

EKSTRAKSI FITUR MOTIF BATIK BERBASIS METODE STATISTIK TINGKAT TINGGI

Mulaab

Jurusan Teknik Informatika Universitas Trunojoyo
Jl. Raya Telang PO BOX 2 Kamal Bangkalan Telp (031)-3011147
e-mail : mulaab@if.trunojoyo.ac.id

Abstrak

Batik merupakan karya seni budaya bangsa Indonesia yang dihasilkan dari proses melukis pada kain. Motif batik merupakan penyederhanaan unsur bentuk alam untuk melambangkan sesuatu. Banyak hal yang mempengaruhi ragam motif batik diantaranya letak geografis, sifat dan tata kehidupan daerah, kepercayaan dan adat istiadat serta alam sekitar termasuk flora dan fauna, simbolisme dengan falsafah kehidupan. Faktor-faktor ini diperlihatkan dengan kekuatan warna dan ragam hias pada masing-masing kain batik.

Pada penelitian ini akan dibahas tentang bagaimana mengekstraksi fitur motif batik pada batik pesisir dengan menggunakan metode statistik tingkat tinggi. Ekstraksi fitur batik dapat digunakan untuk membangun motif batik yang baru dari motif batik yang sudah ada.

Keyword : *Batik, ekstraksi fitur, statistik tingkat tinggi, Independent Component Analysis (ICA), Pearson System*

1. PENDAHULUAN

Batik adalah warisan tradisional yang terkenal dan unik dari Indonesia. Keunikannya berasal dari proses produksinya yaitu yang dikenal dengan "mbatik", motif, dan nilai-nilai yang terkandung di dalamnya. Karena nilai budaya dan nilai seninya, maka batik menjadi produk yang bernilai ekonomis tinggi di era modern ini. Namun, selain sebagai produk ekonomi, batik memiliki karakteristik pada motif nya. Motif motif dan ragam hiasnya, yang lahir dan dibangun dari proses kognitif manusia yang diperoleh dari sekitarnya. Hal inilah dianggap sebagai salah satu aspek yang menarik untuk diteliti menggunakan sains dan teknologi. Pada penelitian ini, mengadopsi tentang bagaimana pemrosesan citra dapat digunakan untuk mengenali fitur dari citra batik. Istilah fitur dari motif batik adalah menyatakan representasi suatu fungsi citra kain batik yang digunakan dalam pemrosesan visual lebih lanjut. Ekstraksi fitur dari citra batik merupakan proses untuk mendapatkan ciri dari dari persepsi visualnya. Selanjutnya untuk mempermudah pemahaman, pada makalah ini akan disusun sebagai berikut; pada bab 2 membahas tentang gambaran umum tentang batik, pengelompokan batik dan ciri khasnya dari masing-masing kelompok. Bab 3 teori menjelaskan tentang representasi feature dan model statistik dari citra. Bab 4 menjelaskan tentang model Independent Component Analysis, pada bab ini akan dijelaskan konsep dasar pemisahan fitur citra, dalam hal ini fitur dianalogikan sebagai sinyal atau sumber. Bab 5 menjelaskan tentang desain sistem yang digunakan untuk melakukan proses ekstraksi fitur citra batik. Bab 6 Adalah proses ekstraksi fitur citra dari 13 ragam motif batik pesir.

2. KAIN BATIK

Definisi batik (Hamzuri, 1985) adalah suatu cara membuat desain pada kain dengan cara menutup bagian-bagian tertentu dari kain dengan malam. Batik pada awalnya merupakan lukisan atau gambar pada mori yang dibuat dengan menggunakan alat bernama canting. Dalam perkembangan selanjutnya dipergunakan alat-alat lain yang lebih baik untuk mempercepat proses pengerjaannya misalnya dengan cap. Membatik sendiri adalah suatu pekerjaan yang mengutamakan ketiga tahapan proses, yaitu pemalaman, pewarnaan dan penghilangan malam. Berapa banyak pemalaman atau berapa kali penghilangan malam akan menunjukkan betapa kompleks proses yang dilakukan, sehingga akan menghasilkan lembaran batik yang kaya akan paduan warna. Batik merupakan salah satu bahan busana yang banyak dikenakan orang Jawa. Seni batik berbeda dengan seni yang lain, dilihat pada kedalaman maknanya. Macam-macam corak batik memiliki arti sendiri-sendiri dimana terdapat perbedaan batik mana yang boleh dikenakan oleh golongan raja/bangsawan dan rakyat biasa. Penggunaan batik sangat dipengaruhi oleh aktivitas masyarakat berbentuk ceremonial, ritual dan historis kultural, serta hal-hal yang bersifat dan berunsur filosofis.

