

SISTEM PENGENALAN KAIN ENDEK KHAS BALI BERDASARKAN FITUR TEKSTUR

I Gusti Agung Gede Arya Kadyanan¹

Program Studi Teknik Informatika, Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan
Alam, Universitas Udayana
Email : gungde@unud.ac.id

ABSTRAK

Penelitian ini merupakan lanjutan penelitian pada periode sebelumnya yang telah berhasil melakukan rekonstruksi citra kain Endek berdasarkan fitur tekstur. Dalam penelitian ini telah berhasil dibangun sistem pengenalan citra kain Endek khas Bali. Dalam upaya membangun suatu platform pengembangan sistem perolehan citra warisan budaya seperti citra batik dan lukisan, diperlukan suatu pendekatan yang handal. Dengan memanfaatkan Keyblock yang merupakan generalisasi sistem temu kembali berbasis konten pada domain citra. Codebook dibangun berdasarkan pengelompokan nilai keyblock yang mewakili karakteristik citra pada data, kemudian digunakan untuk proses encoding dan decoding. Proses encoding merupakan upaya untuk merepresentasikan setiap sub area pada citra kain Endek dengan suatu nilai indeks yang sesuai dengan codebook. Citra hasil encoding berubah menjadi bentuk matrik 1-dimensi, bentuk ini merupakan analogi dari keyword pada sistem pencarian informasi berbasis teks. Sebanyak 60 citra kain Endek dilibatkan dalam penelitian ini, dengan 40 citra sebagai citra pelatihan dan 20 citra sebagai citra uji.

Kata kunci: *keyblock, vektor quantization, codebook, Generalized Lloyd Algorithm, kain Endek*

ABSTRACT

This study is a continuation of research in the previous period which have successfully reconstructed the Endek image based texture features. In this research has been successfully constructed Balinese Endek image recognition system. In an effort to build an image acquisition system development platform of cultural heritage such as batik and painting image, we need a reliable approach. By utilizing Keyblock which is a generalization of content-based retrieval system in the image domain. Codebook is constructed by grouping keyblock value that represents the characteristics of the image data, then used for encoding and decoding process. The encoding process is an attempt to represent each sub area in the Endek image with an index value corresponding to the codebook. Image encoding results turned into a one-dimensional matrix form, this form is an analogy of the keywords in the text-based information retrieval system. A total of 60 images Endek cloth included in this experiment, with 40 image as a training image and 20 images as query images.

Keyword: *keyblock, vektor quantization, codebook, Generalized Lloyd Algorithm, Endek*

1. PENDAHULUAN

Sistem perolehan teks dan perolehan citra sama-sama merupakan suatu sistem perolehan informasi, karena itu teori dan teknik pada perolehan teks yang lebih dulu berkembang berpotensi juga untuk digunakan pada perolehan citra. Namun generalisasi perolehan informasi dari domain teks ke domain citra tidaklah mudah. Salah satu hambatan besar adalah perbedaan karakteristik intrinsik antara teks dan citra dalam merepresentasikan dan mengekspresikan suatu informasi.

Dalam penyajian informasi secara sintaks, dokumen teks bersifat 1-dimensi, sedangkan citra merupakan bentuk 2-dimensi. Dalam ekspresi informasi secara semantik, satuan informasi dalam bentuk kata dari dokumen teks, terutama kata kunci, memiliki semantik yang terkait dengan dokumen teks. Sebaliknya, satuan informasi dalam bentuk piksel maupun segmen dari citra, tidak mengandung semantik yang secara langsung dengan citra tersebut. Maka pertanyaan besarnya adalah bagaimana membangun sistem pengenalan citra kain Endek yang serupa dengan kata kunci pada dokumen berbasis teks.



Gambar 1. Tipikal 6 dari 60 citra kain Endek yang digunakan dalam penelitian.

2. TINJAUAN PUSTAKA

2.1. Fitur Tekstur

Tekstur merupakan properti yang dimiliki semua citra. Tekstur berisi informasi penting tentang susunan struktur dari suatu permukaan dan hubungannya terhadap lingkungan sekitarnya. Adapun teknik penyimpanan citra digital dalam komputer sebagai array 2-dimensi. Jika $L_x = \{1, 2, \dots, N_x\}$ dan $L_y = \{1, 2, \dots, N_y\}$ adalah data spasial, maka $L_x \times L_y$ adalah himpunan sel-sel resolusi spasial. Citra digital I adalah sebuah fungsi yang memetakan nilai keabuan (*gray tone*) $G \in \{1, 2, \dots, N_g\}$ ke setiap sel resolusi; $I : L_x \times L_y \rightarrow G$. Konsep keabuan berdasar pada variasi kegelapan (*shades*) warna keabuan sel-sel resolusi pada sebuah citra, sedangkan tekstur menekankan pada distribusi spasial (statistik) dari nilai keabuan.

