objek bentuk yang sama, maka hasil pada moment invarian akan cenderung sama.



Gambar 6 Moment invarian dari dua macam stempel

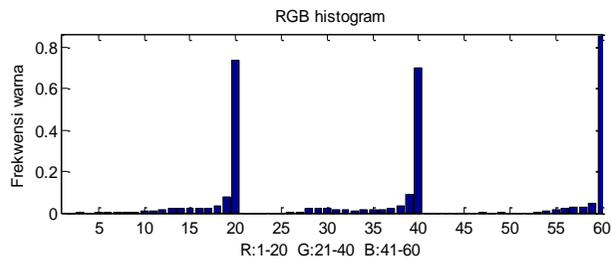
5 Pengujian Keaslian

Warna Stempel pada dokumen dapat menjadi objek pencari untuk identifikasi keaslian dokumen. *Artificial Neural Networks* digunakan untuk identifikasi warna stempel sehingga warna stempel dapat terklasifikasi.



Gambar 7 Stempel berwarna untuk uji ANN

Pelatihan ANN akan menghasilkan output yang baik jika input memiliki pola tertentu untuk dikenali. Pola yang baik akan memiliki perbedaan antara satu dengan yang lainnya. Nilai-nilai RGB setiap piksel dalam gambar dikatalogkan kedalam satu histogram frekuensi, dimana setiap komponen ruang warna (R,G,B) diwakili oleh N kelompok. Ketiga histogram itu kemudian ditambahkan sebagai salah satu vektor, maka vektor input akan berukuran $X = 3N$ sebagai masukan ANN.



Gambar 8 Histogram warna RGB yang sudah dinormalisasi dan disatukan dalam satu vektor

Tingkat pengenalan data uji bernilai 100% dari keseluruhan citra uji, yang berarti pola *neural network* dapat mengenali semua stempel uji.

Tabel 1 Pengujian Neural Networks

Kelompok Citra	Jumlah Kesalahan Pengenalan Citra	Prosentase Keberhasilan
Stempel 1	0/32	100%
Stempel 2	0/32	100%
Stempel 3	0/32	100%
Stempel 4	0/32	100%
Stempel 5	0/32	100%
Stempel 6	0/32	100%
Total	0/192	100%

Pengujian keaslian dilakukan dengan memindai menggunakan scanner stempel berwarna yang sebelumnya telah diujicoba pada *Artificial Neural Networks(ANN)*, stempel uji kemudian dicetak dengan printer. Stempel hasil printer bukan merupakan stempel asli.

Tabel 2 Pengujian keaslian stempel

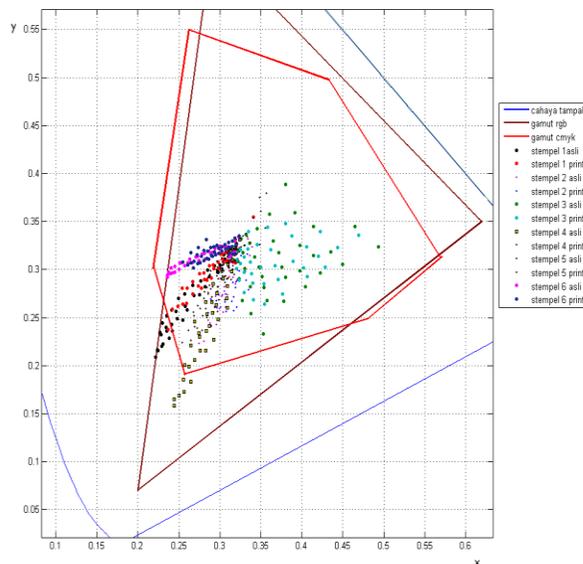
Kelompok Citra	Jumlah Pengenalan Kesalahan	Prosentase pengenalan kesalahan
Stempel Warna 1	21/32	65,6%
Stempel Warna 2	2/32	6,25%
Stempel Warna 3	0/32	0%
Stempel Warna 4	23/32	71,6%
Stempel Warna 5	0/32	0%
Stempel Warna 6	0/32	0%
Dokumen Uji	25/31	80,1%

Pada pengujian keaslian diharapkan citra stempel hasil cetak dengan printer menghasilkan pola keluaran pada *neural network* yang berbeda dengan stempel aslinya. Dapat dilihat pada kelompok stempel 1 memiliki prosentase pengenalan kesalahan sebesar 65,6% pada kelompok stempel 1 dan 71,6 % pada kelompok stempel 4, pada stempel 4 ini mengandung perak nitrat 3,8% di dalam tinta. Nilai prosentase diatas menunjukkan bahwa AAN dapat membedakan stempel yang dicetak dengan printer dan yang aslinya.

