PENCITRAAAN DOKUMEN BERBASIS CITRA INER MENGGUNAKAN TEKNIK KOMPRESI PARTISI REKTANGULAR

Oleh : Acep Lili Khoerudin

Abtraksi

Pencitraan dokumen merupakan replikasi digital dari dokumen tradisional berbasis kertas. Migrasi dokumen dari dokumen tradisional ke dokumen elektronik akan memudahkan proses manajemen dokumen. Fungsi-fungsi pencitraan dokumen yang mengacu pada manajemen dokumen menyediakan fasilitas untuk melakukan pencitraan dokumen, yaitu pemindaian (scanning), pengindeksan (indexing), pengarsipan (stroring), pengambilan (retrieving) dan pendistribusian (distribution).

Pendahuluan

Sejak kemunculan *word processor* dan penggunaan CAD (*Computer Aided Designed*) dalam desain, penggunaan dokumen elektronik (dokumen yang bersumber dari komputer) menjadi dikenal luas. Penggunaan dokumen elektronik lebih menguntungkan dibandingkan dengan dokumen tradisional. Keuntungan ini dijadikan alasan untuk melakukan migrasi dokumen dari dokumen tradisional berbasis kertas ke dokumen elektronik.

Manajemen dokumen elektronik mengacu pada pengelolaan dokumen kertas dengan perekaman melalui proses pemindain (*scanning*), pengindeksan, pengarsipan, pengambilan, dan pendistribusian dokumen secara elektronik. Pencitraan dokumen adalah replika digital dari dokumen kertas yang dapat memudahkan akses dan manajemen dokumen.

Pencitraan dokumen (*document imaging*) akan menghasilkan citra digital yang membutuhkan kapasitas penyimpanan data yang relatif besar, sehingga untuk menyimpannya diperlukan teknik penyimpanan yang efisien. Pada umumnya isi dokumen kertas terdiri dari teks, gambar dan grafik hitam putih, apabila dokumen ini mengalami proses pemindaian maka akan menghasilkan citra hitam putih (*Binary Image / Bi-Level Image*). Citra seperti ini dapat disimpan secara efisien dengan menggunakan teknik kompresi citra biner.

Berlandaskan pemikiran yang tercantum pada latar belakang tersebut diatas muncul masalah, bagaimana membangun aplikasi perangkat lunak pencitraan dokumen dan *file* penyimpanan pencitraan suatu dokumen sekecilnya dengan teknik kompresi partisi rektangular serta mekanisme pencarian suatu bagian dokumen (citra dari dokumen-citra).

Permasalahan yanga akan dibahas terbatas pada pencitraan dokumen menggunakan teknik kompresi partisi rektangular. Masukan sistem pencitraan dokumen berformat bitmap dengan 8 bit kedalaman piksel. Untuk mendapatkan suatu dokumen citra dilakukan pencarian denga metoda sekuensial berdasarkan struktur file dokumen-citra dan kata kunci yang merupakan key word.

Pencitraan dokumen

Proses dalam pencitraan dokumen meliputi [1] :

1. Akuisisi Input
   Akuisisi input dari dokumen kertas menjadi bentuk digital dilakukan melalui proses digitasi dengan perangkat *scanner*. Proses digitasi citra dapat dijelaskan melalui gambar 2.2 :
   a. Cahaya yang dipantulkan dari objek melewati sistem optik dan membentuk distribusi energi radian citra L.
   b. Intensitas energi cahaya dialihragamkan menjadi arus listrik V oleh sensor.
   c. Arus listrik tresbut didigitalisasi oleh *analog-to-digital converter*.
2. Pemrosesan (Opsional)
   Pemrosesan dalam apliksi pencitraan dokumen meliputi :
   a. Peningkatan kualitas citra
      Peningkatan kualitas citra dilakukan dengan tujuan untuk mengoreksi defisiensi yeng terdapat pada dokumen asal dan dokumen hasil digitasi sehingga dapat dihasilkan citra yang lebih layak dibaca / diinterprestasi. Degradasi

kualitas disebabakan oleh *noise* dan *distorsi* non linier akibat kualitas dokumen asal yang rendah (kurang kontras) pengaturan kontras dan pencahayaan yang tidak optimal dan kesalahan kuantisasi pada proses digitasi. Selain proses digitasi, tidak lagi terdapat degradasi pada citra.

b. Indeks dokumen
Pengindeksan merupakan pengategorian dokumen kedalam beberapa kriteria. Pengindeksan dokumen bertujuan untuk memudahkan penelusuran atas dokumen dalam aplikasi pencitraan dokumen. Metode pengindeksan yang dipergunakan dalam pencitraan dokumen [10] :

1. Pengindeksan sequensial berdasarkan tanggal pemindaian.
2. Nested folder
3. Database entry / Record Key / Unique Identifier.
4. Kata / Subjek kumci
5. Full Text / Pencarian Fuzzy
6. Struktur dokumen
7. Kartu Katalog
8. Hyperlinks
9. Kombinasi dari metode yang ada.

