

계층적 픽셀 예측과 컨텍스트 적응적 산술 부호화를 이용한 이미지 적응 무손실 압축

*심재훈¹ 김세윤² 조남익¹¹서울대학교 전기정보공학부 뉴미디어통신공동연구소²GAUSS Labs

*joinjhs@ispl.snu.ac.kr light4u@gmail.com nicho@snu.ac.kr

Image-Adaptive Lossless Compression based on Hierarchical Prediction

*Shim, Jae Hoon¹ Kim, Seyun² Cho, Nam Ik¹¹Department of ECE, INMC, Seoul National University

요약

본 논문에서는 계층적 픽셀 예측과 컨텍스트 적응적 산술 부호화를 이용한 이미지 적응 무손실 압축 알고리즘을 제안한다. 입력 RGB 이미지는 먼저 가역적 색상 변환이 적용된다. Y 채널 이미지는 기존의 무손실 압축 인코더로 압축되고, U와 V 채널 이미지는 Y 채널 이미지를 기반으로 예측된다. 원본과의 차이는 컨텍스트 적응적 산술 부호화를 통해 압축된다. 본 논문에서 제안된 알고리즘에서는 입력 이미지의 성질에 따라 산술 부호화에 사용되는 인코더의 개수를 적응적으로 변화시킨다. 또한 저주파 성분에 상대적으로 많은 자원을 집중시킴으로써 압축 성능을 향상시켰다. 제안된 방법은 기존에 사용되던 압축 방식들과 비교했을 때에도 의미 있는 성능을 보였다.

1. 서론

이미지 압축의 중요성은 사진과 영상 기술이 발전함에 따라 지속적으로 증가해왔다. 딥 러닝을 사용한 압축 방식들은 성능 면에서 기존 압축 방식들을 뛰어넘고 있지만, 딥 러닝이라는 분야 특성상 많은 자원을 소모한다. 현재 대부분의 모바일 기기들은 딥 러닝을 사용할 만한 충분한 전력을 공급할 여유가 없기 때문에, 아직도 JPEG과 같은 기존 코덱들이 널리 사용되고 있다.

앞서 언급했듯이, JPEG은 가장 많이 사용되는 압축 방식 중 하나이다. 하지만 손실 압축이라는 특성상 JPEG 방식을 사용하면 저장된 이미지를 복원했을 때 아티팩트가 발생하여 원본 이미지가 손상된다. 이와 같은 손실 압축 방식은 압축률이 높지만, 복원된 이미지가 원본과 다르다는 단점이 있다. 따라서 이 방식은 의료나 과학 영상에서와 같이 이미지 손상이 허용되지 않는 상황에서 적절한 압축 방식이 아니다. 이 경우 무손실 압축을 사용하는 편이 더 적절하다고 할 수 있다.

디지털 이미지의 발명으로 다양한 무손실 이미지 압축 방식들이 제안되어왔다. 그중에서도 PNG, JPEG, JPEG-LS, LOCO-I, CALIC 등의 코덱들과 JPEG2000, BPG, JPEG-XR, WebP, FLIF와 같은 압축 방식들의 무손실 압축 버전들은 가장 널리 쓰이는 방식들이다.

원본 영상이 복원 영상과 완전히 일치한다는 점 때문에 무손실 압축 방식에서는 손실 압축과 다른 접근이 이루어진다. 대표적인 것은 색상 변환으로, RGB 채널 사이의 상관관계를 줄이기 위해 일반적으로

YUV나 YCbCr 등으로의 색상 변환이 이루어지지만 일반적으로 이러한 변환은 비가역적이기 때문에 무손실 압축 방식에서는 RCT라고 불리는 가역적 색상 변환을 적용한다.

RCT 이외에도 산술 부호화는 무손실 압축에서 널리 쓰이는 기술이다. 원본 이미지의 픽셀 확률 분포를 예측하고 예측한 정보를 바탕으로 이미지를 효율적으로 압축하고 완전히 복원할 수 있도록 하는 것이 산술 부호화의 역할이다. 하지만 일반적인 자연적 이미지들은 복잡한 경우가 많기 때문에 확률 분포 예측이 쉽지 않다. 따라서 본 논문에서는 보다 복잡한 이미지의 확률 분포 예측이 가능한, LCIC에서 제안된 적응적 산술 부호화를 사용하였다. LCIC에서는 컨텍스트에 따라 다른 산술 부호화를 적용하였다.