Menurut Nian S. Djoemena (Joemana, 1986), secara garis besar terdapat 2 golongan ragam hias batik, yaitu ragam hias *geometris* dan ragam hias *non-geometris*.

Yang termasuk golongan geometris adalah:

1. *Garis miring* atau *parang*
2. *Garis silang* atau *ceplok*
3. *Anyaman* dan *Limar*

Yang termasuk golongan non-geometris adalah:

1. *Semen*, terdiri dari flora, fauna, meru, lar dan sejenis itu yang ditata secara serasi.
2. *Lunglungan*
3. *Buketan*, dari kata bahasa Prancis atau Belanda *bonquet* jelas merupakan ragam hias pengaruh dari luar dan termasuk ragam hias pesisir

Sejak zaman penjajahan Belanda, batik ditinjau dari daerah penghasilnya, dibagi menjadi dua kelompok, yaitu:

- a. Batik Vorstenlanden
Yaitu batik dari daerah pedalaman (Surakarta dan Yogyakarta). Di zaman penjajahan Belanda, kedua daerah ini merupakan daerah kerajaan dan dinamakan daerah Vorstenlanden, hingga saat ini kedua kerajaan itu masih memiliki kharisma.
- b. Batik Pesisir
Batik pesisir merupakan batik yang pembuatannya dikerjakan diluar daerah pedalaman (Surakarta dan Yogyakarta), yang termasuk daerah pesisir adalah daerah yang terdapat disepanjang pantai utara Jawa, seperti Jakarta, Indramayu, Cirebon, Pekalongan, Lasem, Garut, Madura dan Jambi.

Pembagian asal batik ini, terutama berdasarkan sifat corak dan warna dasarnya, serta keunikan dari daerah masing-masing

Secara garis besar ciri khas dari kedua kelompok tersebut, yaitu:

1. Batik Pedalaman (Vorstenlanden), khususnya daerah Surakarta dan Yogyakarta, memiliki ciri-ciri sebagai berikut:
 - Ragam hias motif batiknya bersifat simbolisme berlatar belakang kebudayaan Hindhu-Jawa.
 - Warna sogan, indigo (biru), hitam dan putih
2. Batik pesisir memiliki ciri-ciri sebagai berikut:
 - Ragam hias motif batiknya bersifat natural dan mendapat pengaruh kebudayaan asing secara dominan.
 - Warna beraneka ragam

3. REPRESENTASI FITUR CITRA

Pendekatan klasik untuk merepresentasikan citra [Hyv'arinen, 2009] adalah jumlahan dari bobot fitur secara linier. Misalkan dinotasikan masing-masing fitur dengan $A_i(x,y)$, $i=1, \dots, n$. Fitur ini diasumsikan tetap. Pada suatu citra, koefisien dari masing-masing fitur dari citra tersebut dinyatakan dengan s_i . Secara aljabar, dapat ditulis dengan

$$I(x,y) = \sum_{i=1}^n A_i(x,y) s_i \quad (1)$$

n adalah jumlah pixel dalam citra tersebut. Sehingga untuk sebuah citra, s_i dapat ditentukan dengan persamaan sebagai berikut

$$s_i = \sum_{x,y} W_i(x,y) I(x,y) \quad (2)$$

untuk suatu bobot invers tertentu W . W_i disebut dengan deteksi fitur.

Statistik dari fitur

Sifat-sifat statistik yang paling mendasar dari citra diperoleh dengan histogram dari output s_i yang dihasilkan dari detektor-detektor fitur. Misalkan dinotasikan output dari detektor fitur linier tunggal dengan bobot-bobot $W(x,y)$ adalah s , maka

$$s_i = \sum_{x,y} W_i(x,y) I(x,y)$$

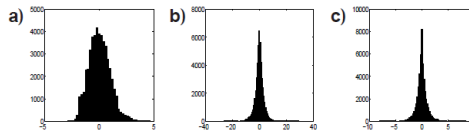
Intinya dari persamaan diatas adalah mencari statistik dari output, dengan input dari detektor adalah potongan-potongan dari sebuah citra. Potongan-potongan citra artinya subcitra (windows) yang diperoleh secara acak dari citra. Sehingga fitur-fitur s adalah variabel acak, dan untuk setiap potongan input didapatkan dari suatu variabel random.

