論 文

論文 80-17-2-3

한글 Graphic Image Data의 統計的 特性에 관한 硏究
(A Study on the Statistical Characteristics of Hangul Graphic Image Data)

金 在 錫*, 金 在 均**
(Kim, Jae-Seok and Kim, Jae-Kyoon)

要 約

Graphic image data의 效率的인 coding을 위해서, 한글 image와 英文 image의 統計的인 特性들이 測定 比較되었다. 또한 Markov model 에 의한 run length의 確率分布와 測定된 run length의 確率分布가 比較 檢討되었다.

測定된 run length 分布는 negative-power 分布에 近似하며, 이것은 한글 image 에서 더욱 뚜렷한 것으로 나타났다.

代表的 네가지 run length code 의 coding 特性이 比較되었는데, 英文 image 보다 한글 image 의 coding 特性이 더욱 우수한 것으로 밝혀졌다.

Abstract

For efficient coding of graphic image data, the statistical characteristics for both Korean lettered images and English lettered images are measured and compared.

Also, the measured run length distribution is compared with the run length distribution based on Markov model.

It is shown that the measured white run length distribution is more like a negative-power distribution than an exponential distribution. This fact is stronger in the Korean lettered images than is the English lettered images. The performances of four typical run length codes are compared for the same set of graphic data files, and it is shown that the codes perform better in the Korean lettered images than in English lettered images.

1. 序 論

Facsimile 로서 傳送되는 書類, 圖面등의 graphic image data는 그 data 量이 방대하여 일반적으로 傳送時間이 길다. data 量을 減縮하여 이 傳送時間을 줄이기 위해서는 image data 의 統計的 特性이 이해되어야 한다. 英文 image 의 경우에는 黑色 pel (picture element) 과 白色 pel 의 run length 確率分布등 여러 特性들이 測定되어 왔으며 이 資料에 근거하여 效率的인 coding 方法들이 제시되어 왔다.

일찌기 연속적인 image data는 first-order Markov source 로 假定되어 run length 分布가 exponential 分布를 갖는 것으로 이해되었다.[1] 그러나 英文 image에 있어서 white run length의 실제 分布는 특히 run length 가 긴 경우 exporential 分布

*正會員, 韓國電子技術硏究所
(Korea Institute of Electronics Technology)
**正會員, 韓國科學院 電氣 및 電子工學科
(Korea Advanced Institute of Science)
接受日字; 1979年 11月 29日

와는 상당한 차이가 있는 것으로 나타났다.

일반적으로 run length 分布를 알고 있을때 構成할 수 있는 가장 optimal 한 code는 Huffman code이지만 構成하기가 비교적 복잡해서, 대체로 조직적인 suboptimal code 들을 실제로 많이 사용하고 있다.[2] 그러나 한글 image에 대해서는 그 特性을 測定하지 않고 英文 image에 의해 얻어진 結果들이 그대로 이용되었다.[3]

본 硏究에서는 英文 image data와 함께 한글 image data의 여러 特性들을 測定하여 그 結果를 比較檢討함으로써, 한글 image data의 效率的인 coding 方法에 올바른 資料를 얻고자 한다. 실험에 사용할 image data source 로서는 한글 타자문 인쇄文등 12종, 英語 타자文 인쇄文 14종, 기타 圖面등 4종으로 총 30종의 data source를 選擇하였다. 이들 data source 는 CCD scanner 로서 digitize되었으며, 그 解像度는 水平方向으로 5 pel/mm, 垂直方向으로 3.3 line/mm이다.

각 data file의 run length 確率分布 white pel 의 比率, 평균 run length, pel 당 entropy 등이 測定되었고, image data에 대한 first-order Markov model의 妥當性이 檢討되었다. 또한 代表的인 4종의 code에 대한 simulation結果를 測定 比較함으로써 실제 image data에 대해 有用한 code選別의 資料를 제시하였다.

2. Markov model

어느 한 画面을 scanning할때 한 line 상에서 연속적으로 나타나는 pel 사이에는 統計的인 相關關係가 존재하는데,[2] J.Capon은 이 연속적인 image data 를 first-order Markov source로 modeling 함으로써 run length의 exponential 確率分布의 特性을 제시하였다.[1] First-order Markov model이란 한 scan line 상에서 어느 한 pel의 현 狀態는 바로 앞에서 나타난 한 pel 의 狀態에만 依存한다고 假定한 것이다.