본 논문에서는 산술 코더의 개수를 적응적으로 변화시키는 것으로 LCIC를 발전시키고자 하였다. 산술 코더의 개수 K가 증가할수록 압축 성능이 늘어날 것으로 예상할 수도 있지만, 실제 상황에서는 K가 너무 커진다면 각 코더에서 사용할 수 있는 데이터의 수가 줄어들기 때문에 성능이 감소하게 된다. 지역적 픽셀의 분산(픽셀 활동)은 고주파 성분이 큰 이미지의 경우 높고, 저주파 성분이 큰 이미지의 경우에는 낮게 형성되는데, 이를 이용하여 본 논문에서는 이미지의 픽셀 활동을 이용하여 적절한 K를 결정하는 K 결정기를 제안하였다. 픽셀 활동이 입력으로 주어지면 K 결정기는 이를 이용하여 히스토그램을 그리고, 이 히스토그램의 평균값을 이용해 K를 결정하게 된다. 또한 산술 코더의 절반을 초저주파 성분에 할당하는 것으로 압축 효율을 더욱 증가시켰다.

2. 사용된 기존 방법

2.1 Hierarchical Decomposition

입력된 RGB 영상을 YUV로 변환하면 휘도 성분인 Y와 색차 성분인 U, V로 분리되는데, 그 중 색차 성분 $X \in U, V$ 는 각각 두 개의 소영상 X_e, X_o 로 분리된다. X_e 는 짝수 행의 픽셀 데이터를 의미하며, X_o 는 홀수 행의 픽셀들을 말한다. X_e 는 그 후 열 기준으로 분리되어 $X_e^{(2)}, X_o^{(2)}$ 로 나누어진다. 홀수 행의 픽셀들과 짝수 행의 픽셀들은 상관관계가 매우 크기 때문에 하나의 데이터를 이용하여 다른 하나를 압축할 수 있게 된다. $X_o^{(2)}$ 는 $X_e^{(2)}$ 를 이용해서 압축되고 $X_o^{(1)}$ 는 $X_e^{(1)} = X_e^{(2)} \cup X_o^{(2)}$ 를 이용해서 압축이 진행된다. 색차 성분은 U, V의 두 개가 있기 때문에 최종적으로 압축되는 이미지들은 $U_o^{(1)}, V_o^{(1)}, U_o^{(2)}, V_o^{(2)}$ 의 네 장이 된다.

2.2 픽셀 예측

X_o 의 각 픽셀들 $x_o(i, j)$ 에 대해서 수평 예측치와 수직 예측치가 정의되는데, 수평 예측치는 바로 왼쪽의 픽셀을 사용하고 수직 예측치는 위와 아래의 픽셀들을 사용한다. 수평 예측치 $\hat{x}_h(i, j)$ 와 수직 예측치 $\hat{x}_v(i, j)$ 는 아래와 같이 정의된다.

$$\begin{aligned} \hat{x}_h(i, j) &= x_o(i, j-1) \\ \hat{x}_v(i, j) &= \text{round}\left(\frac{x_e(i, j) + x_e(i+1, j)}{2}\right) \end{aligned} \quad (1)$$

수평 예측치는 수평적인 성분이 큰 경우에만 사용되고, 대부분의 경우 수직적인 성분이 많기 때문에 수직 예측치를 사용하게 된다.

2.3 산술 부호화

산술 부호화는 엔트로피 부호화의 한 종류이다. 확률 분포를 예측할 수 있다면 산술 코더는 자주 나올 것으로 보이는 값에 작은 비트를 할당하고 그렇지 않은 경우 큰 비트를 할당하는 것으로 압축 영상 크기의 기대값을 낮춘다. 샤논의 정리[8]에 의하면 정보 손실이 없는 경우 압축된 비트 스트림의 크기에는 하한값이 존재한다. 예측된 확률 분포가 실제 분포에 근접할수록 압축된 비트 스트림의 크기는 이 하한값에 근접하게 된다.