Statistik dari output s dapat dilihat dari histogramnya. Misalkan dalam gambar 1 didapatkan output s dari 3 macam detektor yang berbeda, yang pertama adalah *dirac detector* yaitu detektor dimana hanya satu bobot $W(x,y)$ yang nilainya tidak nol, yang kedua adalah detektor dengan fungsi fourier satu dimensi (one-dimensional grating), dan yang ketiga adalah Gabor edge detector (Fungsi gabor).



Gambar 1. Tiga macam detektor (filter)
a. Dirac detector b. Fourier one-dimensional (grating) c. Gabor detector

Dari ketiga detektor fitur tersebut didapatkan histogramnya seperti pada gambar 2 berikut



Gambar 2. Histogram dari output 3 macam detektor : a. output Dirac detector b. Output Fourier one-dimensional (grating) c. Output Gabor detector

Dari tiga histogram terlihat, bahwa penggunaan detektor (filter) fitur yang berbeda maka akan menghasilkan statistik dari citra yang berbeda. Sehingga ini menjadi dasar untuk mengoptimalkan nilai statistik dari fitur-fitur s_i . Artinya bagaimana mendapatkan fitur-fitur yang secara statistik adalah saling bebas. Fitur-fitur dalam sebuah citra memiliki ketidakbergantungan secara linier dan dapat diekstrak. [Hoyer dkk, 2000]

4. INDEPENDENT COMPONENT ANALYSIS

Independent Component Analysis (ICA) [Comon, 1994] adalah metode statistik untuk mencari transformasi linier non singular dari data multivariate sehingga variabel-variabel yang ditransformasi se-independent mungkin. Metode ini pada awalnya untuk Blind Source Separation dimana fungsinya untuk menemukan kembali sekumpulan sumber-sumber signal yang telah dicampurkan (mixed) dan pencampurannya dinyatakan dalam serangkaian sensor. Model ICA secara klasik dinyatakan dengan

$$x = As \tag{3}$$

Dimana $S = [s_1, s_2, \dots, s_M]^T$ adalah vektor sumber yang tidak diketahui, Matrix $A \in \mathbb{R}^{M \times M}$ adalah nilai real yang tidak diketahui dan mixing matrik non-singular. Data $x = [x_1, x_2, \dots, x_M]^T$ kadang-kadang disebut dengan output-output sensor. Berikut asumsi-asumsi untuk model diperlukan :

1. Sumber-sumber sinyal (fitur) secara statistik se-independent mungkin
2. Paling banyak satu sumber memiliki distribusi gaussian
3. Mixing Matrix A dapat diinverskan

Tugas dari ICA adalah menemukan kembali signal aslinya dari pengamatan-pengamatan x tanpa mengetahui A dan atau s . Sehingga,

$$y = Wx = WAs \tag{4}$$

Dimana $W \in \mathbb{R}^{M \times M}$ adalah matrik pemisah, $y = [y_1, y_2, \dots, y_M]^T$ adalah estimasi dari skala dan permutasi vektor dari s . Inti dari ICA bagaimana memaksimalkan hasil dengan sumber-sumber sinyal nongaussian.

Mengukur Nongaussian sumber sinyal.

Salahsatu parameter untuk mengukur nongaussian dari sumber sinyal adalah Negatif Entropy atau disingkat Negentropy [Hyvärinen dkk, 1997]. Pengukuran negentropy didasarkan pada Entropy sesuai dengan teori informasi. Entropy dari suatu variabel random didefinisikan sebagai derajat informasi dari variabel random yang diamati. Ukuran dari suatu variabel yang acak, variabel yang tidak dapat diprediksi dan tak terstruktur maka nilai entropinya besar. Entropy dari variabel random diskrit x dinyatakan dengan

$$H(x) = -\sum_{i=1}^n P(x = a_i) \log_2 P(x = a_i) = E_x(I(x)) \tag{5}$$

Dimana a_i adalah nilai-nilai yang mungkin dari x , $I(x)$ adalah *information content*. $P(x = a_i)$, $P(y = a_i)$ adalah fungsi densitas probabilitas. Untuk variabel acak kontinu, disebut dengan *Differential Entropy*. Untuk variabel random kontinu y , *differential entropy* ditulis dengan :