3. Kompresi
Kompresi citra biner dalam aplikasi pencitraan dokumen didasarkan pada fakta bahwa sebagian citra terdiri dari piksel putih dan hitam. Piksel hitam muncul sebagai representasi dari karakter, simbol atau grafik. Secara umum kompresi citra biner terbagi menjadi tiga kategori, yaitu:
   1. Pengkodean transisi antara hitam dan putih
   2. *White Skipping*
   3. Pengenalan pola

Teknik Kompresi partisi rektangular

Teknik kompresi partisi rektangular pada dasarnya adalah teknik dua dimenasi yang diimplementasikan dalam teknik satu dimenasi sederhana. Teknik kompresi ini terdiri dari dari dua tahap yang independen (lihat gambar 2.4), yaitu :

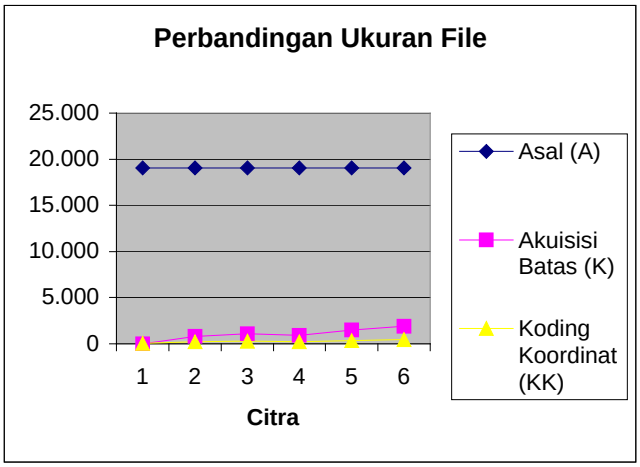
1. Akuisisi informasi batas rektangular, dimana piksel batas (boundary pixsel) yang membentuk rektangular pada arah x dan y didefinisikan dan kemudian dieliminasi sehingga diperoleh titik sudut dari masing-masing rektangular.
2. Kompresi koordinat, dengan tujuan agar efisiensi kompresi dapat ditingkatkan lebih lanjut dengan mengurangi redundansi atau korelasi antar koordinat dari piksel batas yang diperoleh.

Data Pengukuran Pencitraan Dokumen

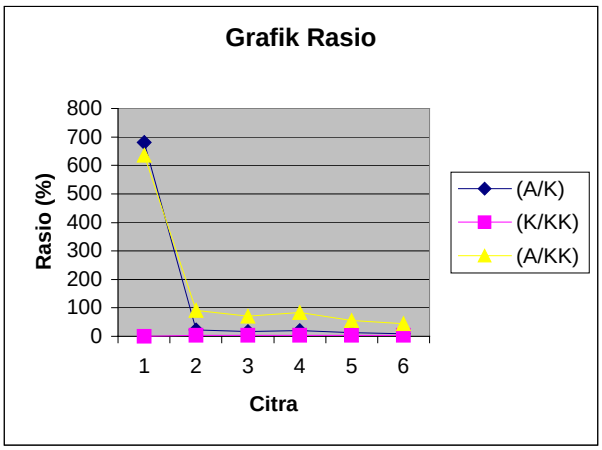
Data yang terdapat pada tabel 1 dan tabel 3 merupakan data hasil pengukuran terhadap parameter-parameter yang terdapat pada teknik kompresi partisi rektangular pertahapan proses dengan alat bantu menggunakan software pencitraan dokumen yang telah dibangun. Data yang tedapat pada tabel 2 merupakan data yang diperoleh dari tabel 1. data tabel 4 merupakan data waktu pencarian hasil pengukuran pada suau dokumen citra. Gambar 1 merupakan grafik perbandingan ukurab file citra asal dan citra hasil proses akuisisi batas yang datanya diambil dari tabel 1. gambar 2 meruapakn grafik rasio kompresi yang datanya bersumber dari tabel 2. gambar 3 meruapakn grafik hubungan rasio kompresi dan jumlah titik batas yang datanya diambil dari tabel 1 dan tabel 2.