이미지에는 고주파 성분이 강한 픽셀과 저주파 성분이 강한 픽셀이 있는데, 두 종류의 픽셀의 분포는 확실히 다르기 때문에 다른 종류의 픽셀들에 다른 산술 코더를 사용하는 경우 압축률을 높일 수 있다. 산술 코더를 선택하는 기준은 pixel activity를 기준으로 결정된다. 자세한 과정은 다음 장에서 서술한다.

3. 방법 제안

3.1 K 결정기

pixel activity $\sigma(i, j)$ 는 $X_e^{(2)}$ 에서 정의되는데, 이 값은 분리된 이미지에서의 지역적 분산을 의미한다. 이 pixel activity 값을 기준으로 하여 각 픽셀들은 서로 다른 산술 코더를 이용하여 압축된다.

$$\sigma(i, j) = |x_e(i, j) - x_e(i+1, j)| \quad (2)$$

$$q_{n-1} \leq \sigma(i, j) < q_n \quad (3)$$

본 논문에서는 소영상 $X_o^{(1)}, X_o^{(2)}$ 의 각 픽셀에 대한 컨텍스트 $c(i, j)$ 를 정의한다. 총 압축되는 소영상의 숫자는 $U_o^{(1)}, V_o^{(1)}, U_o^{(2)}, V_o^{(2)}$ 의 네 장이므로 네 가지의 컨텍스트가 아래와 같이 정의된다.

$$c_{u1}(i, j) = |u_o^{(1)}(i, j) - u_o^{(1)}(i+1, j)| \quad (4)$$

$$c_{u2}(i, j) = |u_o^{(2)}(i, j) - u_o^{(2)}(i+1, j)| \quad (5)$$

$$c_{v1}(i, j) = |v_o^{(1)}(i, j) - v_o^{(1)}(i+1, j)| \quad (6)$$

$$c_{v2}(i, j) = |v_o^{(2)}(i, j) - v_o^{(2)}(i+1, j)| \quad (7)$$

$U_o^{(1)}, V_o^{(1)}$ 의 크기는 $H/2 \times W$ 이기 때문에 (4), (6)에서의 i, j 의 범위는 $i = 1, \dots, H/2, j = 1, \dots, W$ 이다. 또한 $U_o^{(2)}, V_o^{(2)}$ 의 크기는 $H/2 \times W/2$ 이므로, (5), (7)에서의 i, j 의 범위는 $i = 1, \dots, H/2, j = 1, \dots, W/2$ 가 된다.

컨텍스트 값이 높다는 것은 픽셀 값의 변화율이 높다는 것이고, 다시 표현하면 고주파 성분이 높다는 것이다. K 결정기는 정의된 네 개의 컨텍스트 값들을 이용해서 이미지의 고주파 성분이 얼마나 높은지 결정한다. 결정의 기준이 되는 지표는 아래와 같이 세 종류를 설정했다.

$$\text{avgmean} = \frac{1}{6}(2\overline{c_{u1}} + \overline{c_{u2}} + 2\overline{c_{v1}} + \overline{c_{v2}}) \quad (8)$$

$$\text{minmean} = \min(\overline{c_{u1}}, \overline{c_{u2}}, \overline{c_{v1}}, \overline{c_{v2}}) \quad (9)$$

$$\text{maxmean} = \max(\overline{c_{u1}}, \overline{c_{u2}}, \overline{c_{v1}}, \overline{c_{v2}}) \quad (10)$$

여기서 $\bar{c} = \sum_{i,j} c(i, j)$ 로 정의했다. (8)에서 계수에 가중치가 붙은 이유는 이미지의 크기가 다름을 보정해주기 위함이다.