$$H(y) = -\int f(y) \log f(y) dy \quad (6)$$

dimana $f(y)$ adalah fungsi padat probabilitas dari y . Salah satu aturan dasar dari teori informasi bahwa variabel gaussian memiliki entropy yang sangat besar diantara variabel-variabel random dengan variance yang sama. Artinya bahwa distribusi Gaussian adalah sangat random dan tidak teratur. Oleh karena itu entropy dapat digunakan untuk mengukur nongausianitas. Untuk variabel Gaussian maka nilai entropy sama dengan nol dan untuk variabel nongaussian maka nilai entropynya adalah non negatif. Negentropy dari variabel y dinyatakan dengan :

$$J(y) = H(y_{\text{gauss}}) - H(y) \quad (7)$$

dimana y_{gauss} variabel Gaussian dengan matrik kovariannya sama dengan y . Negentropy sangat baik untuk mengukur nongausianitas, tetapi sulit dalam masalah perhitungannya. Untuk mengestimasi negentropy perlu mengestimasi fungsi padat probabilitas, sehingga diperlukan penyederhanaan pendekatan untuk menghitung negentropy. Perhitungan sederhana untuk entropy dari suatu variabel random (mean nol dan variannya satu) adalah :

$$J(y) = \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} kurt(y^2) \quad (8)$$

Pendekatan lain untuk menghitung entropy berdasarkan pada prinsip maximum-entropy, disini pendekatan tingkat tinggi diganti dengan *expectation* dari fungsi non quadratic atau moment non-polynomial. Fungsi polynomial y^3 dan y^2 dapat diganti dengan fungsi lain G Metode ini menyederhanakan pendekatan negentropy didasarkan pada nilai harapan $E\{G\}$, sehingga pendekatan baru menjadi

$$J(y) \approx [E\{G(y)\} - E\{G(v)\}]^2 \quad (9)$$

untuk suatu fungsi non kuadrat G , dimana v adalah variabel Gaussian dengan mean nol dan variansinya sama dengan satu. Pemilihan fungsi G , dipastikan bahwa estimasi dari $\{G(y)\}$ tidak sulit untuk diimplementasikan dan juga tidak sensitif terhadap outlier. Persamaan 9 dikenal sebagai fungsi kontras secara umum yang dikenal sebagai bentuk umum dari kurtosis.

Algoritma FastICA

Hyvarinen menetapkan algoritme untuk menyelesaikan masalah Independent Component Analysis yang disebut sebagai fixed-point maximum negentropy disebut FastICA, [Hyvarinen, 1999] dan dinyatakan sebagai berikut :

1. Mulai dengan inisialisasi secara random w yang memiliki $\|w\|=1$
2. Update $w^+ \leftarrow E\{xg(w^T x)\} - E\{g'(w^T x)\}w$
3. Normalisasi $w^+ \leftarrow w^+ / \|w^+\|$
4. Ulangi langkah 3, sampai konvergen.

Dengan g adalah fungsi score dan g' adalah turunan dari fungsi score.

Statistik tingkat tinggi dengan Pearson System

Pearson System adalah kelompok distribusi parametrik yang memungkinkan digunakan untuk memodelkan berbagai macam distribusi dari sumber-sumber bunyi. Pearson system sangat penting dalam statistik dan banyak dipelajari [Andreev dkk 2005] untuk penelitian masalah pemilihan distribusi data. Pearson system didefinisikan dengan persamaan differential

$$f'(x) = \frac{(x-a)f(x)}{b_0 + b_1x + b_2x^2} \quad (10)$$

dimana $a, b_0, b_1,$ dan b_2 adalah parameter-parameter distribusi. Dalam pendekatan Maximum Likelihood untuk ICA hipotesa distribusi sumber-sumber bunyi digunakan sebagai fungsi contrast. Fungsi score dari sistem Pearson adalah

$$\varphi(x) = -\frac{f'(x)}{f(x)} = -\frac{(x-a)}{b_0 + b_1x + b_2x^2} \quad (11)$$

karena sederhananya fungsi score ini, maka system Pearson muncul dalam ICA. Untuk turunan dari fungsi score pada system Pearson diperoleh

$$\varphi'(x) = \frac{b_0 + ab_1 + 2ab_2x - b^2x}{(b_0 + b_1x + b_2x^2)^2} \quad (12)$$