Tabel 1. Ukuran Citra dan Hasil Kompresi

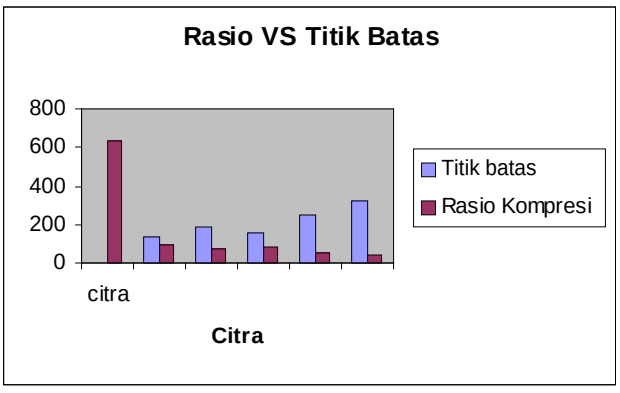
Citra #	Objek	Titik Batas	Ukuran Citra (byte)		
			Asal (A)	Akuisisi Batas (K)	Koding Koordinat (KK)
1	Kotak	4	19.078	28	30
2	Segi Tiga	138	19.078	832	210
3	Oval	184	19.078	1.108	266
4	if4-1	153	19.078	922	228
5	if4-2	247	19.078	1.486	340
6	if4-3	319	19.078	1.918	434
Rata-rata		174,17	19.078	1.049,00	251,33



Gambar 1. Grafik Perbadingan Ukuran File



Gambar 2. Grafik Rasio Kompresi



Gambar 3. Grafik Hubungan Rasio kompresi dan Jumlah Titik Batas

Tabel 2. Rasio Kompresi

Citra #	Rasio		
	(A/K)	(K/KK)	(A/KK)
1	681,36	0,93	635,93
2	22,93	3,96	90,85
3	17,22	4,17	71,72
4	20,69	4,04	83,68
5	12,84	4,37	56,11
6	9,95	4,42	43,96
Rata-rata	127,50	3,65	163,71

Tabel 3. Perolehan Titik Batas Pada Proses Akuisisi Batas

Citra #	Citra			Reduksi Elemen (%)		Rasio
	Asal	Reduksi (V)	Reduksi (H)	Asal-V	Asal-H	
1	12.561	318	4	122,43	125,57	0,000318
2	5.104	276	138	48,28	49,66	0,027038
3	9.852	318	184	95,34	96,68	0,018676
4	2.785	354	153	24,31	26,32	0,054937
5	2.785	447	247	23,38	25,38	0,088689
6	2.785	447	319	23,38	24,66	0,114542

Tabel 4. Data Hasil Pencarian Segmen Dokumen

Hal (n)	Waktu (millisecond) (t_0)	Ukuran Citra (byte) (s_0)	s_n+s_{n-1} (s_1)	t_n-t_{n-1} (t_2)
1	0	3.788	3.788	0
2	170	3.962	7.750	170
3	390	3.594	11.344	220
4	550	4.152	15.496	160
5	769	21.764	37.260	219
6	1.759	19.804	57.064	990
7	2.689	30	57.094	930
8	2.689	230	57.324	0
9	2.689	342	57.666	0
10	2.750	436	58.102	61
11	2.750	32	58.134	0
12	2.750	560	58.694	0
13	2.800	32	58.726	50
14	2.809	268	58.994	9
15	2.960	210	59.204	151
Rata-rata		3.946,93		197,33

Kesimpulan

Hasil analisis performansi teknik kompresi partisi rektangular dapat disimpulkan sebagai berikut :

1. Rasio kompresi citra teknik kompresi partisi rektangular dipengaruhi oleh jumlah titik batas yang diperoleh pada tahap akuisisi batas.
2. Rasio kompresi teknik kompresi partisi rektangular yang tingggi diperoleh untuk citra yang didominasi garis lurus (vertikal dan horizontal) dan daerah hitam yang luas.
3. Rasio kompresi teknik kompresi partisi rektangular yang rendah diperoleh untuk citra yang didominasi karakter.

Hasil analisis waktu pencarian secara sequencial pada pencitraan dokumen dapat disimpulkan sebagai berikut :

1. Waktu pencarian secara sequencial tergantung pada akumulatif ukuran data sebelumnya, dan posisi relatif data dari data pertama. Semakin besar ukuran kapasitas data dan posisinya semakin jauh maka waktu pencarian yang diperlukan semakin lama.
2. Peningkatan 20 byte ukuran data akan menambah waktu pencarian 1 millisecond pada pencitraan dokumen ini.

Struktur file dokumen-citra pada pencitraan dokumen ini mempengaruhi ukuran file citra dokumen, yang tentunya berpengaruh terhadap lamanya waktu pencarian..