표 1 각 지표 값들과 최적의 K 값 사이의 스피어맨 상관계수

지표	avgmean	minmean	maxmean
상관계수	0.754	0.755	0.748

표 1은 Kodak 데이터셋의 20개 이미지 별 최적의 K값(가장 높은 압축률을 보이는 K값)과 각 지표들과의 스피어맨 상관계수의 절대값[12]을 나타낸 표이다. 스피어맨 상관계수는 값이 높을수록 상관관계가 큰 상관계수이다. 0.6보다 값이 높다면 두 데이터 간에 상관관계가 매우 크다고 볼 수 있다. 앞서 설정한 세 지표 모두 0.6보다 크기 때문에 설정한

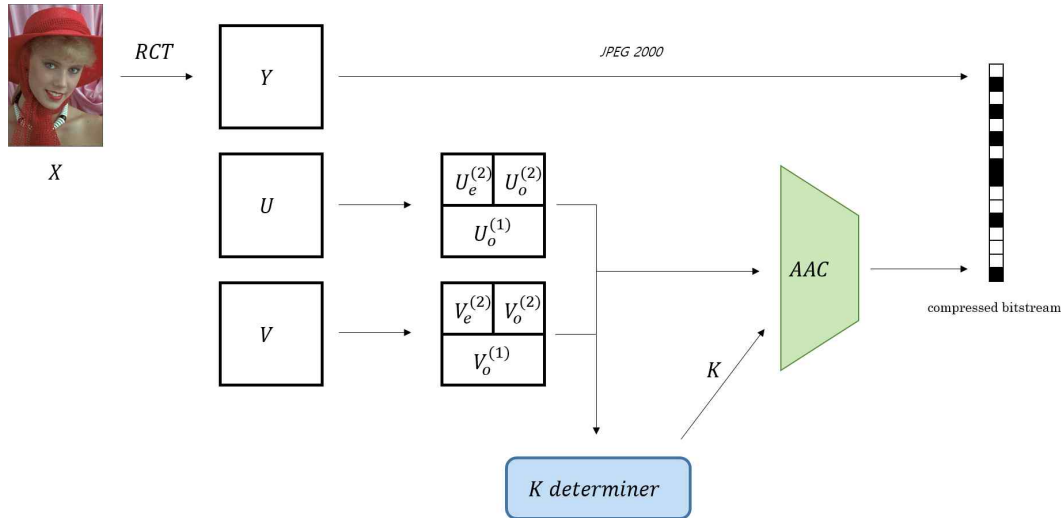


그림 1 전체 압축 과정

지표로 K 를 결정할 수 있는 근거가 된다. 본 논문에서는 상관계수가 가장 큰 minmean을 이용하여 K 를 결정하였다. K 의 결정 과정은 산술 코더를 결정하는 과정이 나타난 식 (3)과 거의 유사한 과정을 거친다. 경계 값들은 실험적으로 결정되었고, K 의 범위는 $K \in \{4, 5, 6, \dots, 14\}$ 로 설정하였다.

3.2 초저주파 성분 산술 부호화 할당

자연적인 영상의 경우 저주파 성분이 대부분을 차지하고, 고주파 성분은 상대적으로 희귀하다. 예시로 Kodak 데이터셋의 경우 80% 이상의 pixel activity σ 값들이 10보다 작고, 95% 이상의 σ 값들은 20보다 작았다. 다시 말하면 20부터 255의 값들은 5% 미만의 비율을 가진다고 볼 수 있다. [6] 논문에서는 각각의 산술 코더가 같은 수의 값들을 압축하도록 설정되어 있지만, 이 방법을 사용할 경우 저주파 구간의 여러 값들이 하나의 산술 코더로 압축된다는 단점이 있다.

더 정확한 저주파 성분의 확률 분포를 얻기 위해, 본 논문에서는 절반의 산술 코더를 초저주파 성분에 할당한다. 구체적으로는, $K/2$ 개의 산술 코더는 각각 pixel activity 값 $0, 1, \dots, K/2 - 1$ 에 대응되게 된다. 나머지 절반의 산술 코더는 기존 방식을 적용한다.