이때 각 run length들은 서로 獨立的이라는 假定이 이에 포함되어 있다. 이 model을 假定할 경우에는 image data의 모든 統計的인 情報를 평균 run length 만의 函數로 표시할 수 있다.[4]

바로 이전의 pel이 white 일때 현재 pel이 black 일 확률을 P(blw)와 같이 표시하면, white run length R_w가 x가 될 확률 $P[R_w=x]$은 다음과 같이 쓸 수 있다.

$$P[R_w=x]=P(wlw)^{x-1}\ P(blw),\quad x\geq 1 \cdots\cdots(1)$$

그러므로 white run의 평균 run length가 transition probability의 逆數로 표시됨을 알 수 있다.[5]

$$E(R_w)=\sum_{x=1}^{\infty} x\cdot P[R_w=x]=\frac{1}{P(blw)} \cdots\cdots(2)$$

(2) 식과 P(wlw)=1−P(blw)의 관계식을 이용하면 (1) 식의 white run length 分布가 평균 run length 만의 函數로 표시될 수 있다.

$$P[R_w=x]=[1-P(blw)]^{x-1}\cdot P(blw)$$
$$=\left[1-\frac{1}{E(R_w)}\right]^{x-1}\cdot\frac{1}{E(R_w)} \cdots\cdots(3)$$

이 식은 run length 가 exponential 하게 감소하는 分布임을 말해 주고 있다.

마찬가지로 pel 당 entropy 값도 평균 run length 만의 函數로 표시할 수 있다.

$$h(R_w)\triangleq\frac{H(R_w)}{E(R_w)}=\frac{1}{E(R_w)}\cdot\{-\sum_{x=1}^{\infty}P(R_w=x)\cdot\log[P(R_w=x)]\}=\frac{1}{E(R_w)}\cdot\{E(R_w)\cdot\log_2 E(R_w)-[E(R_w)-1]\cdot\log_2[E(R_w)-1]\}$$
$$\cdots\cdots(4)$$

J.Capon이 이 model 을 제시한 이후 有用하게 사용되어 왔으나, 그후 實驗的으로 英文 image의 white run 은 exponential 分布* 보다는 negative power 分布** 에 가깝다는 것이 判明되었다.[2]

* $P(L)=b\cdot a^{-L}$ (단 a, b 는 상수)
** $P(L)=a\cdot\left(\frac{1}{L}\right)^{b}$ (단 a, b 는 상수)

3. 確率測定 및 考察

確率測定에는 30종의 한글, 英文, 圖面들이 사용되었으나, 다음의 代表的 7종에 대한 測定結果가 표 1로 표시되었다.

a) 英文 *typewritten text*
b) 한글 〃 〃
c) 英文 printed text
d) 한글 〃 〃
e) 英文 handwritten text
f) 한글 〃 〃
g) line drawing (圖面) 등

이들 7종의 data source 는 그림 1과 같다. 이 data source 는 facsimile system의 scanner 部分을 利用하여 digitize 되었다. image sensor 로는 최신 반도체 소자인 CCD가 사용되었는데 그 解像度는 1024

pels/line 이다.[3]

Scanner에 의해 digitize된 data를 computer 에 입력시키기 위해 scanner-to-computer interface 회로가 設計 製作되었다. Computer로 옮겨진 data는 cassette tape에 저장 되어 process를 기다리게 된다.

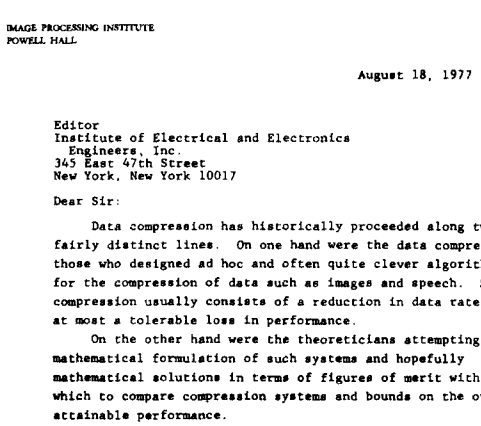


IMAGE PROCESSING INSTITUTE
POWELL HALL

August 18, 1977

Editor
Institute of Electrical and Electronics
Engineers, Inc.
345 East 47th Street
New York, New York 10017

Dear Sir:

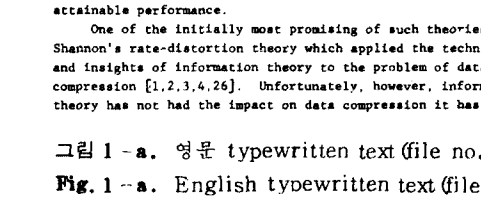
Data compression has historically proceeded along two fairly distinct lines. On one hand were the data compressors, those who designed ad hoc and often quite clever algorithms for the compression of data such as images and speech. Such compression usually consists of a reduction in data rate with at most a tolerable loss in performance.