Parameter a, b₀, b₁, b₂ dapat diestimasi dengan metode moment.

$$b_1 = a = -\frac{\mu_3(\mu_4 + 3\mu_2^2)}{C} \quad (13)$$

$$b_0 = -\frac{\mu_2(4\mu_2\mu_4 - 3\mu_3^2)}{C} \quad (14)$$

$$b_2 = -\frac{(2\mu_2\mu_4 - 3\mu_3^2 - 6\mu_2^3)}{C} \quad (15)$$

dimana

$$C = 10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3 \quad (16)$$

Pada metode moment, *moment* dapat diestimasi dengan menggunakan *moment sample* yang dihitung dari data

$$\hat{\alpha}_1 = \bar{x} = \sum_{i=1}^n x_i / n \quad (17)$$

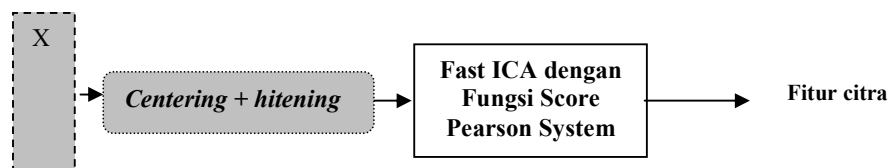
$$\hat{\alpha}_2 = \hat{\sigma}^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / n \quad (18)$$

$$\hat{\alpha}_3 = \sum_{i=1}^n (x_i - \bar{x})^3 / (n\sigma^3)$$

$$\hat{\alpha}_4 = \sum_{i=1}^n (x_i - \bar{x})^4 / (n\sigma^4) \quad (19)$$

5. DESAIN SISTEM

Ekstraksi fitur citra yang akan dilakukan pada penelitian ini, dapat dijelaskan dengan menggunakan model Gambar 3 sebagai berikut :

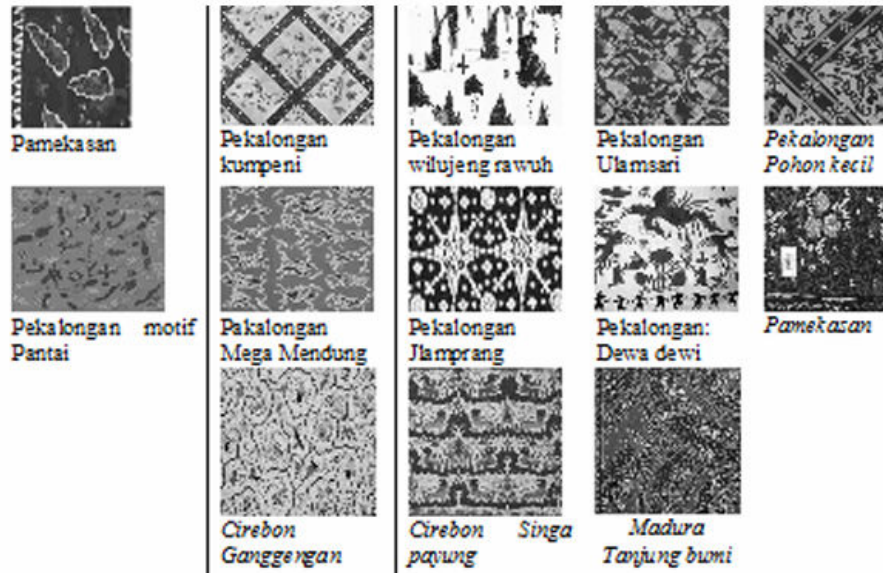


Gambar 3 Ekstraksi fitur-fitur citra batik

Secara garis besar model ekstraksi fitur diatas adalah data x merupakan potongan potongan citra batik, dimana didalamnya mengandung fitur –fitur citra. Selanjutnya dilakukan proses centering dan whitening agar didapatkan data fitur yang tidak saling berkorelasi. Kemudian proses tersebut dilanjutkan dengan pemisahan fitur-fitur citra dengan menggunakan fungsi score Pearson system pada algoritma FastICA-nya.

5. EKTRAKSI FITUR PADA CITRA BATIK

Citra batik yang digunakan adalah citra batik yang ditemui pada pesisir. Untuk kasus disini, ragam batik yang diambil adalah batik pesisir dari madura, Cirebon, dan pekalongan. Pada penelitian ini diambil 13 ragam motif yang akan akan diekstraksi fiturnya.