3.3 전체 압축 과정

전체 압축 과정은 그림 1에 표현되었다. 먼저 RGB 이미지는 YUV 채널로 가역적 색상 변환이 이루어지고, Y채널은 JPEG2000 코덱으로 압축된다. U와 V채널은 $U_e^{(1)}, V_e^{(1)}, U_o^{(1)}, V_o^{(1)}, U_o^{(2)}, V_o^{(2)}$ 의 소영상들로 분리된다. 그 후 $U_o^{(1)}, V_o^{(1)}, U_o^{(2)}, V_o^{(2)}$ 를 이용하여 minmean을 계산하여 이를 기반으로 K 결정기가 K 를 결정하고, AAC가 각 소영상들을 압축해서 전송하게 된다.

4. 실험 결과

4.1 실험 세부 사항

실험에는 DIV2K[2] 데이터셋과 Kodak[1] 데이터셋에서 추출한 이미지들이 사용되었다. 두 데이터셋 모두 영상 처리 분야에서 널리 사용되고 있다. DIV2K 데이터셋은 1000개의 고화질 2K 이미지들로 이루어져 있고, 800개는 training set, 100개는 validation set, 100개는 test set이다. 실험에는 test set에서 20개 이미지를 추출하여 사용하였다. Kodak 데이터셋은 25개의 512×768 크기의 이미지들로 이루어져 있는데, 이 중 20개를 추출하여 사용하였다.

4.2 실험 결과

표 2 DIV2K 데이터셋에서의 압축 성능 비교표

방법	bpp	압축 시간
PNG	12.69	0.001
BPG	13.26	1.364
WebP	9.34	12.785
FLIF	8.74	17.831
LCIC	9.35	4.064
제안 방법	9.34	4.104

표 2는 DIV2K 데이터셋에서의 다른 무손실 압축 방식들과의 압축률, 압축 시간 비교표이다. PNG[4]와 BPG lossless[3]의 경우 2초 이내의 짧은 압축 시간을 보여주지만, 압축률은 13bpp(bits per pixel)을 넘기 때문에 고화질 이미지를 저장하는 데에 적합하지 않다. WebP[13]와 FLIF[11]는 압축률이 굉장히 높은 반면, 실제 활용에는 적합하지 않은데 이는 고화질 이미지 하나 당 압축 시간이 10초를 넘어가기 때문이다. 제안된 압축 방식은 BPG와 PNG에 비해 효율적인 압축과, WebP와 FLIF에 비해 빠른 속도를 자랑한다. 또한 LCIC보다 개선된 압축률을 보이고 있다. 제안된 방식은 속도나 압축률 둘 다 뛰어나다고 볼 수 있다.

4.3 조건 변화 실험

본 논문에서 제안하는 두 가지 방법, K 결정기와 초저주파 성분 산술 부호화 할당은 압축률 향상에 모두 기여하고 있다.

표 3 DIV2K 데이터셋에서의 조건 변화 실험.

방법	K=4	K=6	K=14
LCIC(기준)	10.0640	10.0470	10.1293
초저주파 할당	10.0646	10.0439	10.0362
초저주파 + K 결정기	10.0339		

표 4 Kodak 데이터셋에서의 조건 변화 실험

방법	K=4	K=6	K=14
LCIC(기준)	9.0917	9.0007	9.2186
초저주파 할당	9.0483	8.9959	8.9551
초저주파 + K 결정기	8.9556		

표 3과 4는 DIV2K와 Kodak 데이터셋에서 K를 변화시켰을 경우와 본 논문에서 제안된 방법들을 적용했을 때의 bpp 변화를 나타낸 표이다. 기준이 되는 LCIC는 K=6의 경우 가장 뛰어난 성능을 보이고, 실제로 해당 논문에서도 K=6을 설정하였다. 초저주파 할당을 적용했을 경우 K=14로 갈수록 성능이 향상되는 것을 볼 수 있다. K 결정기는 각 이미지 별로 최적의 K를 찾기 때문에 적용한 경우(맨 아랫줄) 더욱 향상된 성능을 보인다. Kodak 데이터셋의 경우 K=14일 때보다 K 결정기를 사용했을 때 성능이 매우 작게나마 감소하지만, 이는 K 결정기가 항상 완벽한 선택을 할 수 없기 때문이다. 다만 이 경우에도 성능의 차이는 무시할 수 있을 만큼 작기 때문에 결과적으로 본 논문에서 제안한 방식들은 효과적으로 압축률을 상승시킨다고 볼 수 있다.