On the other hand were the theoreticians attempting a mathematical formulation of such systems and hopefully mathematical solutions in terms of figures of merit with which to compare compression systems and bounds on the optimal attainable performance.

One of the initially most promising of such theories was Shannon's rate-distortion theory which applied the techniques and insights of information theory to the problem of data compression [1,2,3,4,26]. Unfortunately, however, information theory has not had the impact on data compression it has had

그림 1-a. 영문 typewritten text (file no.1)

Fig. 1-a. English typewritten text (file no.1).



문 고 부

대학 124-3467 1978. 10. 15.

수신 수신처참조

제목 74학년도 예산보고 지시

1. 74학년도 과별인 및 과별인이 설치 경영하는 대학의 예산서를 사립학교법 제 45조 및 동시행령 제 14조의 규정에 의거 1974. 3. 15. 이전에 보고하여야 하는 바

2. 에년 일부 대학에서 보고기일을 도과함으로써 일부에 지장을 초래하고 있으니 기일을 도과하는 일이 없도록 유의하시고

3. 예산서에 사학기관 재무회계 규칙 제 15조에 규정한 부속서류를 필히 첨부할것이며

4. 정당한 사유없이 보고기일을 도과함때는 관계직을 엄중 문책할 것이니 과별 유의 조치하시기 바랍니다. 끝.

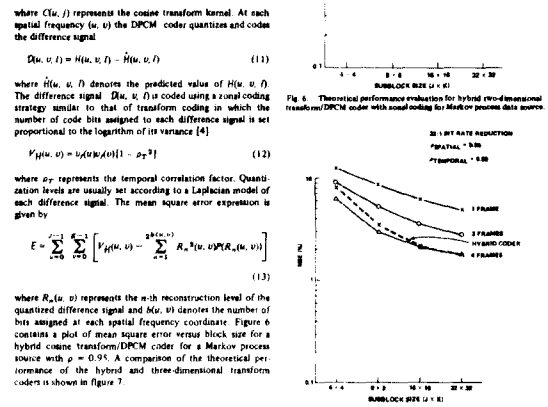
문 고 부 장 관

그림 1-b. 한글 typewritten text (file no. 2)

Fig. 1-b. Korean typewritten text (file no.2).

각 data source의 크기는 21.5㎝ × 27.5㎝를 표준으로 하였으나 scanner system의 特性으로 인하여 scan line의 길이는 左側 半分인 512 pel로 제한되었으며, 各 data file당 700 scan line이 實驗에 使用되었다.

各 data file에서 測定된 값은 white run과 black run 각각에 대한 run length 分布와 총 run 數, white pel과 black pel의 分布比率, 平均 run length,



where $C(u, j)$ represents the cosine transform kernel. At each spatial frequency (u, v) the DPCM coder quantizes and codes the difference signal

$$D(u, v, l) = H(u, v, l) - \hat{H}(u, v, l) \quad (11)$$

where $\hat{H}(u, v, l)$ denotes the predicted value of $H(u, v, l)$. The difference signal $D(u, v, l)$ is coded using a zonal coding strategy similar to that of transform coding in which the number of code bits assigned to each difference signal is set proportional to the logarithm of its variance [4]

$$V_H(u, v) = \nu_t(u)\nu_t(v)[1 - \rho_T^2] \quad (12)$$

where ρ_T represents the temporal correlation factor. Quantization levels are usually set according to a Laplacian model of each difference signal. The mean square error expression is given by