2.2. Pengertian *Keyblock*

Terinspirasi dengan sistem temu kembali informasi berbasis teks, pengusul konsep *keyblock* Shu dan Zhang (2002) mencoba untuk menggunakan teori-teori yang dipakai pada temu kembali teks ke dalam domain citra dengan menempatkan *codebook* sebagai *dictionary* dan *keyblock* sebagai *keyword*-nya.

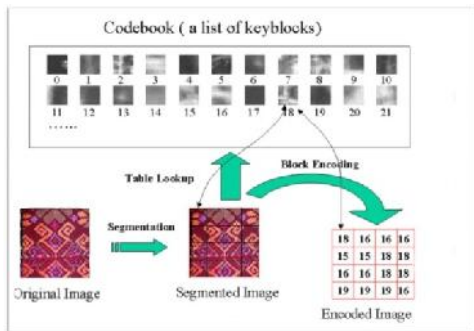
2.3. *Pairwise Nearest Neighbor Algorithm* (PNNA)

Pairwise Nearest Neighbor Algorithm (PNNA) (Eakins dan Graham, 1999) adalah

algoritma pengklasteran sederhana yang dimulai dengan sebuah himpunan pelatihan. Pada setiap iterasi, dua vektor terdekat disatukan dan digantikan dengan *centroidnya* sehingga jumlah data berkurang satu. Proses ini terus berulang sampai ukuran *codebook* yang diinginkan tercapai.

2.4. Encoding Citra Satelit

Pada gambar 2 berikut dapat dilihat skema prosedur *Encoding* citra kain Endek.



Gambar 2. Prosedur *Encoding* (dimodifikasi dari Zhu, Rao, dan Zhang (2000))

2.5. Root Mean Square error (RMS_error)

Untuk melihat kualitas *codebook* yang dihasilkan, maka dilakukan proses *decoding*, yaitu merekonstruksi seluruh citra pelatihan hasil proses *encoding*. Selanjutnya, dihitung *RMS_error* intensitas keabuan antara citra asli dan citra hasil rekonstruksi dengan rumus berikut ini.

$$MSE = \frac{1}{N_1 N_2} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} (x[n_1, n_2] - \hat{x}[n_1, n_2])^2 \quad (1)$$

Dimana,

N_1 : ukuran baris

N_2 : ukuran kolom

n_1 : koordinat baris

n_2 : koordinat kolom

x : nilai keabuan piksel citra asli

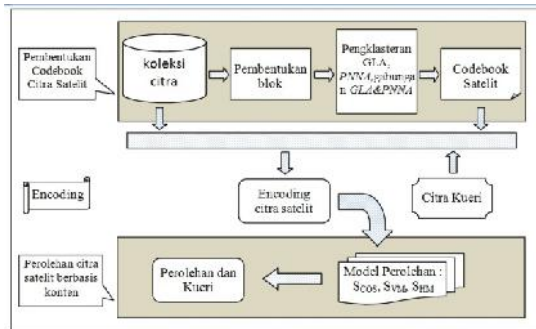
\hat{x} : nilai keabuan piksel citra rekonstruksi

2.5. Model Perolehan Citra Kain Endek

Model yang digunakan dalam perolehan citra *kain Endek* dengan teknik perolehan citra yaitu *quintuple* [D, K, W, Q, S] dimana:

- $D = \{d_1, \dots, d_j, \dots, d_n\}$ adalah daftar citra *kain Endek* terkompresi VQ, dan n disini merupakan jumlah citra *kain Endek* di basis data.
- $K = \{k_1, \dots, k_i, \dots, k_t\}$ adalah daftar *keyblock* pada *codebook*, dan t merupakan jumlah *keyblock* di *codebook*.
- $W : K \times D \rightarrow R^+$ adalah pemetaan yang memetakan pasangan *keyblock* dan citra pada sebuah angka positif. Bila $w_{i,j} = W(k_i, d_j)$, maka W direpresentasikan sebagai matrik $(w_{i,j})_{t \times n}$ disebut bobot matrik yang tiap elemennya adalah indeks *keyblock* pada sebuah citra dimana t adalah jumlah *keyblock* di *codebook* dan n adalah jumlah citra di basis data . Setiap citra d_j , memiliki sejumlah vektor ciri $d_j = \{w_{1,j}, \dots, w_{t,j}\}$.
- $Q = \{q_1, \dots, q_c, \dots, q_l\}$ adalah himpunan citra kueri. Setiap citra kueri memiliki ciri vektor $q = \{w_{1,q}, \dots, w_{t,q}\}$ yang sama dengan vektor ciri d_j .

- $S : Q \times D \rightarrow R^+$ ukuran kesamaan antara kueri dan citra. Nilai ini digunakan untuk membuat rangking citra yang diperoleh.