5. 결론

본 논문에서는 적응적 산술 코더의 개수 K를 변화시키는 무손실 압축 방식을 제안한다. 최적의 K를 찾기 위해, 분리된 소영상들에서 컨텍스트들을 정의하고 이를 지표로 이용한다. 본 논문에서는 또한 산술 코더를 초저주파 대역에 집중시키는 것으로 압축 성능을 더욱 향상시켰다. 제안된 방법은 압축된 데이터의 크기나 압축 시간으로 볼 때 기존 방식들보다 범용성이 뛰어나다고 볼 수 있다. K 결정기와 초저주파 할당 산술 부호하는 모두 성능 상승에 기여하며, 고화질 이미지의 경우에도 뛰어난 성능을 보인다.

감사의 글

본 연구는 산업통상자원부와 한국산업기술진흥원의 “지역혁신클러스터 육성사업(R&D, P0002072)”과 2022년도 BK21 FOUR 정보기술 미래인재 교육연구단에 의하여 지원되었음.

This research was financially supported by the Ministry of Trade, Industry, and Energy (MOTIE), Korea, under the "Regional Specialized Industry Development Program (R&D, P0002072)" supervised by the Korea Institute for Advancement of Technology (KIAT).P0002072, and supported by the BK21 FOUR program of the Education and Research Program for Future ICT Pioneers, Seoul National University in 2022.

참고문헌

[1] Kodak Lossless True Color Image Suite . Accessed: Jun. 21, 2021. [Online]. Available: <http://r0k.us/graphics/kodak/>.

[2] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. IEEE, July 2017.

[3] Umar Albalawi, Saraju P Mohanty, and Elias Kougiianos. A hardware architecture for better portable graphics (bpg) compression encoder. In 2015 IEEE International Symposium on Nanoelectronic and Information Systems, pages 291-296. IEEE, 2015.

[4] Thomas Boutell and T Lane. Png (portable network graphics) specification version 1.0. Network Working Group, pages 1-102, 1997.

[5] Frederic Dufaux, Gary J Sullivan, and Touradj Ebrahimi. The jpeg xr image coding standard [standards in a nutshell]. IEEE Signal Processing Magazine, 26(6):195-204, 2009.

[6] Seyun Kim and Nam Ik Cho. Hierarchical prediction and context adaptive coding for lossless color image compression. IEEE Transactions on image processing, 23(1):445-449, 2013.

[7] Soo-Chang Pei and Jian-Jiun Ding. Improved reversible integer-tointeger color transforms. In 2009 16th IEEE International Conference on Image Processing (ICIP), pages 473-476. IEEE, 2009.

[8] Shannon. A mathematical theory of communication. In Bell System Technical Journal, volume 27, pages 379--423, 623-656. IEEE, 1948. [9] Athanassios Skodras, Charilaos Christopoulos, and Touradj Ebrahimi.

The jpeg 2000 still image compression standard. IEEE Signal processing magazine, 18(5):36-58, 2001.

[10] Athanassios Skodras, Charilaos Christopoulos, and Touradj Ebrahimi. The jpeg 2000 still image compression standard. IEEE Signal processing magazine, 18(5):36-58, 2001.

[11] Jon Sneyers and Pieter Wuille. Flif: Free lossless image format based on maniac compression. In 2016 IEEE International Conference on Image Processing (ICIP), pages 66-70. IEEE, 2016.

[12] Charles Spearman. The proof and measurement of association between two things. 1961.

[13] WebEngines Blazer Platform Version. 1.0 hardware reference guide, xp-002202892, network engines. Inc., Jun. 1:92, 2000.

[14] Marcelo J Weinberger, Gadiel Seroussi, and Guillermo Sapiro. The locoi lossless image compression algorithm: Principles and standardization into jpeg-ls. IEEE Transactions on Image processing, 9(8):1309-1324, 2000.

[15] Xiaolin Wu and Nasir Memon. Calic-a context based adaptive lossless image codec. In 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, volume 4, pages 1890-1893. IEEE, 1996.