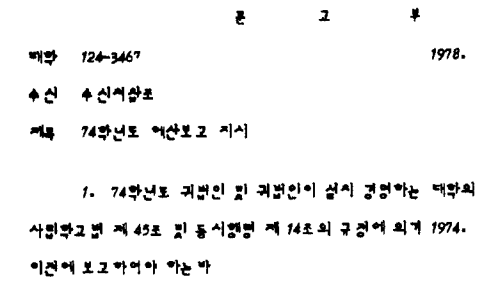
$$E = \sum_{u=0}^{J-1}\sum_{v=0}^{K-1}\left[V_H(u, v) - \sum_{n=1}^{2^{b(u,v)}} R_n^2(u, v)P(R_n(u, v))\right] \quad (13)$$

where $R_n(u, v)$ represents the n-th reconstruction level of the quantized difference signal and $b(u, v)$ denotes the number of bits assigned at each spatial frequency coordinate. Figure 6 contains a plot of mean square error versus block size for a hybrid cosine transform/DPCM coder for a Markov process source with $\rho = 0.95$. A comparison of the theoretical performance of the hybrid and three-dimensional transform coders is shown in figure 7.

Fig. 6. Theoretical performance evaluation for hybrid two-dimensional transform/DPCM coder with zonal coding for Markov process data source.

그림 1-c. 영문 printed text (file no. 3)

Fig. 1-c. English printed text (file no.3).



1978年 5月 電子工學會誌 第15卷 第2號

논문을 고찰하면 먼저 Sodestrand는 이를 위해 가산 출력기의 사용을 제안하였다[5]. 하지만 가산 출력기에서 one-pole rolloff 특성의 전산 출력기를 사용하므로 가산 출력기의 각 입력에 적당한 비의 일정 이득을 주기 위해서는 사용 가능 주파수의 범위가 좁아져야 된다는 단점이 있다. 그후 Mitra와 Aatre는 그들의 논문에서 최프의 다중성을 위해 저항 divider를 사용하였다[6]. 그들이 제시한 최프는 종래의 최프에 비해 낮은 저항비로 설계할 수 있으나 highpass notch filter의 실현이 용이하므로 lowpass의 경우 이득과 Q의 조절이 서로 독립되어 있지 않다는 단점이 있다. Kim과 Ra는 이러한 단점들을 극복하고 2차 전압 전달함수의 형태로 모든 가능한 특성을 독립적인 tuning으로 얻을 수 있는 새로운 configuration을 제시하였다[7]. 그러나 이 최프는 같은 특성의 여파기 설계에 있어 Mitra-Aatre의 최프에 비해 최대 저항비가 커서 저주파와 high Q에서의 이용 범위가 제한되어 있다.

본 논문은 이와 같은 문제점들을 해결하여 낮은 저항비로 독립적인 tuning이 가능한 multiloop의 최프를 제안하였다. 특히 이 최프는 지금까지 높게 저항 여파기에서 다루던 고주파 high Q의 영역 뿐아니라 높은 RC에 의거가 사용되는 영역인 저주파 low Q에도 그 사용이 가능하므로 Butterworth나 Chebyshev approximation을 이용한 highpass와 lowpass filter의 설계에도 적용될 수 있다.

2. 회로 해석

그림 1은 본 논문에서 제안된 최프이다. 이 최프를 해석하기 위해 (먼저 전산 출력기의 이득

$$A_1 = -\frac{A_0\omega_{p1}}{s+\omega_{p1}} = -\frac{GB_1}{s+\omega_{p1}} \quad (1)$$

하면 그림 2의 2개의 Subnetwork에서

$$V_2 = \frac{-a\omega_{p1}}{s+\omega_{p1}+\omega_1}V_1 + \frac{sGB_1}{s+\omega_{p1}+\omega_1}V_3 \quad (2)$$

$$V_3 = \frac{(d-b)GB_1}{s+\omega_{p1}}V_1 - \frac{cGB_1}{s+\omega_{p1}}V_4 \quad (3)$$

$$V_4 = -\frac{s+\omega_{p1}}{GB_1}V_3 + dV_1 \quad (4)$$

를 구할 수 있다. 이 때

$$\omega_1 = \frac{GB_1}{1+\frac{R_3}{R_{1a}//R_{1b}}} \quad (5)$$

$$a = \frac{R_2}{R_{1a}},\ b = \frac{R_2}{R_2+R_3},\ c = \frac{R_3}{R_3+R_4},\ d = \frac{R_{4a}}{R_{4a}+R_{4b}}$$
$$e = \frac{R_{4b}}{R_{4a}+R_{4b}} \quad (6)$$

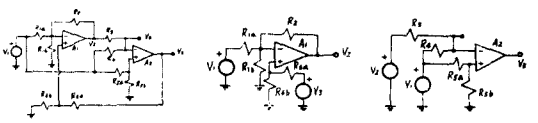


그림 1. 제안 회로
Fig. 1. Proposed network.