Gambar 3. Prosedur Perolehan Citra Kain Endek

2.6.1. Kemiripan Kosinus

Kemiripan kosinus adalah pengukuran kemiripan antara dua vektor dengan mencari nilai kosinus dari sudut antara kedua vektor tersebut. Pada perolehan citra satelit, kemiripan kosinus antar dua citra satelit berkisar antara 0 – 1 dan sudut antara kedua vektor frekuensi tidak lebih dari 90°. Pada dua buah vektor d_j dan q yang merupakan frekuensi kemunculan setiap *keyblock*, maka kesamaan kosinusnya adalah:

$$s_{cos}(q, d_j) = \cos \theta = \frac{d_j \cdot q}{|d_j| |q|} = \frac{\sum_{i=1}^t w_{i,j} * w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} * \sqrt{\sum_{i=1}^t w_{i,q}^2}} \quad (2)$$

2.6.2. Kemiripan Histogram

Perolehan citra dengan model ini menggunakan histogram dari frekuensi kemunculan seluruh *keyblock* pada suatu citra [15], dimana $w_{i,j} = f_{i,j}$, frekuensi k_i muncul pada d_j . Begitu juga $w_{i,q} = f_{i,q}$. Nilai kesamaan dihitung dengan:

$$s_{HM}(q, d_j) = \frac{1}{1 + dis(q, d_j)} \quad (3)$$

2.6.3. Kemiripan Ruang Vektor (Vector Space Model)

Metode perolehan informasi berbasis vektor merepresentasikan dokumen dan kueri dengan vektor dimensi tinggi, serta menghitung kemiripannya dengan *inner product*. Dengan menormalisasi vektor menjadi satuan panjang, *inner product* mengukur kosinus sudut antara dua vektor pada ruang vektor.

Bila ada n citra pada D dan t *keyblock* pada K , maka normalisasi kemunculan *keyblock* k_i pada citra d_j adalah:

$$s_{VM}(q, d_j) = \frac{d_j \cdot q}{|d_j| |q|} = \frac{\sum_{i=1}^t w_{i,j} * w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} * \sqrt{\sum_{i=1}^t w_{i,q}^2}} \quad (4)$$

2.6. Pengujian Precision dan Recall

Efektifitas sebuah sistem perolehan citra berbasis konten biasanya diekspresikan dengan istilah *precision* dan *recall*. Berdasarkan teori perolehan informasi, *precision* didefinisikan sebagai jumlah dokumen relevan yang diperoleh hasil suatu kueri dibagi seluruh dokumen yang diperoleh, sedangkan *recall* didefinisikan sebagai jumlah dokumen yang relevan yang diperoleh hasil suatu kueri dibagi dengan total dokumen relevan yang ada. Berikut adalah rumusnya:

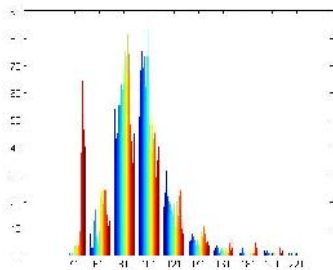
$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

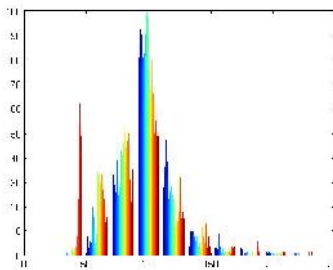
3. METODOLOGI PENELITIAN

Metodologi yang digunakan dalam pengembangan sistem rekonstruksi citra kain Endek berdasarkan fitur tekstur ini adalah studi literatur dan eksperimen. Studi literatur diambil dari jurnal penelitian dan makalah yang berkaitan dengan pendekatan *keyblock* dan sistem perolehan informasi berbasis konten. Eksperimen dilakukan dengan merealisasikan sistem perolehan informasi berbasis konten dengan pendekatan *keyblock* menggunakan MATLAB R2009a.

4. HASIL DAN PEMBAHASAN



Gambar 4. Histogram citra kain Endek asli pertama



Gambar 5. Histogram rekonstruksi citra pertama

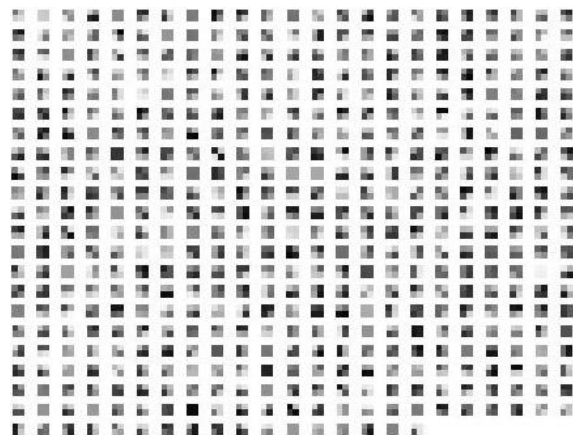
Tabel 2. Rata-rata *RMS_error* pada setiap ukuran *codebook* dan blok