Sedangkan pada kelompok stempel 2,3,4 dan 6 memiliki prosentase kesalahan yang kecil, ini menunjukkan *neural networks* dapat beradaptasi pada kelompok warna tersebut. Kesalahan menunjukkan bahwa mesin cetak dalam hal ini printer tidak dapat menduplikasi warna tersebut baik pada stempel 1, stempel 4 dan stempel pada dokumen uji.

Kesalahan duplikasi warna terjadi karena gamut warna atau *colorspace* yang merupakan batasan warna, masing - masing mode warna RGB ataupun CMYK mempunyai batasan warna (gamut) yang berbeda. Warna CMYK adalah warna yang berdasar pada pigmen warna yang umumnya dipakai dalam teknologi pencetakan. Sedangkan RGB adalah model warna pencahayaan (*additive color mode*) dipakai untuk perangkat masukan seperti

scanner maupun perangkat keluaran seperti display monitor, warna – warna primernya (*red, blue, green*) tergantung pada teknologi alat yang dipakai seperti CCD atau PMT pada scanner atau digital camera, CRT atau LCD pada display monitor.



Gambar 9 Sebaran spektrum warna CIE xzy pada stempel uji

Jika gamut dari dua model warna tidak setara maka kita dapat menggunakan model CIE XYZ sebagai pembandingnya [6]. Pada Gambar 9 dapat dilihat stempel 1 dan 4 sebaran warna melewati gamut warna cmyk. Sehingga pada saat pencetakan dengan printer warna – warna yang melewati batas gamut cmyk akan dikonversi hanya menyerupai tapi tidak sama dengan warna-warna aslinya.

6 Kesimpulan

Fuzzy integral dapat mengextraksi dan meningkatkan kualitas objek pencari, yang dikarenakan objek pencari tidak seragam baik dalam hal tingkat kecerahan warna, kekusaman kertas, dan kontras warna. Pencampuran tinta dengan bahan lain seperti perak klorida, menghasilkan tingkat pengenalan keaslian pada ANN sebesar 71.6%.

7 References

- [1] Aureli, Soria-Frisch., The Fuzzy Integral for Color Seal Segmentation, IEEE, 2003.
- [2] Aureli, Soria-Frisch., Color Seal Extraction From Documents: Robustness Through Soft Data Fusion, Applied Signal Processing, 2005.
- [3] Zadeh, Lotfi A., Fuzzy Set, Information and Control, vol. 8, no.3, 1965.
- [4] Grabisch, M., Murofushi, T. & Sugeno, M., Fuzzy Measures of Fuzzy Events Defined by Ffuzzy Integrals, Fuzzy Sets and Systems 50, 1992.
- [5] Grabisch, M., Fuzzy integral in multicriteria decision making. Fuzzy Sets & Systems, 1995.
- [6] Ebner, Marc., Color Constancy. Jhon Wiley & Sons ltd, 2007.
- [7] Sugeno, M., Fuzzy Measures and Fuzzy Integrals. Trans. S. I. C. E., Vol. 8, No. 2, 1972.

- [8] Munakata T., Fundamental of The New Artificial Intelligence, London: Springer-Verlag, 2008.
- [9] Kosko, Bart., Neural Networks and Fuzzy Systems: A Dynamical System Approach to Machine Intelligence. Prentice-Hall, 1992.
- [10]Gonzales, Rafael C., Digital Image Processing. Printice-Hall, 2001.
- [11]Jahne, Bernd. Digital Image Processing, Springer, 2002.
- [12]Jang, Jyh-Shing Roger., Neuro-fuzzy and Soft Computing : A Computational Approach to Learning and Machine Intelligence, Prentice-Hall, 1997.