그림 2. 회로의 분리
Fig. 2. Subnetwork.

그림 1-d. 한글 printed text (file no. 4)

Fig. 1-d. Korean printed text (file no.4).

entropy/pel 등이다.

표1에서 보면 전체 畵面중에서 white pel이 차지하는 比率이 90~96% 정도임으로 white run의 特性들이 매우 重要함을 알 수 있다. White run의 平均 run length는 英文의 경우 平均 46, 한글의 경우,

平均 55이며 black run에 대해서는 각각 3.3, 3.8 이다. pel당 entropy값은 한글과 英文 image 사이에 큰 차이점이 없는 것으로 나타났다.

그림2는 typewritten 서류에 대한 white run length 分布圖인데, 한글 image의 경우에는 英文 image에 비해 해 run length가 긴 경우의 分布가 더 크고, 分布曲線의 기울기가 더 원만함을 볼 수 있다. 그림 3

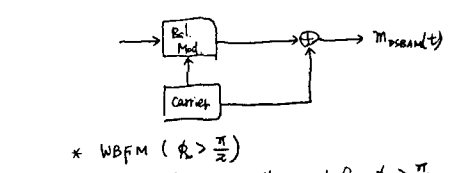


그림 1-e. 영문 handwritteen text (file no. 5)

Fig. 1-e. English handwritteen text (file no.5).

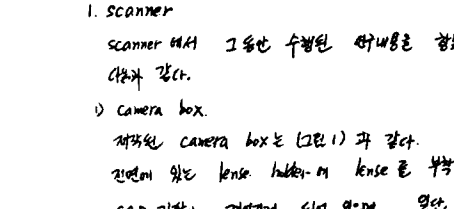
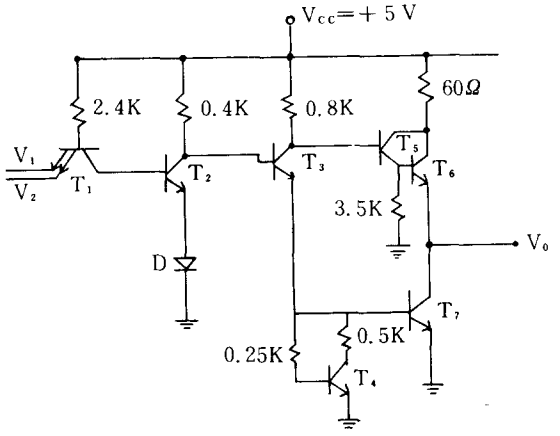


그림 1-f. 한글 handwritten text (file no. 6)

Fig. 1-f. Korean handwritten text (file no. 6).



(TTL AND gate)

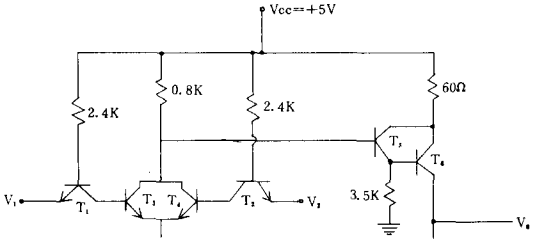


그림 1-g. 圖面 (file no. 7)

Fig. 1-g. Line drawing (file no. 7).

은 그림 2와 같은 data file의 black run length 分布圖인데, 한글과 英文 image 사이에 두드러진 차이점이 보이지 않는다.

여기서 Markov model의 妥當性 여부를 間接的으로 檢討할 수 있다.

즉 Laemmel code가 first-order Markov source에 대해서는 거의 optimal하다는 사실을 利用해서[2] 測定된 image data에 대한 Laemmel code(L_N-code) 결과를, 이 image data와 같은 평균 run length를 갖는 first-order Markov source data에 대한 L_N-code 結果와 比較하는 것이다. L_N-code에서 Markov source data와 실제 image data가 갖는 redundancy를 표2에 %값으로 나타내었다.

Redundancy R은 다음과 같이 定義된다.[2]

$$R = \frac{Q - h(L)}{h(L)} \quad \cdots\cdots (5)$$

여기서 Q는 사용한 code의 bit rate/pel, h(L)은 data의 entropy/pel 이다.

표2에서 볼 수 있듯이 Markov source를 L_N-code로 coding 했을 때 white와 black run 모두 그 re-

dundancy가 8% 이내의 값을 갖는다. 이것은 L_N-code가 first-order Markov source에 대해서 거의 optimal한 特性을 가짐을 말해 준다.