Ukuran blok	Ukuran <i>codebook</i>				
	100	200	300	400	500
2 x 2	6,85	6,71	6,29	6,04	5,93
4 x 4	8,12	8,08	7,92	7,78	7,75
8 x 8	8,45	8,60	8,43	8,47	8,50

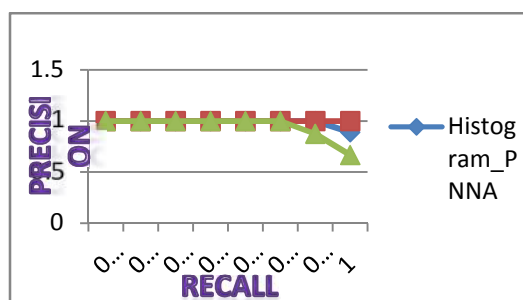
Tabel 3. Rata-rata intensitas keabuan *RMS_error* pada setiap ukuran *codebook* dan blok

Ukuran Blok	Ukuran <i>codebook</i>				
	100	200	300	400	500
2 x 2	6,85	6,71	6,29	6,04	5,93
4 x 4	8,12	8,08	7,92	7,78	7,75
8 x 8	8,45	8,60	8,43	8,47	8,50

Pada Tabel 2 diatas menunjukkan nilai rata-rata *RMS_error* untuk seluruh citra *kain Endek* pada setiap ukuran *codebook* dan blok yang dibentuk. Sedangkan pada Tabel 3. Dapat diketahui rata-rata intensitas keabuan untuk seluruh citra *kain Endek* pada setiap ukuran *codebook* dan blok yang dibentuk.



Gambar 6. Representasi *codebook* dengan blok 2x2 dan jumlah *codebook* 500.



Gambar 7. Hasil perolehan citra kain Endek untuk seluruh data uji

5. KESIMPULAN

Hasil studi ini memberikan kesimpulan sebagai berikut:

- 1) Ciri tekstur citra *kain Endek* dapat diekstraksi dengan pendekatan *codebook* dan *keyblok*.
- 2) Berdasarkan keberhasilan didalam sistem temu kembali berbasis text, dengan ini sistem temu kembali informasi berdasarkan konten pada domain citra *kain Endek* dapat dilakukan dengan merepresentasi citra *kain Endek* diubah menjadi vektor ciri 1-dimensi.
- 3) Akurasi terbaik diperoleh pada ukuran blok 2x2 dan *codebook* 500, dengan melihat tampilan citra *kain Endek* hasil rekonstruksi secara *visual* dan diperkuat dengan menghitung nilai rata-rata *RMS_error* dapat disimpulkan bahwa kerusakan citra *kain Endek* ketika rekonstruksi tidak terlalu signifikan karena tingkat keabuan dan nilai rata-rata *RMS_error* yaitu 5.93 tingkat keabuan masih dianggap dalam batas toleransi.

6. DAFTAR PUSTAKA

- J. Eakins dan M. Graham, "Content-based Image Retrieval," *Technology Applications Programme Report 39*, University of Northumbria at Newcastle, October 1999.
<http://en.wikipedia.org/wiki/CBIR>, Content-based image retrieval" [accessed in October 1, 2009].
- R. M. Haralick, K. Shanmugam, dan I. Dinstein, "Textural Feature for Image Classification," *IEEE Transactions on Sistem, Man dan Cybernetics*, Vol SMC-3 No. 6, November 1973.
- P. Brodatz, "Textures", New York, Dover, 1966.
- L. Zhu dan A. Zhang, "Theory of Keyblok-based Image Retrieval," *ACM Journal*, pp. 1-32, March 2002.
- L. Zhu, A. Rao, dan A. Zhang, "Advanced Feature Extraction for Keyblok-based Image Retrieval," *Proceedings of International Workshop on Multimedia Information Retrieval (MIR 2000)*, Los Angeles, California, USA, November 4, 2000.
- L. Zhu, A. Rao, dan A. Zhang, 2000, "Keyblok: An approach for content-based geographic image retrieval," *Proceedings of First International Conference on Geographic Information Science (GIScience2000)*, Savannah, Georgia, USA, 286-287, 2000.
- L. Zhu, C. Tang, dan A. Zhang, "Using Key blok Statistics to Model Image Retrieval," *Advances in Multimedia Information Processing – PCM 2001, Second IEEE Pacific Rim Conference on Multimedia*, Beijing, China, October 24-26, 2001, *Proceedings 2001*.
- Yates, R.B, dan B.R. Neto. *Modern Information Retrieval*. Addison Wesley, 1999.