Gambar 4. Citra batik pesisir

Potongan-potongan citra batik

Proses pertama kali yang dilakukan dalam melakukan ekstraksi fitur adalah membentuk potongan-potongan citra kecil ukuran 8 x 8 yang diperoleh sekumpulan 13 citra citra batik pesisir. Citra dipotong sesuai dengan ukuran diatas secara acak. Potongan potongan citra dianggap sebagai sample dari data citra batik yang dinyakan sebagai x . Sample data pada penelitian ini diambil sebanyak 1000 potongan citra.

Melakukan preprosesing data

Langkah awal ICA meliputi centering dan whitening pada potongan-potongan citra X . Proses pemutihan digunakan untuk menghilangkan korelasi-korelasi antar fitur pada data citra. Dari hasil centering didapatkan fitur yang saling tidak berkorelasi.

Ekstraksi fitur citra batik

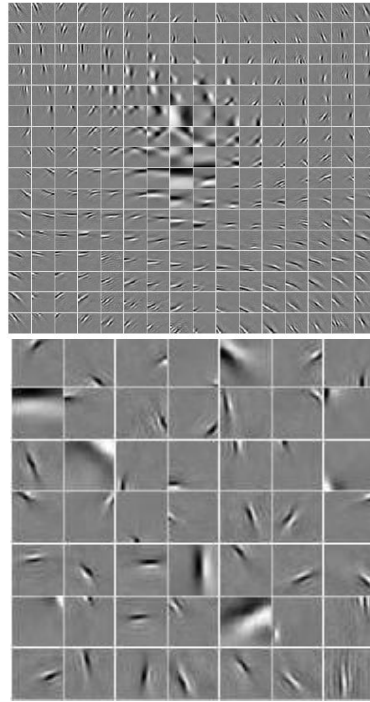
Setelah data menjadi tidak berkorelasi, langkah selanjutnya adalah proses estimasi komponen-komponen independen atau fitur-fitur citra batik. Ekstraksi fitur citra didasarkan pada $y = Wx$, dengan aturan pembelajaran menggunakan FastICA yaitu dengan aturan perubahan w didasarkan pada

$$w^{k+1} \leftarrow E\{zg(w^T z)\} - E\{g'(w^T z)\}w^k \quad (20)$$

dengan $g(.) = \varphi(y) = -\frac{f'(y)}{f(y)} = -\frac{(y-a)}{b_0 + b_1 y + b_2 y^2}$ adalah fungsi score dan $g'(.)$ adalah turunan dari fungsi

score. Proses iterasi pada algoritma FastICA akan berhenti jika $\delta = (|W^{k+1} - W^k|^2)$ menjadi sangat kecil atau mencapai titik konvergen. δ adalah batasan untuk error toleransi yang ditentukan 0.001

Fitur yang dihasilkan dari citra berdasarkan ragam hias batik pesisir dapat dilihat dari gambar 5



Gambar 5. Fitur Citra batik yang dibangkitkan

7. KESIMPULAN.

Pada kesimpulan ini bahwa fitur-fitur dari citra batik dapat dibentuk dengan cara saling bebas antar motif, dan tidak nampak seperti citra batik. Ini hanya merupakan komponen dasar dari citra batik itu sendiri. Jika komponen-komponen ini dibangun maka akan terbentuk motif-motif baru dari batik

6. DAFTAR PUSTAKA

- Hamzuri, 1985, *Batik Klasik*, Djambatan, Jakarta
- Djoemena, Nian S. 1986. *Ungkapan Sehelai Batik*. Jakarta: Djambatan
- Hyvärinen, A., Hurri J., Patrik O. Hoyer. *Natural Image Statistics: A probabilistic approach to early computational vision*, Springer, 2009
- Comon, P, *Independent component analysis, a new concept?*, (1994), " *Signal Processing*, Vol. 36, No.3, hal 287-314
- Hyvärinen, A and E. Oja, 1997, *A fast fixed-point algorithm for independent component analysis*. *Neural Computation*, hal 1483–1492
- Andreev, A., Kanto, A. dan Malo, P., 2005, *Simple Approach for Distribution Selection in The Pearson System*, Helsinki School of Economics, Helsinki
- Hoyer, P.O. dan Hyvärinen, A., 2000, *Independent Component Analysis Applied to Feature Extraction from Colour and Stereo Images*, *Network: Computation in Neural Systems*, Vol. 11, hal 191-210