그런데 Hasler code(H_N-code)에 대한 실제 image data의 redundancy 값이 英文의 경우 平均 33%, 한글의 경우 平均 31%가 되는 것으로 나타났다.

표 1. 측정된 image data의 특성

Table 1. Characteristics for some typical image data.

File No.	Number of runs		Probability (%)		Mean run length		entropy/pel (bit/pel)		Remarks
	N_W	N_B	P (w)	p(B)	E (w)	E (B)	h (w)	h (B)	
1	7,657	7,019	0.937	0.063	43.80	3.21	0.0945	0.8683	English typewritten
2	4,541	3,899	0.956	0.044	75.36	4.03	0.0625	0.7477	Korean 〃
3	11,031	10,408	0.910	0.090	29.52	3.09	0.1469	0.8817	English printing
4	8,012	7,321	0.927	0.073	41.42	3.56	0.1187	0.8203	Korean 〃
5	4,941	4,261	0.961	0.039	69.61	3.27	0.0773	0.8189	English handwritten
6	6,957	6,279	0.946	0.054	48.64	3.10	0.1034	0.8624	Korean 〃
7	4,564	3,881	0.967	0.033	75.82	3.05	0.0738	0.8510	line drawing

(a) White run for file no. 1.

(b) White run for file no. 2.

그림 2. White run length의 확률분포도

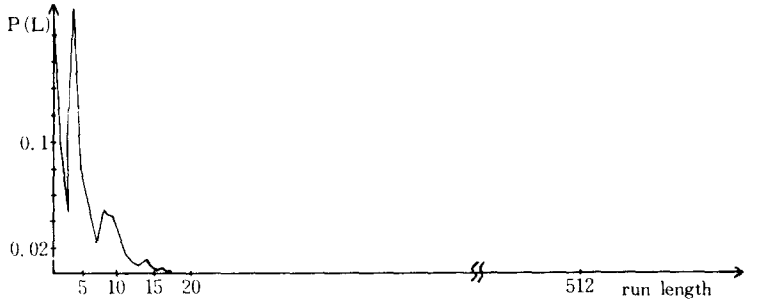
Fig. 2. Probability distribution of white run length.

그러나 실제 image data의 white run은 L_N-code에 대한 그 redundancy값이 英文의 경우 70～115%, 한글의 경우 73～106%나 되어, white run에 대해서는 한글과 英文 image 모두가 Markov model의 適用이 適合하지 않음을 보여 주고 있다.

이것은 한글과 英文 image의 white run length 分布가 그 分布의 모습은 서로 다르지만 exponential 分布보다는 negative-power 分布에 더 가깝다는 것을 말해 주고 있다.

그러나 black run의 경우에는 그 redundancy 값이

(a) Black run for file no. 1.



(b) Black run for file no. 2.

그림 3. Black run length의 확률분포도

Fig. 3. Probability distribution of black run length.

표 2. Laemmel code에 대한 redundancy 값

Table 2. Redundancy values for Laemmel code.

file No.	White run		Black run		Remarks
	Markov source	measured image	Markov source	measured image	
1	7.6(%)	115(%)	3.3(%)	10.2(%)	English typewritten
2	7.8	105	6.5	18.7	Korean 〃
3	7.5	83	3.2	10.7	English printing
4	7.9	73	4.4	13.3	Korean 〃
5	8.2	70	3.5	19.2	English handwritten
6	6.9	74	3.2	13.6	Korean 〃
7	7.9	57	3.2	16.7	line drawing

英文의 경우 10.2 ~ 19.2 %, 한글의 경우 13.3~18.7 %의 값을 갖기 때문에, black run에 대해서는 Markov model의 適用이 매우 妥當함을 보여 주고 있다. 이것은 한글과 英文 image의 black run이 exponential 分布에 近似한 分布를 가짐을 말해 준다.

그러데 실제로는 white와 black run을 區分하지 않고, white run에 optimal한 code로 같이 coding하여 사용하는 경우가 많다. 이럴때에는 white run의 特性이 크게 작용함으로, image data를 Markov model로 假定하는 것은 한글과 英文 image 모두에게 있어서 適合하지 않다고 하겠다.

4. Run length code의 特性比較

본 章에서는 앞 章에서 測定된 image data의 여러 特性들을 사용하여, 代表的인 네가지 기존 run length code의 coding 特性을 比較하였다.

選擇된 네가지 code는 modified-white block skipping code[6] (M-WBS code), linear code의 일종인 Laemmel code (L_N-code), logarithmic code의 일종인 Hasler code (H_N-code), 그리고 Truncated-Huffman code[7]이다.

각 image data를 選擇된 위의 네가지 code로 coding했을경우, 우리가 얻을 수 있는 최소의 bit rate Q_{min}. 과 그때의 optimum coding block size N_{opt}. 을 computer simulation으로 測定한 結果가 표 3과 같다. 여기서 N_{opt}.은 block run과 white run에 同

표 3. 네가지 code에 대한 minimum bit rate ($Q_{min.}$)와 optimum block size(N_{opt}).

Table 3. Minimum bit rate and optimum block size for 4 different codes.

file No.	experimental entropy	modified WBS code		L_N- code		H_N - code		Truncated Huffman code
	entropy	Qmin	Nopt.	Qmin	Nopt.	Qmin	Nopt.	Q_{min}
1	0.1433	0.233	5	0.321	6	0.172	1	0.1698
2	0.0925	0.164	6	0.209	7	0.123	1	0.1104
3	0.2130	0.338	5	0.415	5	0.241	1	0.2294
4	0.1697	0.303	6	0.321	6	0.198	1	0.1852
5	0.1062	0.212	8	0.218	7	0.134	1	0.1264
6	0.1447	0.261	7	0.289	6	0.174	1	0.1634
7	0.0995	0.239	9	0.193	7	0.133	1	0.1288

一한 code를 適用한 경우의 optimum block size 이다.

표 3에 보면 한글과 英文 image 모두에서 Truncated-Huffman code가 가장 우수하고, 그 다음이 H_N-code(그중 H_1-code), M-WBS code, L_N-code (그 중 L_6-code) 順으로 bit rate가 증가하고 있다. 그러나 all-white line이 별로 없는 line drawing image에 대해서는 M-WBS code보다 L_N-code가 더 좋은 特性을 나타내고 있다.

또한 한글과 英文 file에 관계없이 H_N-code와 L_N-code의 평균 optimum block size는 각각 1과 6임을 알 수 있다.

표 3의 각 code로 부터 얻을 수 있는 data 壓縮比의 平均値를 정리하면 표 4와 같다.

표 4. 각 code의 reduction factor의 평균치

Table 4. Average values of reduction factor in 4 different codes.

File	MWBS code	L_6-code	H_1-code	Truncated Huffman code
English	3.92	3.18	5.40	5.70
Korean	4.43	3.70	6.03	6.52
Line drawing	3.23	3.81	5.54	5.67

표 4로부터 H_1-code가 Truncated-Huffman code에 매우 가까운 우수한 壓縮比를 가지며, 같은 code에 있어서도 英文 image 보다 한글 image에서 조금 더 큰 壓縮比를 얻을 수 있음을 알 수 있다.

특히 file #1(英文 image)과 file #2(한글 image)에서 block size의 變化에 따른 각 code의 bit rate 變化는 그림 4, 그림 5와 같다. 이 그림에서보

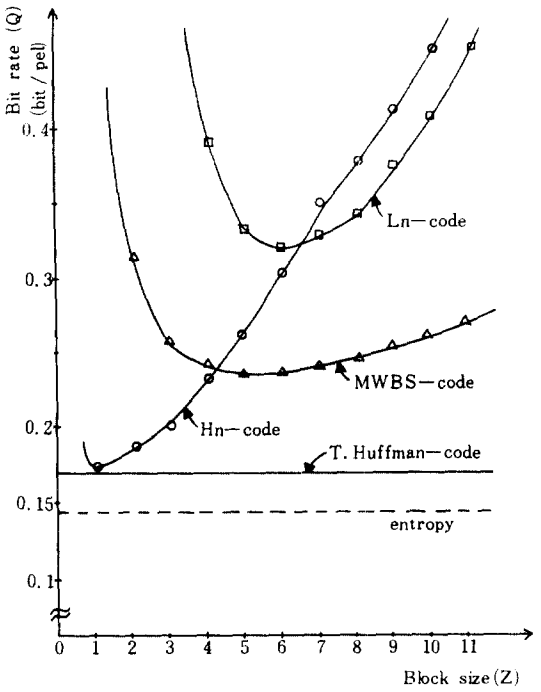


그림 4. 영문 image에 대해 block size에 따른 bit rate 변화

Fig. 4. Bit rate variation by the block size for English typewritten text.

면 한글image가 英文 image에 비해 block size 의 變化에 덜 민감한 특징을 찾고 있음을 알 수 있다.

각 code의 効率性을 알아보기 위해, optimum block size때 각 code의 redundancy를 평균한 結果, Truncated-Huffman code가 0.2, H_1-code는 0.3, M-WBS code는 0.8, 그리고 L_6-code는 0.8 정도인 것으로 나타났다.

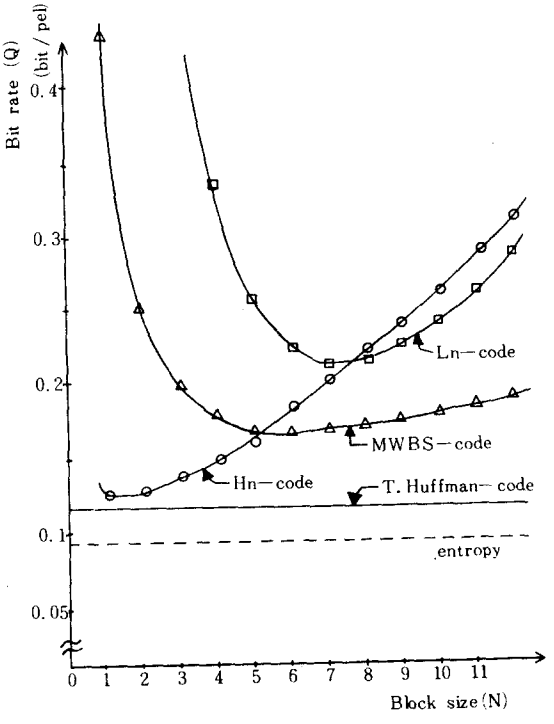


그림 5. 한글 image에 대해 block size에 따른 bit rate 변화

Fig. 5. Bit rate variation by the block size for Korean typewritten text.

5. 結 論

한글과 英文 image data의 統計的 特性들이 測定比較되었다. run length의 確率分布는 대체로 비슷하지만, white run에서는 한글 image가 英文 image보다 negative-power 分布의 性向이 좀 더 강함을 알 수 있었다. 그러나 한글과 英文 image 모두 그 run length 分布가 Markov model에 의한 exponential 分布와는 상당한 차이가 있음이 확인되었다.

네가지 run length code에 대한 여러가지 coding 特性을 比較한 結果, 같은 code로서도 英文 image에서 보다 한글 image에서 data 壓縮比가 더욱 크며, code block size에 덜 민감한 사실을 볼 수 있었다. 이런 사실들은 英文 image의 特性을 調査 研究하여 얻어진 여러 coding 方法들이, 한글 image에 대해서는 오히려 더 効率的임을 말해 주고 있다. 그러므로 英文 image에 의해 얻어진 結果들이 한글 image에 그대로 사용되어도 무방하다고 할 수 있다. 그러나 한글 image가 갖는 特性을 최대한 살리기 위해서는 한글 image에 적합한 coding 方法이 새로이 研究 開發되어야 할 것이다.

比較된 code 중에서는 Truncated-Huffman code와 Hasler code(H_1-code)의 성능이 우수했다.

參 考 文 獻

1. J. Capon "A probabilistic model for run length coding of pictures", IRE Trans. vol. IT-5, pp. 157~163, Dec. 1959.
2. H. Meyr, H. G. Rosdolsky and T. S. Huang "Optimum run length codes," IEEE Trans., vol. Com-22, No. 6. pp. 826~835, June 1974.
3. 한국과학원, "전화 통신망을 이용한 FAX system의 研究開發", July 1978.
4. T. S. Huang "Coding of two-tone images", IEEE Trans. vol Com-25, No.11, pp. 1406~1424, Nov. 1977.
5. Kie-Bum Eom "A study on run length codes -comparison and implementation", M.S. thesis, KAIS, Feb. 1978.
6. T. S. Huang and Shahid Hussain "Facsimile coding by skipping white", IEEE Trans. vol. Com-23, No. 12, pp. 1452~1460, Dec. 1975.
7. H. G. Musmann and Dieter Preuss "Comparison of redundancy reducing codes for facsimile transmission of documents", IEEE Trans. vol. Com-25, No. 11, pp. 1425~1433, Nov. 1977.